Tomorrow's Earth
pp. 287 & 304

Developed economies enable gender differences *p. 307*

Opium poppy genome reveals metabolic secrets *p. 343*

Science

$15
19 OCTOBER 2018
sciencemag.org

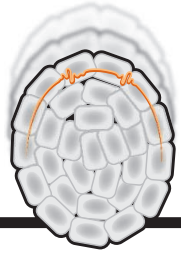
AAAS

NOT SO NUTS

Promise and risk in treating food allergies with their triggers

p. 278

CONTENTS



290 & 339
Cellular group mobility

19 OCTOBER 2018 • VOLUME 362 • ISSUE 6412

Jair Bolsonaro, front-runner for Brazil's presidency, is a scary prospect to some.

CREDITS (FROM TOP): C. BICKEL/*SCIENCE*; VICTOR MORIYAMA/STRINGER/GETTY IMAGES

NEWS

IN BRIEF

IN DEPTH

FEATURE

INSIGHTS

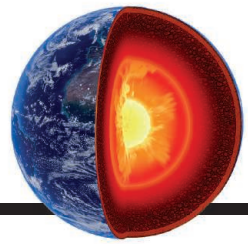
POLICY FORUM

PERSPECTIVES

BOOKS *ET AL.*

LETTERS

CONTENTS



294 & 329
Core constraints

19 OCTOBER 2018 • VOLUME 362 • ISSUE 6412

ON THE COVER

Between 1 and 2% of people in the United States are allergic to peanuts. A new treatment strategy exposes affected children to peanuts or peanut proteins in hopes of desensitizing their immune system. This approach has gained popularity as doctors in private practice offer it and two commercial products are nearing potential approval. But it also comes with risks and uncertainties. See page 278.
Photo: Mypurgatoryyears/iStock.com

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. *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

EDITORIAL

Network for safe and secure labs

The current outbreak of Ebola virus in the Democratic Republic of the Congo is a reminder that dangerous diseases exist in many corners of the world and that they can cause substantial human suffering and financial devastation locally and internationally. In response, institutions and nations are constructing maximum biocontainment laboratories (MCLs) to address these threats. MCLs operate at the highest level of biological containment to diagnose, perform research on, and validate cures for life-threatening diseases like Ebola. There are more than 50 MCLs that are operational, under construction, or in advanced planning around the world. The global proliferation of these facilities raises questions about how to ensure their safe and secure operations while enhancing their contributions to science and global health. One solution is to establish an MCL network that enables the sharing of best practices, collaboration, transparency, and exchange of specimens and technology.

A multitude of challenges are associated with MCLs. Even at the idea stage, a serious issue is the objection of local communities to the construction of an MCL in their neighborhood. Several MCL operations were delayed or never realized because of public concern. Gaining community trust and support is therefore vital to planning and operating MCLs, so a network of such labs would be valuable for sharing experiences and providing guidance in these situations.

Besides the millions of dollars that it costs to build a modern MCL, there are annual operations—maintenance, utility, and security—that can amount to 5 to 10% of the construction costs. Moreover, there is a need for experienced guidance and qualified oversight to ensure that an MCL is built and operated safely and securely. Yet, few such resources exist, and available training opportunities are inconsistent and often costly. An MCL network could fill the personnel pipeline more efficiently by connecting experienced personnel and professional societies to develop standards for globally accepted training and create mentoring opportunities.

A training session in Galveston National Laboratory

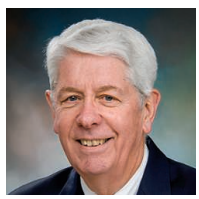
"These labs handle the world's most dangerous pathogens..."

Importantly, MCLs must share a culture of responsibility. These labs handle the world's most dangerous pathogens known, and there must be safeguards to prevent theft or misuse. At the same time, security must be balanced against mechanisms that support collaboration, including specimen sharing. Again, by working together through an MCL network to develop standards and guidelines, a culture of responsibility could be fortified.

We direct a newly constructed MCL in Wuhan, China (Z.Y.), and an established MCL in the United States (J.W.L.), in Galveston, Texas. In preparation for the opening of the new China MCL, we engaged in short- and long-term personnel exchanges focused on biosafety training, building operations and maintenance, and collaborative scientific investigations in biocontainment. We succeeded in transferring proven best practices to the new Wuhan facility. Both labs recently signed formal cooperative agreements that will streamline future scientific and operational collaborations on dangerous pathogens, although funding for research and the logistics of exchanging specimens are challenges that we have yet to solve. Ours is a promising first step in MCL partnerships; however, wider national, regional, and international cooperation is needed. We benefited from meetings jointly sponsored by the U.S. National Academy of Sciences and the Chinese National Academy of Sciences, and from World Health Organization initiatives, but stakeholders are not limited to human and animal health. Our partnership still requires input from foundations and governmental agencies that are involved in security, commerce, and transportation, as well as from the commercial sector.

Not every country requires an MCL, but every country can benefit from the collaborative operation of these labs. We encourage existing MCLs to convene a forum that brings together all stakeholders to conceive of an MCL network so that these critical labs can tackle urgent global health needs safely, securely, and productively.

–James W. Le Duc and Zhiming Yuan



James W. Le Duc
is the director of the Galveston National Laboratory and a professor in the Department of Microbiology and Immunology at the University of Texas Medical Branch, Galveston, Texas, USA. jwleduc@utmb.edu

Zhiming Yuan
is the director of the Wuhan National Biosafety Laboratory of the Chinese Academy of Sciences and a professor at the Wuhan Institute of Virology, Wuhan, China. yzm@wh.iov.cn

10.1126/science.aav7120

5000 The number of faces an average person can recognize, according to an October study in the *Proceedings of the Royal Society B*.

IN BRIEF

Edited by **Catherine Matacic**

METEOROLOGY

Michael's power blindsided forecasters

Hurricane Michael tore houses from their foundations in Mexico Beach, Florida.

Hurricane Michael roared into Mexico Beach, Florida, on 10 October as the strongest storm ever to strike the Florida Panhandle in terms of wind speed, and the third strongest to make landfall in the continental United States. In just 24 hours, the storm—which killed at least 18—jumped from a Category 2 to just shy of a Category 5, the upper limit of the hurricane rating scale. On its 5-day march to the coast, Michael underwent at least three periods of "rapid intensification" that took researchers by surprise. Forecasting a hurricane's path is relatively straightforward, but when it comes to predicting changes in intensity, the underlying physics becomes much more complicated, says Kerry Emanuel, an atmospheric scientist at the Massachusetts Institute of Technology in Cambridge. That's because hurricanes are complex, massive rotating heat engines, fueled by a favorable combination of warm ocean water, moist air, and consistent atmospheric winds. Collecting enough data points is also incredibly difficult. Researchers are concerned that by warming the oceans, climate change will open the door to hurricanes of greater intensity and unpredictability.

Hunger strike fells Indian activist

ENVIRONMENT | One of India's leading environmentalists—Guru Das Agrawal, also known as Swami Gyan Swaroop Sanand—died on 11 October following a 111-day fast to save the Ganges River from sand mining, excessive damming, and pollution. Agrawal, a former professor of environmental engineering at the Indian Institute of Technology in Kanpur, became a Hindu ascetic in 2011, dedicating himself to rejuvenating the dying river. Agrawal commenced his Mahatma Gandhi–style hunger strike on 22 June to force India's government to honor a commitment to save the Ganges. When his letters to Prime Minister Narendra Modi went unanswered, Agrawal also stopped drinking water; police took him to a hospital for force-feeding on 10 October, but he died the next day.

Puerto Rico's insects perish

CLIMATE | Consider it another effect of rising global temperatures: Over the past 4 decades, Puerto Rico's insect and other arthropod populations have plummeted, some declining by 90% to 98%, according to a 15 October study in the *Proceedings of the National Academy of Sciences*. Scientists surveyed the island's Luquillo Experimental Forest from 1976–77 and again from 2011–13. They found that, along with arthropods, their predators—including many lizards, frogs, and birds—have seen their numbers nosedive. The underlying reason, they write, is climate change. As maximum daily temperatures in the forest have risen 2°C, arthropods that evolved in the relatively constant tropical temperatures have struggled to cope.

Gene bank gets seed money

PHILANTHROPY | Many research storehouses of seeds and other plant tissues are in poor condition because of funding shortages. The Crop Trust, a nonprofit in Bonn, Germany, has aimed to change that. Last week, it announced its first long-term award: a 5-year, perpetually renewable grant of $1.4 million annually to the gene bank of the International Rice Research Institute (IRRI) in Los Baños, Philippines.

PHOTO: JABIN BOTSFORD/*THE WASHINGTON POST* VIA GETTY IMAGES

Published by AAAS

CLIMATE

Sea-level rise threatens heritage sites

Homes, roads, and ports aren't the only things threatened by sea-level rise: A new study suggests some 80% of World Heritage sites along the Mediterranean coasts are also at risk. These include the medieval city of Rhodes in Greece, the Kasbah of Algiers in Algeria, and Venice, Italy, and its saltwater lagoon (above). Scientists found that by the year 2100, 40 of 49 such low-lying sites would be endangered by storm surges that exceed the level of a modern-day 100-year flood, and 46 would be threatened by coastal erosion. Nations seeking to protect such sites shouldn't wait 80 years to address these problems, researchers warn in the 16 October issue of *Nature Communications*. That's because 37 of the flood-prone sites and 42 of the erosion-prone spots are already vulnerable, although to a much smaller degree than they will be in the future.

Since 2004, the Crop Trust has raised a $300 million endowment for plant gene banks. Over the next few years, it also plans to permanently fund the 10 gene banks at IRRI's sister institutes in the Consultative Group on International Agricultural Research, a global nonprofit partnership for agriculture.

Global C-sections double

PUBLIC HEALTH | In 15 years, the number of cesarean section (C-section) births around the globe has nearly doubled, going from 16 million—or 12% of live births—in 2000 to 29.7 million—or 21% of live births—in 2015, according to an 11 October study in *The Lancet*. That's well above a commonly used estimate that C-sections are medically necessary in just 10% to 15% of births. Regional disparities are vast: Latin America and the Caribbean have rates of 44.3%, whereas in West and Central Africa, the procedure is used in just 4.1% of births.

Stem cell paper retraction

PUBLISHING | Marking another chapter in one of the decade's most infamous scientific fraud cases, Harvard Medical School and Brigham and Women's Hospital, both in Boston, are recommending the retraction of 31 papers by a former high-profile researcher there. The institutions said in a statement that the papers "included falsified and/or fabricated data," and that they've notified the relevant journals. Cardiologist Piero Anversa had claimed to have found stem cells in the heart that could regenerate cardiac muscle, but in 2014 Harvard revealed it was investigating Anversa after one of his papers was retracted. His lab shuttered the following year, and he no longer works at the hospital. The investigating institutions haven't shared the list of affected papers.

Rocket mishap endangers ISS

SPACE SCIENCE | A booster failure last week on a Russian Soyuz rocket sent the two astronauts aboard into a dangerous emergency landing. They survived, but the 11 October mishap puts the International Space Station (ISS) in a precarious situation. That's because Soyuz is the only launch system that can deliver crews to the ISS after the 2011 retirement of the U.S. Space Shuttle fleet. While it investigates, Russia has suspended Soyuz launches. The Soyuz capsule now docked at the ISS—which is due to return to Earth by January 2019—recently had a small hole in its side repaired. If NASA and Russia decide to use it to bring the ISS's three crew members down, the station will be unattended for the first time since 2000. New crewed launch systems being built for NASA by Boeing and SpaceX won't be ready until mid-2019 at the earliest.

Filipinos chill to vaccines

PUBLIC HEALTH | Reeling from calamitous rates of dengue, the Philippines launched a public vaccination campaign against the mosquito-borne disease in spring 2016, using the Sanofi Pasteur vaccine Dengvaxia. Now, public confidence in all vaccines there has plunged, after the drugmaker reported in November 2017 that Dengvaxia should be given only to people previously infected. The public was outraged, even though most of the 830,000 children who had been vaccinated by that time had probably been infected. On 12 October, scientists with the London School of Hygiene & Tropical Medicine's Vaccine Confidence Project reported in *Human Vaccines & Immunotherapeutics* that after the incident, just 32% of Filipinos "strongly agreed" that vaccines were important, compared with 93% in 2015.

U.S. gets new weather model

METEOROLOGY | In early October, the National Weather Service's National Centers for Environmental Prediction approved its next-generation global weather model for broad use; starting in January 2019, its computer code will power weather forecasts across the United States and elsewhere. The model, meant to put the agency back on equal footing with sophisticated European competitors, is

PHOTO: ISTOCK.COM/GIVAGA

powered by a mathematical engine called FV3, first developed to recreate Earth's atmospheric flows for climate forecasts. The new version, the agency notes, did especially well during a summer comparison with its predecessor at forecasting hurricane paths and fluctuations in air pressure, among other improvements.

Gift aids MIT computer science

PHILANTHROPY | Fueled by a $350 million donation for a computer science building from investment banker Stephen Schwarzman, the Massachusetts Institute of Technology (MIT) in Cambridge plans to "rewire" how it educates students in this foundational subject. The gift brings MIT two-thirds of the way toward its goal of spending $1 billion on research and training in computer science and artificial intelligence to keep up with rising demand for computer science courses. Computing is now being taught "in a haphazard fashion" across many departments, says computer scientist Michael Stonebraker, one of seven MIT faculty members who last year proposed the creation of a separate school of computing. MIT Provost Martin Schmidt says the new building and the hiring of 50 new faculty—half for the new college and the rest across campus—should help address that problem. Schwarzman's gift is part of a $5 billion capital drive MIT launched in May 2016, for which $4.3 billion has been pledged.

Delay for asteroid probe

SPACE SCIENCE | Japan's Hayabusa2 spacecraft, now orbiting the asteroid Ryugu, will be sent to sample the surface in January 2019 instead of late October, the Japanese Aerospace Exploration Agency (JAXA) announced on 14 October. The probe's arrival at the asteroid, followed by the deployment of three small landers, has revealed the top-shaped Ryugu to be far rockier than thought, with few smooth, flat patches big enough for the refrigerator-size probe to touch the surface with its arm and suck in grist. JAXA will use the extra few months to verify the spacecraft's ability to sample plains just 20 meters wide, rather than 100 meters, as originally planned. Return of the spacecraft's samples to Earth is scheduled for 2020.

EPA drops clean air panels

CLIMATE | To the dismay of many environmental researchers, the U.S. Environmental Protection Agency (EPA) in Washington, D.C., last week abolished a panel of outside experts that helped review permissible levels of deadly particulate pollution. It also dropped plans to form a similar panel to offer technical advice on standards for ground-level ozone, a major contributor to smog. The two specialized panels weren't needed, said Andrew Wheeler, EPA's acting chief, because they duplicated the work of their parent panel, the Clean Air Scientific Advisory Committee. That seven-member committee, made up mostly of Wheeler's appointees, will now review air-quality standards. Critics say the moves are designed to reduce independent oversight of several upcoming EPA reviews of major air pollution regulations, and to ensure outcomes that are favorable to industry.

AWARDS

Science magazine's Elizabeth Culotta, Ann Gibbons, Emily Underwood, Jennifer Couzin-Frankel, and John Bohannon won the 2018 National Academies Communication Award for a May 2017 package of articles on human migration; Beth Rakouskas won Art Director of the Year at the 2018 Folio Awards.

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

CONSERVATION

Whooping crane breeding program winds down

This week, 33 whooping cranes were airlifted from Maryland to Louisiana, marking the beginning of the end of the U.S. Geological Survey's (USGS's) 50-year effort to help save these endangered birds. Scientists at USGS's Patuxent Wildlife Research Center in Laurel, Maryland, spearheaded the recovery of *Grus americana*, whose numbers had dropped to fewer than 20 in the wild. To bring back the majestic 1.3-meter-tall birds, biologists developed innovative methods, including using puppets in the shape of crane heads to teach chicks to feed and to follow ultralight aircraft on migratory flights. Their numbers now stand at more than 700, and the center's breeding and caretaking protocol are now standard. "We feel as if our job is done," says the center's director, John French. The airlifted birds—and some remaining to be transferred—will now become part of captive breeding flocks.

PHOTO: JONATHAN L. FIELY/USGS PATUXENT WILDLIFE RESEARCH CENTER

Engineer Chrissy Houlahan is a heavy favorite to win a House of Representatives seat in Pennsylvania.

PHOTO: CHRISSY HOULAHAN FOR CONGRESS

U.S. ELECTION

Scientists who beat the odds seek victory in November

Candidates, all Democrats, learned in the primaries that a science background by itself does not sway voters

By **Jeffrey Mervis**

The wave of scientists running this year for U.S. Congress has crashed on the shoals of political reality. But it's not a total wipeout.

Although nearly two-thirds of the candidates—all Democrats—were knocked out in primary contests, more than a dozen will appear on next month's general election ballots. Their fate could help determine whether Democrats recapture the House of Representatives and become a counterweight to the policies of President Donald Trump.

Trump's victory in November 2016 energized many with scientific, technical, and health care backgrounds to dive into politics for the first time. They promised to rely on research and evidence in making policy, something they say Trump has failed to do. And whereas most novice candidates begin at the local or county level, these scientists set their sights high: a seat in the 435-member House.

Just how many scientists chose to run is open to interpretation. There is no common definition of a scientist, but *Science* identified 40 first-time House candidates with a graduate degree in a science, technology, engineering, or math (STEM) discipline, or a medical degree. The pool grows to 47 if expanded to include candidates without advanced degrees but endorsed by 314 Action, a Philadelphia, Pennsylvania–based political action committee formed in 2016 to help people with STEM and health backgrounds run for office. All are Democrats; *Science* could not find any first-time Republican candidates who have cited STEM backgrounds as an asset.

Overall, 17 of the candidates survived the primaries and advanced to the general election on 6 November (see graphic, p. 272). Only one of the nine candidates from an academic or research setting—and just two of 14 candidates with a doctoral degree in the natural sciences—remain on the ballot. Physicians fared better: Six of nine candidates with M.D.s are still in the running.

As the election nears, analysts say two STEM candidates are favored to win their races, and six more have a credible chance. The rest face uphill battles running in solidly Republican districts or against popular incumbents.

The science candidates span the ideological gamut from fiery progressive to quiet conservative. Their campaign strategies also vary. Those running in districts that Trump carried in 2016 tend to focus on the perceived weaknesses of their Republican opponents and avoid criticizing the president. But those in more Democratic-leaning districts highlight their differences with the president on health care, trade, taxes, and immigration.

Despite the widespread sentiment among Democrats that Congress needs more members with scientific training, the primary election results suggest a STEM background does not, by itself, sway voters. What does matter are high name recognition, strong ties to party regulars, and access to a large pool of donors and cash.

Those factors have helped lift Chrissy Houlahan to the brink of victory in Pennsylvania's sixth congressional district. The industrial engineer, who attended Stanford University in Palo Alto, California, has a master's degree in technology policy from the Massachusetts Institute of Technology in Cambridge. She spent 3 years in the U.S. Air Force working on defense technologies, followed by executive positions at two successful startup companies and a year of teaching high school chemistry.

Houlahan's impressive fundraising helped scare off primary challengers. Now, a weak Republican opponent has solidified her status as the front-runner for the open seat in a district outside Philadelphia where Democrats hold a comfortable majority.

Two other STEM candidates are running in districts that are rated toss-ups—and have attracted attention as bellwethers of the national mood. In Illinois, energy entrepreneur Sean Casten is trying to knock off incumbent Peter Roskam (R) in a race that has drawn millions of dollars from donors outside the sixth congressional district in the Chicago suburbs. And a victory by pediatrician Kim Schrier, running against real estate investor Dino Rossi (R) for an open seat in Washington's eighth district east of Seattle, could help change the demographics of an elite club—doctors who are also members of Congress. There are no women among the 16 current physician-lawmakers, and only two Democrats.

Four science candidates are given reasonable odds despite running against incumbents in districts that traditionally favor Republicans. That list includes pub-

lic health advocate Lauren Underwood, in Illinois's 14th congressional district west of Chicago, and resource economist Kathleen Williams, running for Montana's sole House seat. Both women hope to convince voters that it's time for a change. But whereas Underwood is swimming with the national Democrat party on key issues, in particular health care, Williams is trying to distance herself from the party's leadership and stressing the need for greater bipartisan amity in Congress.

"I had given my 20s to make sure that the American people had a high-quality health care system, and I was not interested in watching the Trump administration come in and try to dismantle that," says Underwood, a registered nurse with a master's degree in public health policy who worked on implementing the Affordable Care Act and on biopreparedness during former President Barack Obama's administration. She returned to Naperville, Illinois, and decided to run against her congressman, four-term Representative Randy Hultgren (R), after she felt he had broken a promise not to support legislation making it easier for health insurers to deny coverage to people with preexisting conditions.

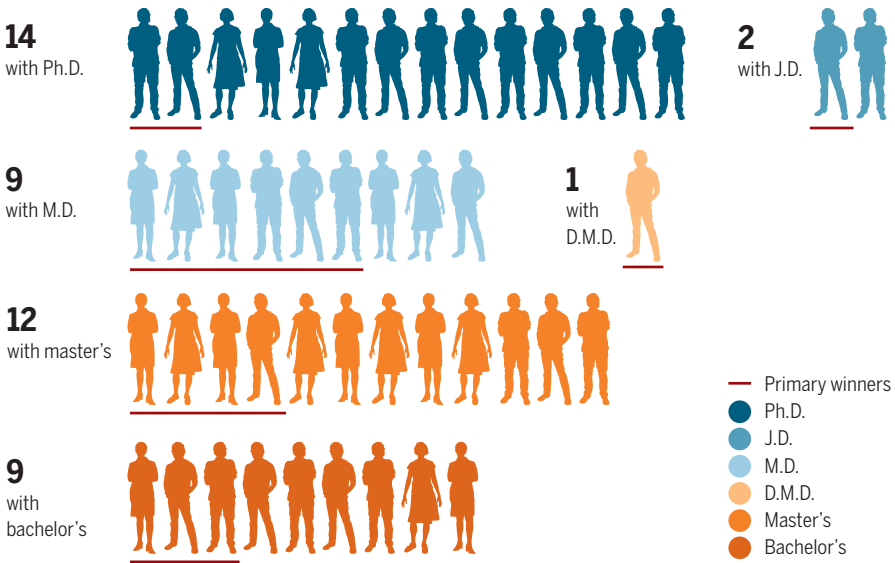
Underwood's stances on health care, tax policy, immigration, environmental protection, and reproductive rights reflect her self-described "progressive" views. Hultgren is equally proud of his record as a fiscal and social conservative. But the contrast blurs when it comes to science. Underwood supports greater federal spending on science, and Hultgren has been named "a champion of science" by research groups because of his vocal support for increased federal funding for basic energy research, reflecting the proximity of his district to two Department of Energy national laboratories. But whereas Hultgren backs Trump's decision to pull out of the Paris climate treaty and repeal Obama's Clean Power Plan, Underwood sharply criticizes those positions.

In Montana, Williams is promising voters in the state's single, state-wide district that they will set her agenda, not national Democratic party leaders. (She's part of a small group of candidates who say they won't vote for Nancy Pelosi [D–CA] as speaker if the Democrats win the House majority.) "Although I have degrees in the sciences [including a master's degree in recreation resources], and water is important to Montana, it's health care that is their priority," she says. "So that's what I'm running on."

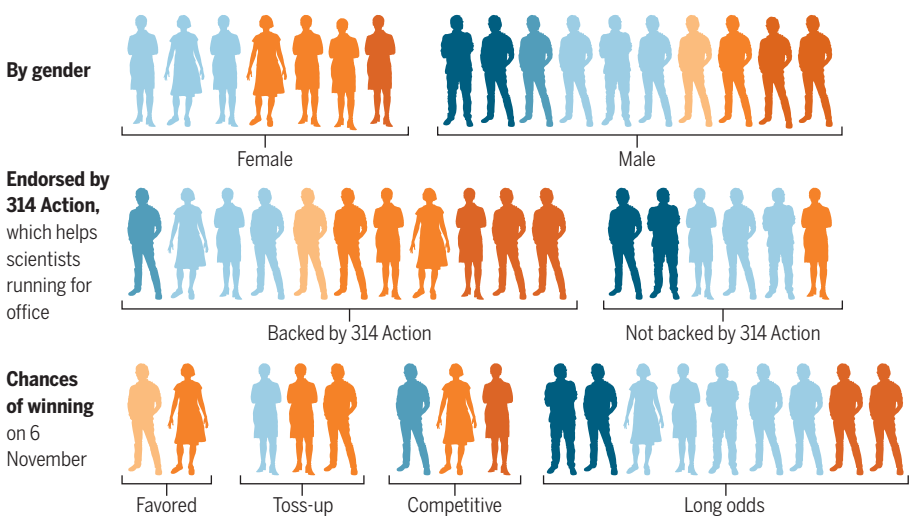
Williams is one of the few STEM candidates with electoral experience, having served three terms in the state legislature.

The hard road to Congress

Some 47 candidates with scientific training, all Democrats, decided to run for a seat in the 435-member U.S. House of Representatives. Most were political novices, and 30 didn't survive primary elections held earlier this year in every state. Based on their highest degree, the group included:



Here's a closer look at the 17 who won their primary contests:



Her Republican opponent, Representative Greg Gianforte, is a successful software entrepreneur who is defending a seat he won last year in a special election. And although Williams accuses Gianforte of being out of touch with local voters, she has generally avoided criticizing him for his support of Trump's major policy initiatives.

The two surviving STEM candidates who have doctorates are considered long shots. In Mississippi, Randy Wadkins has kept his day job as a professor of biochemistry at the University of Mississippi in Oxford while running against Trent Kelly, the Republican incumbent in the state's first district. Trump won the district by a two-to-one margin over Hillary Clinton, and the national Democratic party has largely ignored Wadkins, much to his frustration.

Wadkins says his district contains the same coalition of voters who propelled Democrat Doug Jones to an upset win of a Senate seat in neighboring Alabama earlier this year. But a lack of visibility and his struggle to raise money have made it hard for him to get out his message.

The other remaining STEM candidate with a Ph.D.—Indiana's Mel Hall—doesn't mind being an afterthought to the national party. He also has pledged to oppose Pelosi, although he reminds voters, "I am a Democrat." A former Methodist minister, Hall calls himself "a relentless pragmatist who's driven by data." It's a nod to his decision

CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) CAMPAIGN WEBSITES/314 ACTION/COOK POLITICAL REPORT

to earn a Ph.D. in data science and join a local company that, under his leadership, became the dominant player in patient satisfaction surveys.

Hall's success as a health care executive has allowed him to pour nearly $1 million into his campaign to win Indiana's second district, and keep pace with spending by his Republican opponent, three-term incumbent Jackie Walorski. But Walorski carried the solidly Republican district by 22 percentage points in 2016.

In two other races, science watchers are paying attention not because of the challenger's background, but because the Republican incumbent is a major figure in science policy. In the seventh district west of Houston, Texas, Representative John Culberson, who leads the House spending panel that oversees NASA and the National Science Foundation (NSF), is facing a tough challenge from Democrat Lizzie Fletcher, a local attorney. (A STEM candidate, cancer researcher Jason Westin, had set out to challenge Culberson but finished third in the Democratic primary.) Culberson, who was first elected in 2000, is seen as vulnerable on a range of local issues, including flood control.

In the 48th congressional district in Orange County in California, another longtime Republican incumbent active on national science issues, Representative Dana Rohrabacher, is running neck-and-neck with Democratic challenger Harley Rouda. In 2013 Rohrabacher made an unsuccessful bid to chair the House science committee, and he'd likely try again if Republicans retain control of the House.

The man who beat Rohrabacher was Republican Representative Lamar Smith, a 15-term incumbent who represents the 21st district in central Texas. Over the past 6 years, Smith has been a persistent critic of NSF and vocal opponent of federal investments in climate and social science. And last spring, Joseph Kopser, a 20-year U.S. Army veteran with a degree in aerospace engineering and experience as a clean energy entrepreneur, decided to take on Smith. But 6 months later Smith announced he would retire. So Kopser—who bested another STEM candidate, math professor–turned–minister Mary Wilson, in a primary—now faces Chip Roy, a former chief of staff to Senator Ted Cruz (R–TX). Politicos say it is Roy's race to lose.

STEM candidates are running in eight of the 73 districts that analysts expect to be key to which party gains a House majority. So even absent a science wave, how they fare on election night will shape the makeup of the new Congress that convenes in January 2019. ■

SCIENCE POLICY

Scientists, environmentalists brace for Brazil's right turn

Country's likely next president promises R&D investment but favors development over climate and biodiversity

By **Herton Escobar**

Beset by economic woes and dissatisfied with the left-wing politicians in power for most of the past 15 years, Brazil appears poised to make a hard turn and elect a far-right candidate, Jair Bolsonaro, as its next president. His rapid ascent has unnerved local researchers, who worry about the future of Brazilian science, the protection of the country's biodiversity, and its role in the global struggle against climate change.

"I think we are headed for a very dark period in the history of Brazil," says Paulo Artaxo, a climate change researcher at the University of São Paulo (USP) in São Paulo, Brazil. "There is no point sugarcoating it. Bolsonaro is the worst thing that could happen for the environment."

Bolsonaro has vowed to withdraw Brazil from the 2015 Paris agreement, which requires nations to reduce greenhouse emissions to combat climate change, and he plans to eliminate the Ministry of the Environment and fold its duties into the Ministry of Agriculture, Livestock, and Supply. The "Myth"—as his supporters call him—has also said, while campaigning in the Amazon, that Brazil has "too many protected areas" that "stand in the way of development."

The 63-year-old congressman and former army captain fell just short of winning a majority in the primary election earlier this month, and he heads into a 28 October runoff with a large lead in the polls over left-wing scholar Fernando Haddad. The plight of Brazil's research establishment, which has endured sharp budget cuts in recent years, has had little mention in the campaigning so far. When recently asked about his possible choice for science minister, Bolsonaro named Brazilian astronaut and former air force pilot Marcos Pontes, a member of his party, as his top preference.

A draft campaign document focusing on science—first revealed last week by the Brazilian newspaper *O Estado de S. Paulo*—offers additional insight into his plans. It pledges to more than double the level of R&D investment in the next 4 years, but would focus most of the extra money and attention on applied sciences such as space and robotics, rather than on basic research at universities.

Under the slogan "Brazil above everything, God above all," Bolsonaro's campaign exalts national pride, military discipline, and a zero-tolerance, iron-fist stance against

Jair Bolsonaro leads in the polls for Brazil's presidential election.

PHOTO: ANDRE COELHO/BLOOMBERG/GETTY IMAGES

crime. Famous for inflammatory remarks about women and minorities, Bolsonaro openly cherishes the 21-year military dictatorship that started with a coup in 1964.

Historically, Bolsonaro has had little to do with science, and he recently sparred with the academic community, authoring legislation to favor an unproven cancer therapy. A general he picked to craft his science and education plans defended the teaching of creationism this week, telling *O Estado de S. Paulo* that students need to know that "Darwin existed," but not necessarily to "agree with him."

Climate change is one scientific issue Bolsonaro has touched on. He hasn't specifically questioned that humans are driving global warming, but his son, a popular Brazilian congressman, has done so in a video that celebrates the climate policies of U.S. President Donald Trump. And Bolsonaro has indicated that the Paris agreement's mandate threatens Brazil's national sovereignty, especially in the Amazon region, where deforestation for farming and cattle ranching has driven most of the nation's greenhouse gas emissions.

> ***"I think we are headed for a very dark period in the history of Brazil."***
>
> **Paulo Artaxo**, University of São Paulo

Through law enforcement and mechanisms such as incentives for sustainable practices, Brazilian authorities have substantially reduced Amazon deforestation in the past 13 years, and the nation's commitment to the Paris climate change accord requires it to continue that trend. Bolsonaro's campaign instead promises to promote agriculture and mining in the region. One of the generals helping develop the candidate's policies told *O Estado de S. Paulo* last week that he missed the days when road builders could cut down trees in the Amazon without being bothered by environmental authorities.

Unfettered development of the Amazon would be a "grave mistake," says Eduardo Assad, a climate change and agricultural scientist at the Brazilian Agricultural Research Corporation in Campinas. He adds that studies show Brazil's agricultural production could be doubled by exploiting abandoned or degraded pastures and farmland—"without any additional deforestation."

Haddad, a 55-year-old professor of political science at USP, offers more moderate views, focusing on social justice and sustainable development. But corruption scandals that culminated in the impeachment of Brazil's then-President Dilma Rousseff in August 2016 and the recent imprisonment of former President Luiz Inácio Lula da Silva, who is closely associated with Haddad, have darkened his prospects. In a poll out on 15 October, he trailed Bolsonaro 41% to 59%.

Haddad's campaign has pledged to "rebuild the national science, technology, and innovation system" and provide ample public funding to help double the intensity of the country's R&D expenditure to 2% of gross domestic product (GDP) by the year 2030. In his campaign's draft science document, Bolsonaro pledges even more R&D investment, 2.5% of GDP by the end of his term, in 2022.

Many researchers doubt that either candidate can fulfill such pledges. "I've heard this promise many times before," says Fernando Peregrino, a science policy expert and president of Confies in Brasília, a national network of foundations that support scientific research and higher education. Brazil lacks the economic policies and fiscal stability to provide generous support for R&D, he believes.

Bolsonaro plans to rely heavily on the private sector to boost R&D spending, through economic incentives and partnerships. "Our greatest deficit is in innovation," says economist Marcos Cintra, president of the Brazilian Research and Innovation Agency in Rio de Janeiro, who is helping craft Bolsonaro's R&D proposals.

As for public spending, the campaign document calls for a "greater balance" between "curiosity-oriented research and research directed towards missions and goals." Brazil's Ministry of Education now receives 60% of federal R&D funds, compared with the Ministry of Defense's 1.5%, and Cintra argues defense should get more. "One of the political difficulties is that public research in Brazil still has a strong academic bias, without focus or specific priorities," Bolsonaro's campaign document says.

Luiz Davidovich, president of the Brazilian Academy of Sciences in Rio de Janeiro, agrees that it's important to define national priorities and strategic goals, but says academic and intellectual freedom must also be preserved.

Whoever wins the election, scientists here are unlikely to see any relief soon. Federal funding for the Ministry of Science, Technology, Innovation, and Communication has fallen by more than half since 2013, and the budget proposal for 2019—drafted by the current administration—predicts another 10% cut below this year's.

"Even for the most optimistic of us, it's looking bad," Artaxo says. ■

Herton Escobar is a science journalist in São Paulo, Brazil.

ASTRONOMY

Planet hunter nears its end

Kepler space telescope found trove of exoplanets

By **Daniel Clery**

Kepler, the NASA space telescope that almost single-handedly transformed the study of exoplanets, may have sent its last data home. Managers halted the craft's latest observing campaign last month, when its pointing became unstable—a sign of low fuel supplies. To conserve fuel, they put it in "nap mode" until a scheduled high-speed download opportunity began on 10 October. Observations are still on hold as the mission team sifts the spacecraft's vital signs for hints of the remaining fuel supply, but the end could come suddenly and without warning, says project scientist Jessie Dotson of NASA's Ames Research Center at Moffett Federal Airfield in California. "The only way to tell is to watch the spacecraft," she says. "We'll know when we know."

When the end comes, it will be a bittersweet conclusion to a prolific mission that upended theories. Before Kepler, astronomers had found only a few hundred exoplanets, mostly gas giants uncomfortably close to their parent stars. Kepler found thousands, of many sizes—including 30 tantalizingly similar to Earth. Kepler, which pioneered a novel approach to planet finding, was a gamble, says Bruce Macintosh of Stanford University in Palo Alto, California. But, "The universe cooperated, beyond anyone's wildest dreams."

Kepler's success is a vindication for William Borucki, the NASA space scientist who struggled to persuade the agency to build it. Early exoplanet discoveries came from observations of stars wobbling from the tug of large planets. Borucki suggested planets would also reveal themselves when they passed in front their host star, blocking a tiny fraction of its light. He just needed a photometer 1000 times more sensitive than any available. "You can do an awful lot with dedicated people and technology if you persist," says Borucki, now retired from NASA Ames.

Persist he did. Borucki's proposals were rejected four times before NASA finally approved the mission in 2001. "He really had a long fight," says Didier Queloz, an astronomer at the University of Cambridge in the United Kingdom who in 1995 helped dis-

cover the first exoplanet around a normal star, using the wobble technique.

The $692 million Kepler finally launched in 2009, with a 1.4-meter mirror and a 95-megapixel camera, the largest in space at the time. It stared at the same patch of sky, about the size of a fist at arm's length, simultaneously monitoring the brightness of 150,000 stars. To date, it has captured about 2650 confirmed exoplanets, and nearly 3000 more await confirmation. Among the haul are 30 rocky planets, all less than twice the size of Earth, orbiting in habitable zones where liquid water could exist. Assuming Kepler's star field is typical—and allowing for the fact that it detects only the small fraction of planets that transit—astronomers estimate the Milky Way harbors at least as many planets as stars: 100 billion.

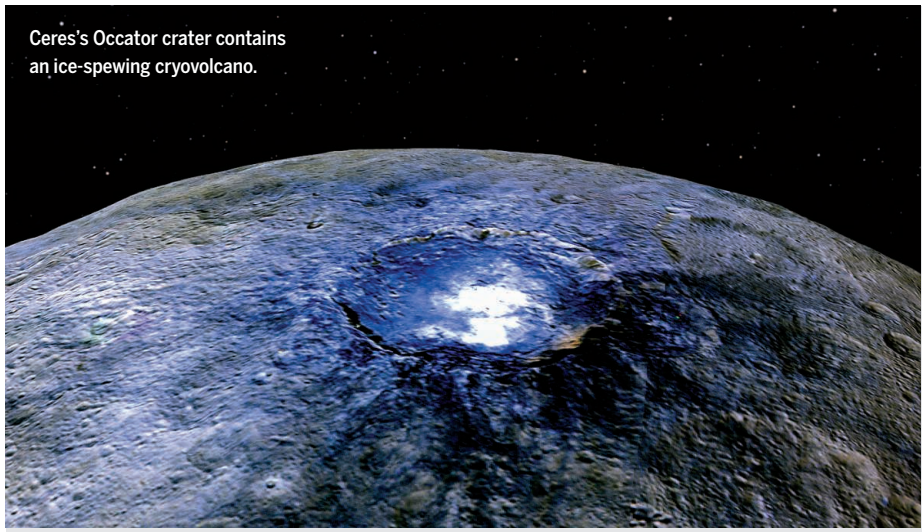
The diversity of planets was a surprise, however. Multiplanet systems proved to be common, but they often looked wildly unlike our solar system in the sizes and orbits of their planets (*Science*, 29 July 2016, p. 438). "Theorists haven't caught up yet," Macintosh says. For instance, the most common planet type in the sample is a variety unknown in our solar system, halfway between Earth and Neptune in size.

The planets came so fast and furiously that, a few years into the mission, the Kepler team ended a data policy that gave mission scientists exclusive access. It began to release processed data after a few months and, last year, it began to release raw data immediately. "It made it easy for people to enter the field," Macintosh says.

There were hiccups, however. Operators found that both the camera's output and the stars themselves were more variable than predicted, making it harder to spot the tiny light dips of true Earth twins. By 2013, two of the four gyros used for pointing had failed, and the drifting spacecraft had to abandon its fixed view of the stars. Mission controllers then took advantage of sunlight pressure to partially steady the craft, but the rate of discovery slowed. As a result, the mission had to abandon a primary goal: assessing the frequency of Earth twins around sunlike stars. But given the wealth of data that hardly seems to matter. "People don't really care," Queloz says.

Successors are ready to pick up the torch. The Transiting Exoplanet Survey Satellite, launched in April to look for nearby exoplanets (*Science*, 30 March, p. 1453), has already found two, with one more awaiting confirmation. And the European Space Agency has two missions in the works: the Characterising Exoplanets Satellite, due for launch in 2019, and Planetary Transits and Oscillations of Stars in 2026. "It's an OK time to say goodbye," Macintosh says. ■

PHOTO: NASA/JPL-CALTECH/UCLA/MPS/DLR/IDA



Ceres's Occator crater contains an ice-spewing cryovolcano.

NASA's asteroid explorer Dawn soon to go dark

By **Paul Voosen**

After an 11-year journey to Vesta and Ceres, the asteroid belt's two largest members, NASA's Dawn spacecraft is expected to run out of thruster fuel in the next few weeks, ending its mission. The robotic explorer, which gave a close-up view of how the presence or absence of water can shape asteroids, will remain tumbling in orbit around Ceres for decades before ultimately crashing into it.

Launched in 2007, Dawn is the only NASA mission to orbit two planetary bodies, a feat made possible by its efficient ion thrusters. In 2011, it arrived at the egg-shaped, 600-kilometer-long Vesta, orbiting for a year before departing for Ceres, where it arrived in 2015.

The two asteroids, which together account for 45% of the belt's mass, turned out to be a tale of contrasts. Parched Vesta has a composition like the terrestrial planets, with an iron core and a dry, rocky surface carved up into canyons, craters, and mountains, remnants of past impacts and volcanism. Dawn was able to verify that a class of meteorites found on Earth are chips off of Vesta, making it a sort of "reverse sample return mission," says Carol Raymond, the mission's principal investigator and a planetary scientist at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California.

Telescopes had already found water-rich minerals on Ceres, a 900-kilometer-wide body classified as a dwarf planet because of its large size and spherical shape. Dawn revealed the remnants of a frozen ocean topped by a heavily cratered crust of clays and salts. "We could not have imagined it would have looked like this," Raymond says.

The young sun could have boiled away much of the water if Ceres had formed in its present location. So, some scientists suspect it was born out past Jupiter and Saturn, only to be drawn in by the gas giants' turbulent early gyrations. Dawn found a marker of this distant birthplace: ammonia, a volatile molecule that could have only formed in the outer solar system.

Dawn also spotted a sign that Ceres remains geologically active: what looked like a lone volcano, 4 kilometers tall, that oozes a "lava" of water, salt, and other materials. Later observations spotted traces of 21 more volcanoes. And more than 100 bright spots rich in salt also suggested the subsurface ice finds its way to the surface here and there, as a slushy brine. One such spot, at Ernutet crater, showed signs of organic molecules—common in the outer solar system and not a signifier of life. But the discoveries have made Ceres a prime candidate for further exploration; a NASA study on a possible lander is expected to conclude next year.

In its final months, Dawn settled into a tight orbit, just 35 kilometers from Ceres's surface. The defunct craft could remain in orbit for a half-century or more, says Marc Rayman, Dawn's mission director at JPL, "an inert celestial monument around the dwarf planet it helped unveil."

The Shanghai Synchrotron Radiation Facility draws users from industry.

FUNDING

China narrows U.S. lead in R&D spending

But lagging support for basic research may hinder ambitions to become a science superpower

By **Dennis Normile**

Already the world's second biggest R&D spender behind the United States, China is steadily narrowing the gap, according to a government report released last week. It shows that the total R&D outlay by government and industry rose 12.3% last year to a record 1.76 trillion yuan ($254 billion). Whereas China spent about 34% as much as the United States in 2012, the proportion is now closer to 45%, figures from the Organisation for Economic Co-operation and Development (OECD) and U.S. agencies show. In purchasing power, China's 2016 spending was equivalent to 88% of U.S. spending, according to OECD.

"The year-to-year growth in R&D spending indicates firm governmental and social support for making China a scientific power," says Xie Xuemei, a specialist in innovation economics at Shanghai University in China. She adds, however, that "there is still a long way to go" for China to match the research capabilities of developed countries. Xie and others note that basic research funding appears to be lagging, at just 5.5% of the total, according to the report, from the ministries of science and finance and the National Bureau of Statistics in Beijing. Observers also question how the burgeoning funds are spent. "It's one thing to get the money, it's another to get the talent and to spend the money effectively," says David Zweig, a political scientist at the Hong Kong University of Science and Technology in China who studies international talent flow.

12.3%
Increase in China's R&D spending in 2017

5.5%
Share of spending on basic research

Accounting differences between China and other countries make international comparisons tricky, says Cao Cong, a science policy specialist at the University of Nottingham's Ningbo, China campus. Government accounting lumps R&D funding with other science and technology expenditures such as support for science communications, administration, and exchanges. On the other hand, when figuring spending on basic research, China often excludes capital investment in major facilities and university faculty salaries, costs that other countries typically include.

This likely means basic research spending is higher than the reported 97.55 billion yuan ($14.1 billion). But it still almost certainly falls short of the share devoted to basic studies in advanced countries. In 2016, that share was 16.9%, or $86.32 billion, in the United States and 12.8% in Japan, according to OECD. The share of R&D going to basic research in China "is not optimal," National Bureau of Statistics official Zhang Peng noted in comments released with the recent report.

Spending on basic research has been dwarfed by the rapid growth in business investment in R&D, most of it applied, which reached 77.6% of total R&D in 2017. Spending by regional and city governments, which topped central governmental spending last year, also tipped the balance away from basic research. Zweig says China's cities and provinces are competing to boost local economic growth by funding infrastructure and facilities that appeal to industry. One example is the Shanghai Synchrotron Radiation Facility. Completed in 2009 with central government construction funding, the accelerator is now jointly operated by the Chinese Academy of Sciences and the Shanghai municipal government. The facility is open to basic researchers, but businesses can also tap its intense beams of x-rays for commercial research, something that helped lure computer chip–makers and life science firms to Shanghai's Zhangjiang Hi-Tech Park, which hosts the synchrotron.

China's overall R&D spending is certain to rise. The government has set a target for R&D of 2.5% of gross domestic product (GDP) by 2020, up from 2.13% in 2017. By comparison, the United States spent 2.7% of GDP in 2016, according to OECD. Basic research should get an increasing share, Cao says. The current level of support "is not enough for China to become a scientific power."

Many are also wondering what a recent bureaucratic shake up will mean for science. This spring the National People's Congress decided to merge the National Natural Science Foundation, which supported a large share of the country's basic research efforts through competitive grants, into the Ministry of Science and Technology, which has managed nationally important big science projects. Some worry that the merger will erode support for basic research by small groups. The possible impact is not yet clear, Cao says, but "is something worth paying attention to." ■

PHOTO: IMAGINE CHINA/NEWSCOM

INFECTIOUS DISEASES

Outbreaks of poliolike disease pose a puzzle

An enterovirus may cause sudden paralysis in children, but proof is elusive

By **Gretchen Vogel**

Does a virus that usually causes mild cold symptoms sometimes paralyze children? That's the question facing scientists again this fall, after dozens of previously healthy kids across the United States suddenly lost muscle control in their arms or legs, a condition called acute flaccid myelitis (AFM) that eerily resembles polio.

Similar waves occurred in 2014 and 2016, and scientists have fingered a relative of the polio virus, called enterovirus D68 (EV-D68), as a possible culprit. But the evidence isn't conclusive yet, and it's unclear why the virus would only paralyze a small minority of children it infects. Solving these mysteries is urgent because the paralysis can be severe and irreversible. AFM is "pretty rare, but it's pretty devastating," says Priya Duggal, a genetic epidemiologist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, who's studying whether some patients may have a genetic vulnerability to the virus. "And it appears that it's cyclical. It's not going away."

EV-D68, which spreads through saliva and mucus, usually causes symptoms such as coughing, sneezing, and mild fever. It isn't routinely diagnosed, so no one knows how common it is. But a U.S. outbreak in late summer 2014 was more serious than usual: Hundreds of children hospitalized for severe breathing problems were diagnosed with the virus. Around the same time, more than a dozen children in the Kansas City, Missouri, area and in Denver, which both had large EV-D68 outbreaks, came down with sudden loss of muscle control. Many had mild fever or cough in the days before their symptoms developed, but were otherwise healthy. By December 2014, 120 children in 34 U.S. states had been diagnosed with AFM—"something that hadn't been reported since the days polio was circulating," says Mark Pallansch, a virologist at the Centers for Disease Control and Prevention (CDC) in Atlanta.

Scientists suspected a link between the cases of paralysis and the EV-D68 outbreak, but only one patient had evidence of the virus in their cerebrospinal fluid, the usual place to look for pathogens infecting the central nervous system, and fewer than half tested positive in stool or respiratory samples. That may be because samples were taken too late to detect the virus, researchers say; AFM typically appears more than a week after initial cough or fever symptoms.

The fall of 2016 saw another wave of the puzzling cases: 149 of them in 39 U.S. states. This time, 29 patients with AFM in 12 European countries tested positive for EV-D68 as well.

In the latest outbreak, 62 cases in 22 states have been confirmed; 65 more are under investigation, CDC officials said on 16 October. They did not say whether any patients have tested positive for enteroviruses, but "the pattern looks very familiar," says Carlos Pardo-Villamizar, a neurologist at Johns Hopkins University School of Medicine, including the timing: Autumn is known as enterovirus season.

Nine-year-old Jayden Broadway from Denver suffered from breathing problems caused by EV-D68 in 2014. The virus may also cause muscle paralysis.

PHOTO: CYRUS MCCRIMMON/GETTY IMAGES

Lab research has strengthened the case that EV-D68 could be involved. Last year, Kenneth Tyler of the University of Colorado (CU) School of Medicine in Aurora and his colleagues reported that several strains of EV-D68 could cause paralysis in mice. The virus seemed to target nerve cells involved in muscle control, which makes the link much more plausible, says Kevin Messacar, a pediatric infectious disease specialist also at CU who was not involved in the study. The study also suggests the virus is attacking nerve cells directly, he says, not just triggering an immune response that damages nerves. Others have demonstrated that the virus can infect neuronal cells in the lab. Meanwhile, scientists have found that EV-D68 can be detected much more reliably in respiratory fluids than in cerebrospinal fluid or stool samples, where the poliovirus shows up, which makes identifying infections easier.

If EV-D68 is causing paralysis, it's not clear why it does so in very few cases. But the same is true for the poliovirus, which causes only mild symptoms, or none at all, in more than 99% of those infected. One theory is that the virus can infect nerves in rare cases when it slips into an injured muscle. Another is that genetic factors play a role; Duggal and her colleagues have enrolled nearly 60 families in their study to test that theory.

Experts warn that as the 30-year campaign to eradicate polio winds down, attention will shift to other viruses that can paralyze children. (One known example is enterovirus 71; it produces blisters on children's mouths, hands, and feet, but also occasionally causes AFM.) "We are seeing more cases of this in the U.S. than we are seeing in polio in the whole world," Tyler says. "You don't need an awful lot of paralyzed children to make this an important problem."

There is no vaccine against EV-D68. A candidate vaccine developed in China has shown promising results in mice, but at the moment, the rate of serious complications from the virus is probably too low to make widespread vaccination worthwhile, says Heli Harvala, a virologist at University College London. Still, "It might be something we need to consider" in the future, she says. "It's good to be ready."

Rather than pinning hopes on a vaccine, doctors should be aware of the symptoms of AFM and know how to test for enteroviruses that might be involved, says Thea Kølsen Fischer, a viral epidemiologist at Nordsjællands Hospital in Hillerød, Denmark. "I would rather see investments in surveillance and diagnostics," which could help front-line clinicians be ready when the next cluster emerges. ■

TOXIN OR TREATMENT?

Ingesting small doses of peanut products guards against allergic reactions—but an undercurrent of anxiety persists

By **Jennifer Couzin-Frankel**

PHOTO: MADDIE MCGARVEY

Jacob Kingsley, 12, visits a bakery that was off-limits before he began oral immunotherapy for a peanut allergy.

Jacob Kingsley was 9 years old when he was handed the poison he'd shunned since before he could walk and told to swallow it as medicine. Obediently, he gulped down a few micrograms of peanut flour—less than 1/1000 of a peanut—diluted in grape Kool-Aid. His mother and a nurse hovered, ready to inject him with epinephrine if an itchy throat and wheezing struck.

Jacob's mother, Jennifer Kingsley, had driven him 2 hours from their home in Columbus to this doctor's office in Cincinnati, Ohio, for the first of dozens of sessions of peanut immunotherapy. Giving Jacob gradually increasing doses of peanuts, she hoped, would desensitize his immune system.

It's a strategy Kingsley hadn't pursued until she reached her breaking point. A year earlier, Jacob had swallowed a handful of popcorn that, unbeknownst to him, was laced with peanut product. He suffered a particularly frightening reaction: two bouts of intense symptoms about 6 hours apart. The incident marked his second peanut-related trip to the emergency room, and Kingsley was terrified that the next encounter could be fatal. "I decided, 'I can't live like this,'" she says. "I was desperate."

As Jacob sat through the hourslong appointment in Cincinnati, playing video games and swigging increasing doses of peanut-spiked Kool-Aid, he joined legions of children writing food allergy's next chapter. Today, more than 3000 people worldwide, most of them children, have undergone peanut immunotherapy, with the goal of protecting them if they accidentally encounter the food. Other children are trying immunotherapy for allergies to milk, eggs, and tree nuts. Some, like Jacob, get treatment in allergists' offices, where doctors share protocols informally and in published papers. Other children have enrolled in clinical trials, including those run by two companies racing to introduce a peanut-based capsule or skin patch. Both plan to apply for approval from the Food and Drug Administration (FDA) this year. The agency's blessing would dramatically boost immunotherapy's credibility and reach.

In a field that for decades has had nothing to offer patients beyond avoidance, immunotherapy marks a seismic shift. As it edges closer to mainstream, "There's mixed feelings, with a whole range of enthusiasm," says Corinne Keet, a pediatric allergist-immunologist at Johns Hopkins Medicine in Baltimore, Maryland. Fear that it might cause harm is mingling with euphoria that children living constrained lives could be set free. Doctors who offer immunotherapy describe families eating in Chinese restaurants for the first time and home-schooled children rejoining their peers.

Like many medical firsts, the therapy is not perfect. "This is version 1.0," says Brian Vickery, a pediatric allergist-immunologist at Emory University in Atlanta. He has conducted peanut immunotherapy trials and worked for 2 years at Aimmune Therapeutics, headquartered in Brisbane, California, one of the companies whose products are nearing approval. Physicians fret about oral immunotherapy's rigors—treatment must continue indefinitely—and its risks, which include the same allergic reactions it aims to prevent. Last year in Japan, a child suffered brain damage during a trial of immunotherapy for milk allergies.

Meanwhile, physicians on the front lines are navigating hazy science. No one knows exactly how immunotherapy works or who's most likely to be helped or hurt by it. "For me," Keet says, "it's really not clear for an average child with peanut allergy whether it will make sense to do oral immunotherapy or not."

LIKE MANY WHO STUDY food allergies, Keet was enticed by their mystery. Animal models are poor. The intensity of allergic reactions varies unpredictably, even in the same person over time. Why one child outgrows an allergy and another doesn't is unknown.

"This was something we didn't cover much in medical school" in the 1990s, says Matthew Greenhawt, a pediatric allergist-immunologist at Children's Hospital Colorado in Denver. Greenhawt's career trajectory tracks with a surge in food allergies, and these days, he can barely keep up with the stream of affected children who visit his hospital. Today, between 1% and 2% of people in the United States, the United Kingdom, and several other countries are allergic to peanuts—a rate that has roughly tripled since the mid-1990s. Other food allergies, such as those to tree nuts, are also on the rise. What's causing the increase is not well understood.

Despite rising caseloads, deaths from food allergies remain rare. Precise numbers are hard to come by, and estimates range from fewer than 10 to more than 150 a year in the United States. But even though an affected child is more likely to be struck by lightning than to die of a food allergy, the risk can feel ever-present. Parents never know when their children will happen upon culprit foods and how they'll be affected if they do. "We live in a complex world—people move food all over the place," says David Bunning, a businessman whose two sons, now adults, have multiple food allergies. "The impact on children in terms of their confidence to explore their environment can be extreme." Bunning's family almost never traveled or ate out. At their grandparents' house, the boys were usually confined to one room where food wasn't allowed.

Bunning now chairs the board of directors at Food Allergy Research & Education (FARE), an advocacy group in McLean, Virginia. Families like his, and the doctors who cared for their children, began to agitate for new treatments about a decade ago. Immunotherapy was the obvious candidate: Injections that desensitize the immune system to pollen, grass, pet dander, and bee venom have been around for decades.

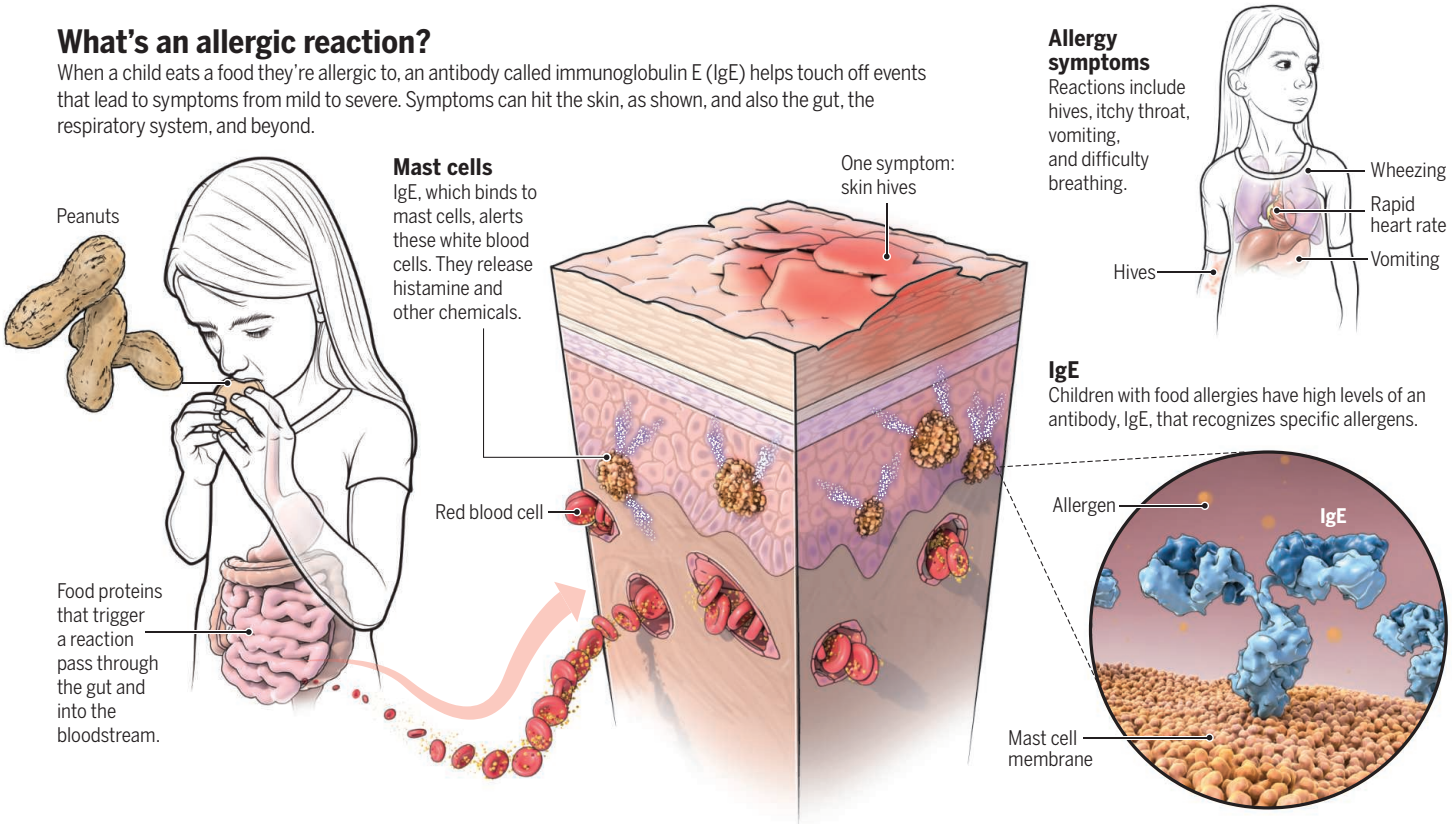
Whether for an allergy to cats or pistachios, immunotherapy aims to disrupt the cells that swing out of control when faced

Fighting fire with fire

Eating gradually increasing doses of a food allergen seems to desensitize the immune system over time. Thousands of children have tried oral immunotherapy, and a capsule to treat peanut allergies might be approved by regulators next year. But there's anxiety about the strategy's risks and unknowns.

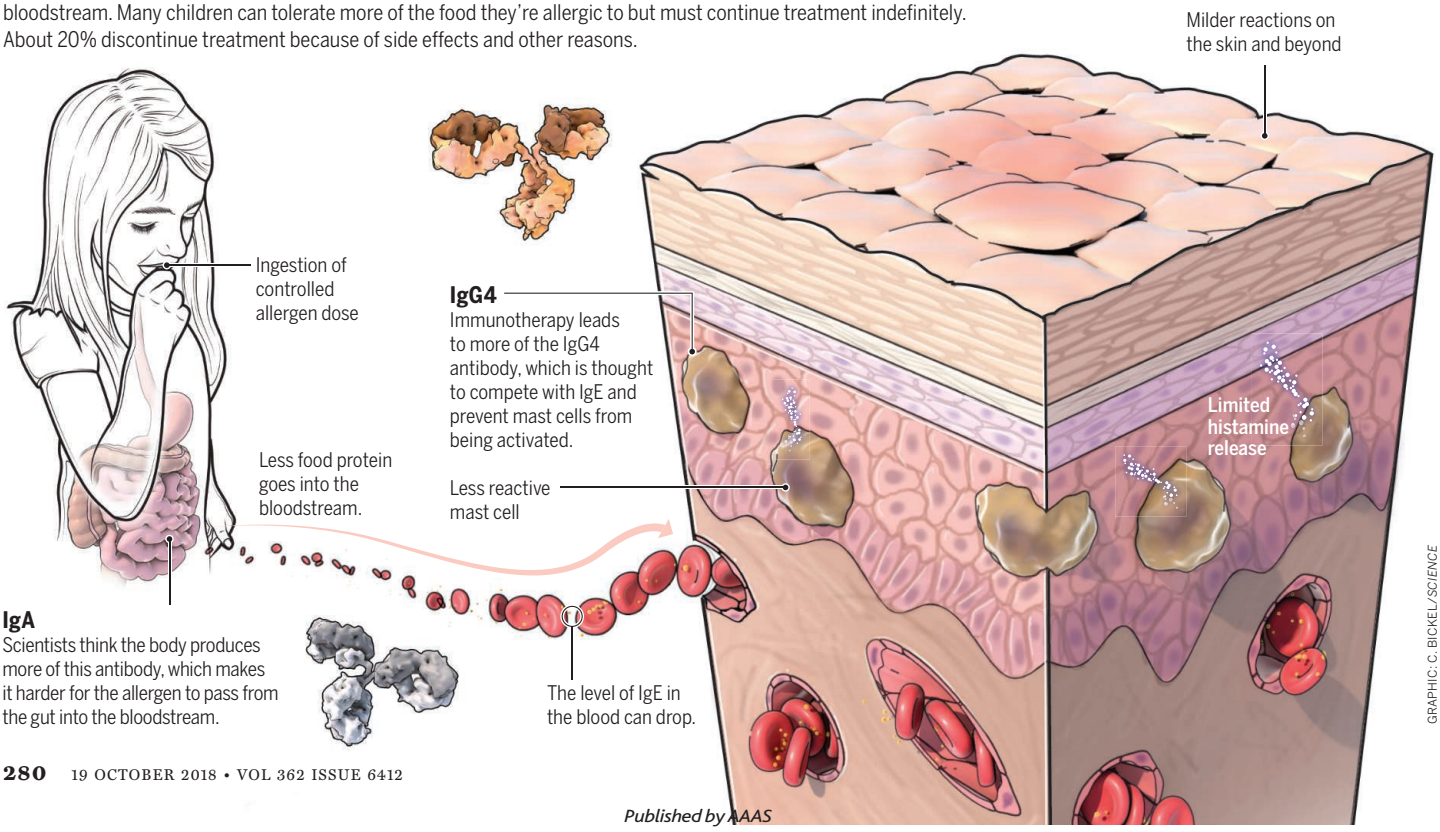
What's an allergic reaction?

When a child eats a food they're allergic to, an antibody called immunoglobulin E (IgE) helps touch off events that lead to symptoms from mild to severe. Symptoms can hit the skin, as shown, and also the gut, the respiratory system, and beyond.



How does treatment work?

Months of oral immunotherapy make mast cells less reactive and seem to reduce how much allergen enters the bloodstream. Many children can tolerate more of the food they're allergic to but must continue treatment indefinitely. About 20% discontinue treatment because of side effects and other reasons.



GRAPHIC: C. BICKEL/*SCIENCE*

with an allergen. When a child who is allergic to a food eats it, food proteins cross from the digestive tract into the bloodstream. An antibody called immunoglobulin E (IgE), which is bound to white blood cells called mast cells in tissues, recognizes the culprits. IgE activates the mast cells, which release histamine and other chemicals. In the skin, that response can lead to hives; in the respiratory tract, wheezing; and in the gut, vomiting. The most serious symptoms, such as a swollen throat or a reaction throughout the body, mark anaphylaxis, which is what families fear the most. Allergy shots blunt production of IgE, in part, researchers believe, by boosting levels of certain T cells that prompt a cascade of immune changes.

Brief testing decades ago indicated that shots for food allergies weren't safe. So around the mid-2000s, scientists began to feed children the allergen instead. One watershed moment came in 2005, when the National Institutes of Health formed a consortium for food allergy clinical trials. A second was in 2011, when advocates sponsored a symposium at Harvard Medical School in Boston to standardize goals and strategy for the pioneering immunotherapy efforts. About 60 people attended. "The patients were very clear," says Carla McGuire Davis, a pediatric allergist-immunologist at Texas Children's Hospital in Houston. They didn't care about eating a peanut butter sandwich; they wanted protection if they accidentally encountered one. Trialists set their end dose at a couple of peanuts and pressed ahead.

The results of early clinical trials were promising, says Hugh Sampson, a pediatric allergist-immunologist at the Icahn School of Medicine at Mount Sinai in New York City, who has studied immunotherapy in food allergies for many years. After 6 to 12 months of treatment, he says, about 70% to 80% of patients could handle higher doses of the food than before. Lab data were encouraging, too: Ingesting allergens over time seems to make mast cells less reactive, inhibiting their release of harmful chemicals. The therapy also produces other immunoglobulins: IgG4, which further inhibits mast cell activity, and IgA, which helps keep food allergens from escaping the gut (see graphic, p. 280).

The 2011 conference inspired the founding of the company now called Aimmune, fueled by more than $3.5 million from FARE. A second company, DBV Technologies, based in Montrouge, France, and New York City, expanded a few years later. Aimmune began to develop an oral product, essentially a capsule of powder derived from peanut flour with proteins held to consistent levels. In February, the company announced in a press release the results of a phase III trial involving 496 children and teenagers, with a regimen stepping up every 2 weeks through 11 dose levels. Among the 372 people in the treatment group, about 20% dropped out for various reasons, including side effects. After about a year, 96% of people who completed treatment could consume one peanut with no more than mild symptoms, 84% could tolerate two, and 63% could tolerate at least three.

DBV's skin patch represents a more conservative strategy: It delivers tiny amounts of peanut protein, the equivalent of one peanut over 3 years. Last year, DBV announced that in its phase III trial of almost 400 patients, after a year, those using the patch could, on average, eat three peanuts over the course of several hours before experiencing clinical symptoms such as vomiting or hives; before the trial, the average was just under one peanut. Outcomes varied substantially from person to person.

If one or both products are approved by FDA in the coming months, expectations are high that they'll be welcomed: Aimmune is now worth about $1.5 billion on the U.S. stock exchange. In 2016, FARE sold its share in Aimmune for $47 million.

"It made me nervous, really nervous, to put something in my daughter's mouth that she was allergic to."

Divya Balachandar, mother of Leena Wong

MEANWHILE, SOME DOCTORS embrace another route: offering peanut immunotherapy in their practices. "I can treat 20 patients with $5.95 of peanut flour," says Richard L. Wasserman, a pediatric allergist-immunologist in Dallas, Texas.

Wasserman ventured into food allergy immunotherapy 11 years ago. He developed a protocol based partly on published case reports and protocols for allergy shots, and he put IVs into his first five peanut allergy patients in case he had only seconds to rescue them from severe anaphylaxis. "When they all sailed through the first day, we stopped doing IVs," he says. "But that's a measure of how concerned I was."

Wasserman has since treated more than 300 children with peanut allergies and more than 400 with other food allergies. Other practitioners are joining in, among them the Cincinnati allergist whom the Kingsley family sought out: Justin Greiwe at Bernstein Allergy Group. Greiwe joined the practice in 2014, straight out of medical training. "It was a little nerve-wracking at the beginning," he says, because no officially sanctioned oral immunotherapy protocol existed. He took precautionary measures, such as lung testing before every treatment, to help ensure patient safety.

Some clinicians—and executives at the companies developing products—aren't happy about the doctor's office treatments. "That gives a lot of us pause," says Sampson, who in addition to his academic post is chief scientific officer of DBV. "We're very afraid that if this goes on enough, somebody is going to have an accident or a fatal reaction, and that's really going to change the FDA's viewpoint" about the products in development, he says.

Wasserman agrees about the need for caution. "Not every practicing allergist should be doing oral immunotherapy," he says. Greiwe suggests the treatment requires a dedicated staff, and he gives every immunotherapy family his cellphone number.

Jacob was one of Greiwe's first immunotherapy patients. His mother remembers Jacob's ears burning—a minor reaction that subsided on its own. "Or he said he hated peanuts and wanted to quit," she says. Worst was about 6 months in, when Kingsley discovered that for 2 weeks, Jacob had hidden his dose to avoid eating it. That was "the only time we ever felt danger," she says. Stopping treatment can quickly alter the immune system, says Cecilia Berin, an immunologist at Mount Sinai, because immunotherapy requires constant exposure. When Jacob squirreled away his daily dose, the changes induced in his immune system almost certainly started to fade out, putting him at risk. Greiwe restarted him on a lower dose and, his mother says, "We got through it."

EVEN CHILDREN WHO faithfully follow instructions face risks. The immune system can react to even subtle pressures, and the list of what can provoke a reaction to treatment is long. Exercising within a couple of hours of the dose can do it; so can a cold, a stomach virus, menstruation, or a hot shower. An asthma attack can trigger a reaction—many children with allergies have asthma as well—and so can stress. "We had a patient who had just played the violin on a stage, came down, and about 15 minutes later … took the dose and had a reaction," Davis says.

Berin posits that external pressures such as physical activity or illness make the gut more permeable, pushing more of the immunotherapy dose into the bloodstream. But that remains hypothesis. Regardless, it's becoming clear that "there are people who react years down the road to a maintenance dose," Keet says. For Jacob, such a moment came 9 months in. One evening while watching a movie, he downed his peanut

Food allergies are becoming more common, and a handful of foods accounts for the vast majority of allergies. But small doses of the foods can blunt allergic reactions.

M&M's and later ran outside with his cousins to dance in a rainstorm. He broke out in hives head to toe. Kingsley dialed Greiwe's number, and Jacob got a double dose of an allergy medication.

The most tragic data point to date is the case in Japan. A child had enrolled in a trial of immunotherapy for milk allergies at the Kanagawa Children's Medical Center in Yokohama. He'd raised what he could ingest from less than 8 milliliters to 135 milliliters—about half a glass of milk. After 3 months on that maintenance dose, he swallowed it and soon complained of pain. Within minutes, he had stopped breathing. His heartbeat was later restored in the emergency room, but he'd gone too long without it and sustained severe brain damage, according to a statement from the hospital's president, Sumimasa Yamashita, in November 2017. Kanagawa Children's Medical Center declined to comment, saying only that the incident remains under investigation.

In its statement, the hospital noted the boy had suffered an asthma attack the day before the catastrophic dose. He also was on a protocol that aimed to rapidly escalate the volume of milk he could drink over less than 3 weeks. But why the child reacted so disastrously to that glass of milk is unknown.

"What people don't understand is this level of protection fluctuates," says Mimi Tang, a pediatric allergist-immunologist at Murdoch Children's Research Institute in Melbourne, Australia. "It is not guaranteed, nor is it constant."

One of the few long-term analyses was published in 2013 in *The Journal of Allergy and Clinical Immunology*. Keet, pediatric allergist-immunologist Robert Wood at Johns Hopkins Medicine, and their colleagues sought out 32 children who'd been in a milk immunotherapy trial. Three to 5 years later, "The results were surprising in a sobering kind of way," Wood says. Only about a quarter "were doing great … tolerating unlimited quantities of milk without side effects." Another quarter had abandoned the protocol and returned to strict avoidance. The rest were eating dairy products inconsistently, with intermittent or even frequent allergic reactions. "It's hard to know which comes first, whether they got complacent" about ingesting it "or backed off because [they were] having too many symptoms," Wood says.

MORE AND MORE families are willing to live with those uncertainties because the alternative is greater anxiety. "We were scared senseless," says Divya Balachandar, whose daughter Leena Wong, now 7 years old, had her first episode of anaphylaxis at age 4 after being touched by a cashew. Testing revealed Leena also was allergic to sesame, eggs, milk, other tree nuts, and peanuts. Balachandar, a pediatric pulmonologist in New York City, and her husband enrolled Leena in a federally funded oral immunotherapy trial for peanut allergy in 2015. "It made me nervous, really nervous, to put something in my daughter's mouth that she was allergic to," Balachandar says. She gravitated toward a trial over treatment with a local allergist because, she says, "there were no rules" about how to treat in private practice. By this spring, Leena could eat two spoonfuls of peanut butter—about 25 peanuts—without a problem. She started second grade sitting with her classmates at lunchtime, liberated from a separate nut-free table.

Both companies developing peanut-based treatments say they had more volunteers for their trials than they could accommodate. Private practitioners usually have a waiting list; Greiwe's runs more than 4 months. At Stanford University in Palo Alto, California, which has a large food allergy research program, more than 2000 patients are waitlisted to enroll in the university's clinical trials, says Sharon Chinthrajah, an allergist-immunologist there.

More treatments are on the horizon. In Australia, Tang is working with a company that's testing an approach she pioneered, a combination of a probiotic and oral peanut immunotherapy. The probiotic should tilt the body toward producing the subset of T cells that tolerate the allergen and away from making cells that attack it, she says. Chinthrajah and others are enthusiastic about combining oral immunotherapy with a monoclonal antibody called omalizumab, which is FDA approved to treat allergic asthma. Clinical trials are also gearing up to test other monoclonal antibodies that target molecules involved in allergic inflammation.

Jacob's and Leena's families are eager to see what comes next. Jacob is also allergic to pistachios and cashews, but because he finds those foods easier to avoid than peanuts, the family has rejected immunotherapy that targets them. Leena's family is the opposite. With her older sister and her parents, Leena attends Indian functions regularly, where tree nuts are a common ingredient in sauces. In August, another episode of anaphylaxis landed her in the emergency room: She began to vomit and suffered chest tightness and eye swelling after eating Indian food her parents suspect contained cashews—despite having triple-checked with the restaurant that it did not. "I would love to do tree nuts," Balachandar says, once immunotherapy "becomes more available and better understood."

Physicians with deep roots in food allergy immunotherapy hope those new to it tread carefully. Doctors who offer such treatments "have to know the data cold," including published results and side effects that may crop up, Greenhawt says. Still, he's thrilled that peanut immunotherapy treatments may soon be approved. The other day, talking with a peanut-allergic 4-year-old and his mother, Greenhawt shared what the next year might bring. "I said, 'I'm going to see you a year from now; hopefully, we will have two products that are approved, and we can talk about which one might be best for you.'" The mother looked startled and delighted, Greenhawt says. "I've never seen somebody smile as brightly as that." ■

PHOTO: SCIENCE PHOTO LIBRARY/SCIENCE SOURCE

POLICY FORUM

CONSERVATION POLICY

Endangered species recovery: A resource allocation problem

Explicit articulation of values and objectives is critical

By **Leah R. Gerber, Michael C. Runge, Richard F. Maloney, Gwenllian D. Iacona, C. Ashton Drew, Stephanie Avery-Gomm, James Brazill-Boast, Deborah Crouse, Rebecca S. Epanchin-Niell, Sarah B. Hall, Lynn A. Maguire, Tim Male, Don Morgan, Jeff Newman, Hugh P. Possingham, Libby Rumpff, Katherine C. B. Weiss, Robyn S. Wilson, Marilet A. Zablan**

Many nations have laws to identify and protect imperiled species and their ecosystems. In the United States, actions taken under the Endangered Species Act (ESA) have prevented many extinctions, but few listed species have recovered to the point where they can have the ESA protections removed (*1*, *2*). One reason for this [among many (*3*)] is a shortfall in funding, raising a conundrum for agencies responsible for species recovery: Should resources be allocated toward species facing imminent extinction or species whose long-term survival can most benefit from investment? Some argue that the latter strategy is ethically unsound because it may abandon species with little hope of long-term recovery [for example, (*4*)], even when science suggests that the former strategy may miss opportunities to prevent species from ever experiencing the risk of imminent extinction (*2*). We suggest that framing recovery prioritization as a resource allocation problem provides a structure to facilitate constructive debate about such important questions. We discuss here the merits of an explicit resource allocation framework and introduce a prototype decision tool [(*5*); see supplementary materials for details] that we developed with the U.S. Fish and Wildlife Service (USFWS) to facilitate transparent and efficient recovery allocation decisions.

THE KNAPSACK PROBLEM

It is an inevitable consequence of limited resources that not all recovery efforts can be fully funded; therefore, some projects may be left behind. In some cases, managers make such choices (and neglect some projects) in ad hoc fashion, making implicit judgments about values and objectives with little transparency. We consider investment

See supplementary materials for author affiliations.
Email: leah.gerber@asu.edu

in endangered species recovery as a type of "knapsack problem" (*6*), in which a portfolio of recovery actions is chosen to best achieve a set of fundamental objectives, subject to a budget constraint. The science of decision analysis (*7*) provides tools to frame and solve such problems, allowing managers to allocate limited funds so that every dollar contributes as much as possible to the program objectives (*8*). Scenario-based decision-support tools can allow managers to compare how resource allocation strategies with different objectives, budgets, logistical constraints, and competing societal values may influence conservation outcomes and fundamental objectives before committing resources. Such approaches require decision-makers to clearly articulate objectives and constraints and to evaluate alternative allocation strategies to find which strategy best meets multiple objectives.

The first step in any structured resource allocation process is to identify which values the allocation seeks to reflect, because explicit articulation of objectives ensures transparent consideration of trade-offs among them. For investment in recovery, objectives may include minimizing the number of extinctions, maximizing the number of species recovered, favoring some species over others (owing to public support, taxonomic uniqueness, or other values), or minimizing the effect on other human activities (*6*, *9*). The ESA (and similar legislation in other countries) provides some guidance about the fundamental aims but usually not in enough specificity to guide allocation decisions across species and programs. Thus, the implementing agencies are left to define the objectives of recovery allocation themselves.

Even if the objectives for recovery actions were clearly articulated, deciding how to allocate limited resources to meet those objectives remains a vexing problem. First, the social, economic, and environmental values at stake are often contested and competing, requiring difficult trade-offs, even when resources are not limiting (*10*). Second, information on costs and benefits of different actions are hard to estimate and are often expressed as different metrics, making it difficult to weigh alternative allocation strategies (*10*). Third, the underlying complexity and uncertainty of the dynamic system (for example, regarding threats driven by climate change) means that effective implementation of conservation actions can be difficult to predict (*4*). Fourth, allocating resources (in the case of ESA) across more than 1500 federally listed species and thousands of individual recovery actions is a numerically daunting problem that cannot be solved transparently without a structured and explicit decision process. Fortunately, decision science provides tools to help overcome these challenges: Multicriteria decision analysis (*11*) allows a decision-maker to balance multiple objectives, and combinatorial optimization (*6*) allows the solution of complex knapsack problems.

RECOVERY ALLOCATION UNDER THE ESA

Very few federal and state agencies charged with management of endangered species have yet to adopt systematic methods for allocating recovery funds. Although the methods have existed in the academic literature for some time (*6*, *9*, *10*, *12*), it is challenging to implement them in any specific regulatory and institutional setting. With support from the National Socio-Environmental Synthesis Center, we worked with the USFWS to develop a tool to compare different funding allocation strategies for recovery of threatened and endangered species at the national and regional levels (*5*). This work was motivated, in part, by past critiques of USFWS recovery allocation processes [for example, (*13*)]. Our prototype Recovery Explorer tool (*5*) can be used, for instance, to examine how different values-based inputs (for example, desires for taxonomic representation or regional parity in funding) influence optimal allocation and recovery outcomes or the effect of uncertainty in technical inputs (for example, extinction risk or cost) on allocation and outcomes. The tool is meant to be exploratory, not prescriptive, allowing decision-makers to examine alternative approaches to resource allocation by making the important components of the decision process transparent.

Scientists face difficult decisions about how to allocate limited funding to the recovery of a few iconic and costly species (such as the black-footed ferret, facing page), and many lesser-known but less costly species (such as the Delhi Sands flower-loving fly, the deciduous tree *Kokia cookei*, and the Ozark hellbender, top to bottom, above).

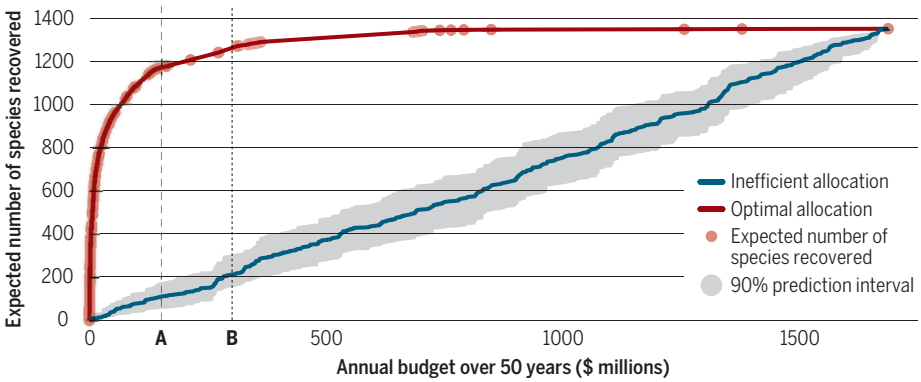
PHOTOS: (TOP TO BOTTOM) ERIC PORTER/USFWS; DAVID EICKHOFF/FLICKR/CC BY; JOEL SARTORE/CONTRIBUTOR/GETTY IMAGES

A fundamental tenet of decision analysis is to distinguish between science and values, and any decision-support tool requires both values-based and technical input. One of the advantages of a structured decision-making process is that the decisions and rationale regarding all of these components are made transparent. Values-based inputs to our decision framework include identification and specification of the recovery benefit metrics (for example, preventing extinction versus promoting recovery), the time frame over which benefits are measured, whether and how to weigh species differently to reflect taxonomic uniqueness or societal importance, and whether to constrain allocation to achieve parity across geographic regions. Science-based inputs include quantitative estimates of the benefit metrics with and without funding for each species, estimates of the cost of recovery for each species, and estimates of the likelihood of recovery conditional on funding for each species.

Our framework focuses on identifying cost-effective recovery plans, based on the assumption that managers want to complete the projects that achieve the most recovery benefit possible (however defined) with their available budget. The exact solution of knapsack problems requires combinatorial optimization, which is computationally expensive for large problems like this. An approximate solution can be found from the ranked list of project efficiencies, by adding plans from highest to lowest efficiency until the budget is exhausted (as in our algorithm) or by deleting plans from lowest to highest efficiency until the budget is achieved. This work represents the first effort to seek a solution to the specific problem of funding recovery efforts at the national scale in the United States, considering all species for which the USFWS has recovery plans.

Example of the benefit of optimal allocation

At an annual budget of $150 million (dashed line A), inefficient allocation of resources would recover about 104 species, whereas an optimal allocation is predicted to recover 1168 species. When annual funding is doubled to $300 million (dotted line B), the number of species recovered under the optimal allocation strategy increases to 1242. Data are from the U.S. Fish and Wildlife Service's database on recovery plans (*https://ecos.fws.gov/ecp0/ore-input/ad-hoc-recovery-actions-public-report-input*). Methods used to generate the "optimal allocation" curve are described in the supplementary materials; the "inefficient allocation" curve is based on a random selection of species for funding (mean and 90% prediction intervals).



ENCOURAGING CONSTRUCTIVE DEBATE

Use of the resource allocation framework has the potential to identify how more species can be recovered over time while preventing an equal or greater number of extinctions than less-efficient decision processes (see the figure), given a fixed budget. The framework can also demonstrate the expected gains from increased funding or, conversely, the expected losses from decreased funding. These two questions can be addressed simultaneously within a common framework for discussion.

Debate about how to value the outcomes of recovery allocation is contentious, passionate, and important. Explicit resource allocation tools provide a way to investigate alternative valuing systems, allowing users to compare the outcomes of focusing on imminent risk or long-term potential, of favoring charismatic or keystone species over others, and even of seeking a balance among such objectives. The approach we have developed with the USFWS allows such a conversation to begin. There are other features that might need to be added in future versions, such as the ability to use a benefit metric that weights multiple objectives. Such additions will be most valuable if they arise from deliberations by potential users.

Outside the United States, there have been sizable increases in recovery funding where managers responsible for species recovery and the agencies setting policies and budgets have developed and used such tools (*12*). For example, in New South Wales, an Australian state with >1000 threatened species, a transparent prioritization process demonstrated the responsible use of funds, showed benefits expected from a change in funding, and led to an unprecedented $100 million in additional funding (*14*). We believe that if similar decision processes were used to inform work under the ESA, alternative allocation strategies could maintain the success of the ESA at preventing extinctions while achieving far more recoveries at current levels of investment.

Some critics of applying a prioritization approach to endangered species suggest that instead of worrying about divvying up limited funds as species move toward extinction, we should be developing ecosystem-based approaches to address the root cause of species endangerment. We agree that a focus on mitigating threats is needed and further suggest that a transparent and cost-effective resource allocation will provide greater confidence in the use of funds and articulate what could be gained from increased investment, allowing agencies and conservation partners (for example, philanthropists, volunteers, and nongovernmental organizations) to evaluate, and possibly increase, their investment (*15*).

Underlying any allocation method is a set of predictions about how investment in a portfolio of species will result in the desired outcomes. As a first prototype, we used information from existing recovery plans and assumed that if actions suggested in those plans are undertaken, they will result in recovery of the species at the estimated cost. This assumption may not be warranted for all species. Further, we have assumed that species are largely independent and investment in one does not promote recovery of another, but there are cases where such synergies are possible. We have also only investigated full investment in each species, but partial investment might also be a valuable approach. Thus, a technical question arises for the USFWS and other agencies wishing to use such a resource allocation method: Are more detailed estimates of the costs and probability of success required?

Values-based judgments are embedded in any important public decision. It is up to society to determine the relative priority of investment in conservation of threatened and endangered species. The role of conservation scientists is to work with practitioners to guide decisions that are most likely to meet stated objectives and communicate associated risks and opportunities. Transparent tools for examining the effect of different allocation strategies on the outcomes of recovery spending allow agencies and the public to explore the effects of these choices. Resource allocation is not about saving some species and letting others go extinct; it is about finding a way to better order the work so that as many species as possible are recovered given the limited resources available at any moment in time. Embracing a thoughtful and consistent approach to the allocation of recovery funding improves the transparency in decisions, which is important when analyzing myriad management options under the ESA. ■

REFERENCES AND NOTES

1. D. Evans *et al.*, *Issues in Ecology* (no. 20) (2016).
2. L. R. Gerber, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3563 (2016).
3. J. M. Scott, D. D. Goble, A. M. Haines, J. A. Wiens, M. C. Neel, *Conserv. Lett.* **3**, 91 (2010).
4. J. A. Vucetich, M. P. Nelson, J. T. Bruskotter, *Front. Ecol. Evol.* **5**, 45 (2017).
5. A. I. Drew, S. Avery-Gomm, R. Maloney, L. R. Gerber, Recovery prioritization explorer (2017); https://shiny.sesync.org/apps/RecoveryExplorer/.
6. M. L. Weitzman, *Econometrica* **66**, 1279 (1998).
7. M. C. Runge, *J. Fish Wildl. Manag.* **2**, 220 (2011).
8. J. Boyd, R. Epanchin-Niell, J. Siikamäki, *Rev. Environ. Econ. Policy* **9**, 23 (2015).
9. H. P. Possingham *et al.*, *Trends Ecol. Evol.* **17**, 503 (2002).
10. H. B. Wilson, L. N. Joseph, A. L. Moore, H. P. Possingham, *Ecol. Lett.* **14**, 886 (2011).
11. R. L. Keeney, *Value-Focused Thinking: A Path to Creative Decisionmaking* (Harvard Univ. Press, 1992).
12. L. N. Joseph, R. F. Maloney, H. P. Possingham, *Conserv. Biol.* **23**, 328 (2009).
13. U.S. Government Accountability Office (GAO), "Endangered species: Fish and Wildlife Service generally focuses recovery funding on high-priority species, but needs to periodically assess its funding decisions" (Publication GAO-05-211, GAO, 2005).
14. Office of Environment and Heritage (OEH), "More plants and animals to be saved from extinction" (OEH, New South Wales Government, Australia, 2016).
15. J. R. Bennett, R. Maloney, H. P. Possingham, *Proc. R. Soc. B* **282**, 20142693 (2015).

ACKNOWLEDGMENTS

This work benefited from support from the National Socio-Environmental Synthesis Center (SESYNC)–NSF award DBI-1052875. C.A.D. was supported in part by a research contract from the Arizona State University Center for Biodiversity Outcomes. S.A.-G., H.P.P., and G.D.I. were supported by the Australian Research Council Centre of Excellence for Environmental Decisions. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/284/suppl/DC1

10.1126/science.aat8434

GRAPHIC: J. YOU/*SCIENCE*

PERSPECTIVES

SUSTAINABILITY

Well-being in metrics and policy

Well-being metrics provide key insights for economic and environmental sustainability

By **Carol Graham[1], Kate Laffan[2], *and* Sergio Pinto[3]**

This century is full of progress paradoxes, with unprecedented economic development and improvements in longevity, health, and literacy coexisting with climate change, persistent poverty in the poorest countries, and increasing income inequality and unhappiness in many wealthy ones. Economic growth and the traditional metrics used to assess it—particularly gross domestic product (GDP)—are necessary but not sufficient to guarantee growth that is inclusive and politically and socially sustainable. Well-being metrics, derived from large-scale surveys and questionnaires that capture the income and nonincome determinants of individual well-being, often provide a different picture of what is happening to people. These metrics can provide insight into policies to sustain human welfare in the future.

The United States has one of the wealthiest economies in the world, yet life expectancy is falling owing to deaths driven by suicides and drug and alcohol overdose. This particularly affects Caucasians with less than a college education. An increasing proportion of this group—15% of males in their prime years (25 to 54 years old)—has dropped out of the labor force. Poor Caucasians report much less hope for the future and more stress than do poor African Americans and Hispanics, who face higher objective disadvantages (*1*). This toxic combination yields a loss of welfare and productive potential and the resurgence of nativism and support for antisystem populists who promise a return to the past (*2*).

China is perhaps the most successful example of rapid growth and poverty reduction in modern history. GDP per capita increased fourfold between 1990 and 2005, and life expectancy increased from 67 to 73.5 years. Yet life satisfaction fell dramatically, and suicide increased, reaching one of the highest rates in the world (*3*). The unhappiest cohorts were educated workers in the private sector, who benefited from the growing economy but suffered from long working hours and lack of sleep and leisure time.

The most recent economic success story is India. Our calculations, based on data in (*4*), show that life satisfaction dropped by 10% from 2006 to 2017. Again, unhappiness and ill-being coincide with the positive story that economic indicators are telling.

Many studies demonstrate that nonincome factors—such as norms, expectations, and stigma—matter more to human welfare than standard economic models assume. For example, promotions have much more lasting effects on life satisfaction than do salary increases (*5*), and the unemployed are less unhappy when local unemployment rates are higher because they experience less stigma (*6*). Higher levels of happiness and optimism tend to lead to better individual future outcomes, including in income, health, and friendship (*7*). Related research highlights the interaction of genetic determinants of well-being (such as certain aspects of intelligence and of the immune system) with environmental ones in determining longer-term earnings and health outcomes (*8*). Experimental studies show that provoking optimistic thoughts, such as recalling a time that individuals felt proud, or providing an asset, such as a cow, that provides hope to the poor lead to investments by individuals in work, health, and education, which result in better futures (*9*, *10*). These links between well-being, productivity, and health are critical to future sustainability.

TOMORROW'S EARTH

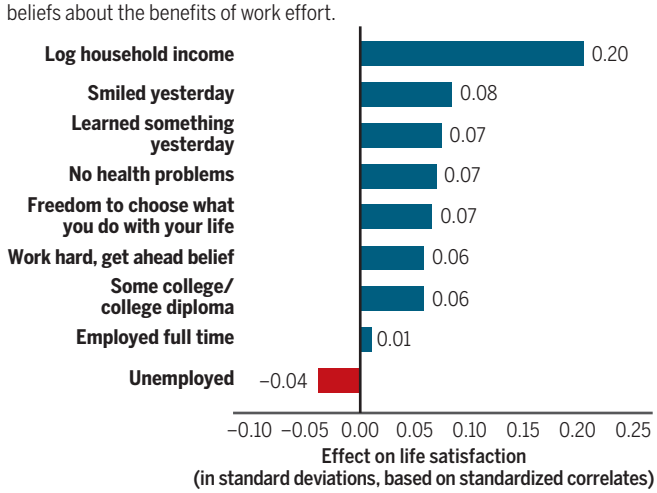
Read more articles online at scim.ag/TomorrowsEarth

There is now best practice for well-being questionnaires, with consensus on three distinct dimensions of well-being: hedonic, evaluative, and eudaimonic. Hedonic metrics capture individuals' affective states—such as enjoyment, stress, or anger—and the role they play in daily living. They evaluate daily quality of life, such as the effects of various health conditions, and assess the effects of engaging in activities such as smoking or exercising. Evaluative metrics, which are the most common, assess individuals' satisfaction with their lives over their lifetime, including whether they can choose the kinds of lives they want to lead. Eudaimonic metrics ask whether individuals have purpose or meaning in their lives. Income correlates more closely with evaluative metrics than with hedonic ones because poverty is bad for all dimensions of well-being, but beyond a certain point, more money will not improve moods, friendships, or commutes.

Common patterns in the distribution of responses across large populations and over time, as well as validation from psychological measures of well-being, support an underlying consistency in well-being metrics. Respondents are not asked whether certain things make them happy; rather, well-being questions are asked directly—for example, "generally speaking, how satisfied are you with your life as a whole?"—followed by a range of questions about socioeconomic and demographic factors that are

Correlates of life satisfaction around the world

The life satisfaction of individuals worldwide correlates with income, health, employment, and education as well as with positive moods, freedom, and beliefs about the benefits of work effort.



The figure is based on standardized coefficients using 2009–2012 Gallup World Poll data (*16*).

CREDITS: (TOP TO BOTTOM) ADAM SIMPSON/HEART AGENCY; A. KITTERMAN/*SCIENCE*

associated with life satisfaction. Controlling for these, we can explore things that vary across people and places, such as inequality, commuting time, volunteering, freedom and learning, and attitudes about work (see the figure).

Well-being metrics have limitations. They can only infer causality when tracking the same respondents over time, or after an intervention. Otherwise, there is often two-way causality: Happier people are more likely to marry each other and to make friends as well as to derive benefits from those relationships. Another issue is scale interpretation: We cannot assume that a score of eight for life-satisfaction is equivalent to double a rating of four. Adaptation also poses challenges. People are remarkably adaptable to many experiences (good and bad). When a poor person living in terrible conditions responds that they are "quite happy," we do not know whether that is true or whether they have learned to live with insurmountable conditions. One resolution is to compare the same respondent's scores on daily experience questions with those on life satisfaction. People in compromised conditions may respond positively to the former yet tend to score lower on the latter because they lack the means to choose the kinds of lives they want to lead. Additionally, we can use vignettes, which ask respondents to rate alternative scenarios that, for example, tease out how women think they should respond in the context they live in versus how they actually feel, to allow for scale differences across cultures, races, and gender.

Well-being metrics can contribute to policy design, monitoring, and evaluation in a range of areas. For example, most economic models assume that inflation and unemployment affect welfare equally badly, yet unemployment rates have more negative effects on well-being. Places and people with good public health, education, and welfare systems have higher levels of well-being. Yet, standard economic models do not incorporate these factors.

Well-being metrics can inform on social issues. There is a consistent "U"-shaped relationship between age and life satisfaction (controlling for health and income), with most dissatisfaction (and stress) occurring in middle age (40 to 54 years old) (*11*). This is not a marginal issue for policy; for example, the highest rates of overdose and suicide in the United States are in these middle-age years. Another is gender rights. Although women are typically happier than men (except when their rights are compromised), they may experience happiness losses when rights equalize because the shifting of social norms can result in intrahousehold conflicts or other negative outcomes not accounted for in policy design (*12*).

Policies based solely on income-based cost-benefit analysis can fail to capture important side effects. For example, closing rural post offices may make sense from a budget perspective; they are expensive to reach and do not deliver much mail. But, well-being surveys in the United Kingdom showed that the daily post office visit is an important social event for isolated residents, particularly the elderly. Understanding such aspects of well-being could avoid policies with unforeseen negative side effects.

Well-being metrics can also influence environmental sustainability. Airport noise and air pollution, as well as transient conditions such as flooding and drought, have substantial life satisfaction costs. Moreover, individuals living in greener urban areas and on coasts report higher life satisfaction and less mental distress. Policy calculations of the benefits of pro-environmental behaviors such as recycling account for time and money costs. Yet these behaviors are positively associated with well-being, which could alter policy calculations (*13*).

"These links between well-being, productivity, and health are critical to future sustainability."

Many scholars believe that happiness should be the primary objective of policy, citing the limits of economic growth. Indeed, Easterlin *et al.* found that decades of growth did not yield increased average national happiness (*14*). There is extensive debate over the accuracy of this paradox, which depends on the country sample—because the gains in happiness from growth are greater for poor countries—and on the time range of the analysis.

Supporters of the proposal to make happiness the central objective of policy cite the common drivers of happiness among people worldwide, which in turn allows for assessing the relative importance of related social indicators. Happiness metrics are "democratic" because they are based on responses from individuals, not what scholars think influences happiness. These metrics also identify important—and often costly—trends in ill-being and mental illness.

Skeptics believe that well-being metrics and the research based on them are valuable inputs into policy-making but should not be the measures of success. They allow scholars to assess the relative importance of various conditions or institutional arrangements and can complement welfare assessments based on economics. Additionally, making happiness an objective of policy raises challenges, including differences in peoples' conceptions of happiness, adaptation, changing expectations that influence individuals' evaluations, and possible political manipulation of the information by cherry-picking findings.

This is not a hypothetical debate. Bhutan made Gross National Happiness its official national development strategy in 2008. The U.K. government does not endorse that strategy but has included well-being questions in its Annual Population Survey since 2012. In 2013, the Organization for Economic Cooperation and Development issued guidelines for well-being metrics in statistics, and a U.S. National Academy of Sciences panel provided recommendations for well-being metrics in policy (*15*).

There are also myriad local efforts to measure well-being to inform policy. For example, the What Works Well-Being Program assesses the well-being effects of interventions in deprived communities in the United Kingdom, and the U.S. city of Santa Monica uses a well-being index to guide better municipal policies. Informing sustainability objectives is a natural extension because these efforts assess the financial and well-being benefits and costs of interventions such as more green spaces, opportunities for volunteering, and commuting arrangements.

Well-being metrics can serve as warning lights; they point to vulnerabilities in particular places or groups of people and to positive trends that could provide broader lessons. GDP and well-being indicators can and should coexist to play a role in public and policy debates. ■

[1]*Global Economy and Development Program, The Brookings Institution, Washington, DC 20036, USA.* [2]*Psychological and Behavioral Science Department, London School of Economics, London WC2AZAE, UK.* [3]*School of Public Policy, University of Maryland, College Park, MD 20742, USA. Email: cgraham@brookings.edu; k.m.laffan@lse.ac.uk; stpinto@umd.edu*

REFERENCES

1. C. Graham, S. Pinto, *J. Popul. Econ.* **31**, 1 (2018).
2. S. Monnat, D. Brown, *J. Rural Stud.* **55**, 227 (2017).
3. C. Graham *et al.*, *World Dev.* **96**, 231 (2017).
4. J. Lall, *Gallup.Com*, July 1 (2018); https://news.gallup.com/opinion/gallup/236357/indians-life-satisfaction-goes-bust-economy-booms.aspx.
5. R. DiTella *et al.*, *J. Econ. Behav. Organ.* **76**, 834 (2010).
6. A. Clark, A. Oswald, *Econ. J.* **104**, 648 (1994).
7. C. Graham *et al.*, *J. Econ. Behav. Organ.* **55**, 319 (2004).
8. J. E. De Neve, A. Oswald, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19953 (2012).
9. C. C. Hall *et al.*, *Psychol. Sci.* **25**, 619 (2014).
10. J. Haushofer, E. Fehr, *Science* **344**, 862 (2014).
11. D. Blanchflower, A. Oswald, *Soc. Sci. Med.* **66**, 1733 (2008).
12. R. Lalive, A. Stutzer, *J. Polit. Econ.* **92**, 933 (2010).
13. T. Kasser, *Phil. Trans. R Soc. A* **375**, 2095 (2017).
14. R. A. Easterlin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22463 (2010).
15. A. Stone, C. Mackie, Eds., *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Human Experience* (The National Academies, 2013).
16. C. Graham, M. Nikolova, *World Dev.* **68**, 163 (2015).

10.1126/science.aau5234

QUANTUM COMPUTING

Computational complexity, step by step

For certain linear algebra tasks, quantum circuits are proven to outperform classical ones

By **Ashley Montanaro**

It is widely believed that large-scale quantum computers, when they are built, will outperform their standard, "classical" counterparts. This supposition has inspired huge public interest and very substantial state and private investment in the development of quantum computing hardware. Yet, is it actually correct? On page 308 of this issue, Bravyi *et al.* (*1*) prove the first rigorous separation between two analogous and natural quantum and classical computational-complexity classes.

Quantum computers can solve certain problems, such as integer factorization (*2*) and simulation of quantum systems (*3*), exponentially faster than our best classical algorithms. Perhaps surprisingly, there is no rigorous mathematical proof that no better classical algorithm exists. Indeed, it is often extremely challenging to prove separations between computational-complexity classes. In particular, proving the existence of an exponential quantum speed-up would be a major breakthrough in complexity theory.

Here, the authors sidestep this barrier by considering the setting of "low-depth" computations. These are computations that can be interpreted as a short sequence of groups of elementary operations (quantum or classical logical gates), where each group of operations takes place in parallel. Bravyi *et al.* describe an explicit family of problems that can be solved with a quantum circuit of constant depth, whereas any classical circuit must have depth that scales logarithmically with the size of the problem. Further, the quantum circuit contains only operations that act across nearest neighbors on a two-dimensional square lattice,

The problem the authors use, called "hidden linear function" (HLF), can be defined in terms of linear algebra over finite fields, or perhaps more naturally in terms of the correlations arising from measuring a certain family of entangled quantum states. An HLF problem instance is specified in terms of an $n \times n$ binary matrix A, which we can think of as the adjacency matrix of a graph. This matrix A defines a certain quadratic function of n bits, which turns out to behave like a linear function when restricted to the binary null-space of A. The HLF problem is to determine this linear function.

Bravyi *et al.* first show that the HLF problem can be solved using a quantum computer by preparing the entangled "graph state" (*4*) corresponding to A, and then performing a particular local measurement on each qubit of this quantum state (see the figure). The remaining technical challenge is to show that the HLF problem cannot be solved by sub–logarithmic-depth classical circuits, even in the special case where A is a subgraph of a square lattice. A beautiful aspect of their proof is how it shows that no low-depth classical circuit can reproduce the required quantum correlations. The approach is inspired by, and can be seen as a generalization of, the well-known concept of Bell inequalities separating quantum mechanics from local hidden-variable theories (*5*).

Bell inequalities certify that there are some quantum correlations that cannot be reproduced by separated classical players who are only allowed to perform local operations. These local operations can be seen as very simple classical computations. Here, the authors generalize this idea to more complicated computations, namely low-depth circuits. They consider correlations resulting from graph states corresponding to simple cycles. Generalizing a result previously shown in the context of quantum foundations (*6*), Bravyi *et al.* show that these correlations cannot be reproduced with low-depth and geometrically local classical circuits, where the gates act locally with respect to the graph under consideration.

This result is then lifted to the case of geometrically nonlocal circuits through a careful technical argument. This argument shows that, given any low-depth classical circuit, it is possible to find a subgraph A of the square lattice with respect to which that circuit acts locally, but where long-range correlations are required to reproduce the quantum measurement statistics and solve the corresponding HLF problem. It is even possible to infer an average-case hardness result from this, which proves the existence of a family of problems such that most of them cannot be solved by a low-depth classical circuit.

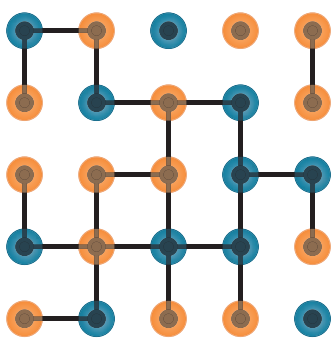
The HLF problem itself has a relatively time-efficient classical algorithm, so it will not enable us to prove directly that quantum computers can solve certain problems exponentially faster than classical computers can. Although the complexity-theoretic result of Bravyi *et al.* is far away from such a quantum-classical separation, it represents substantial progress in this direction. The true classical complexity of solving the HLF problem—what is the most powerful class of classical computations that is not able to solve it?—remains to be determined. The techniques introduced here may help find other problems of more practical import that are classically hard but solvable with low-depth quantum circuits. ■

Qubits for linear algebra

To solve the hidden linear function problem for a matrix A corresponding to a subgraph G(A) of a square lattice, a graph state can be prepared using entangling operations, indicated by black edges, across qubits.

Touching bases

Each qubit is measured in one of two measurement bases, indicated by orange or blue circles.



GRAPHIC: N. CARY/*SCIENCE*

School of Mathematics, University of Bristol, Bristol BS8 1TW, UK. Email: ashley.montanaro@bristol.ac.uk

REFERENCES AND NOTES

1. S. Bravyi, D. Gosset, R. König, *Science* **362**, 308 (2018).
2. P. W. Shor, *SIAM Rev.* **41**, 303 (1999).
3. I. M. Georgescu, S. Ashhab, F. Nori, *Rev. Mod. Phys.* **86**, 153 (2014).
4. M. Hein, J. Eisert, H. J. Briegel, *Phys. Rev. A* **69**, 062311 (2004).
5. N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).
6. J. Barrett, C. M. Caves, B. Eastin, M. B. Elliott, S. Pironio, *Phys. Rev. A* **75**, 012103 (2007).

ACKNOWLEDGMENTS

A.M. is supported by the UK EPSRC (EP/L021005/1), is cofounder of the quantum software company PhaseCraft, and consults on quantum computing.

10.1126/science.aau9555

CELL BIOLOGY

Supracellular contractions propel migration

Cytoskeletal cords connecting cells at the back of cell groups enable directional migration

By Igor Adameyko[1,2]

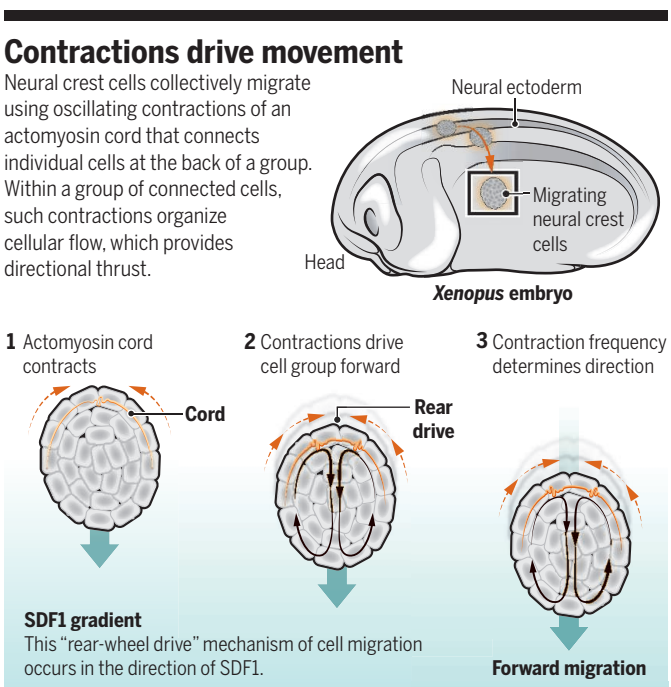
Constructing multicellular bodies, starting from a single-cell zygote, often requires the movement of cells across considerable distances, which is achieved through cell migration. During embryonic development, as well as in healing and regeneration, cells travel across diverse terrains, which dictates the character of navigation (*1*). Cancer cells metastasize and migrate into healthy organs, and knowledge of their migration strategies could be important to identify targets to treat advanced disease (*2*). Some migratory cells cover large distances individually (*3*), whereas others migrate in groups, with leaders and followers being directed by chemical signals (chemotaxis) (*4*, *5*). The exchange of information and resulting motility of such groups has been enigmatic. Moreover, the driving force of collective cell migration has been considered a sum of migratory and signaling activities of individually participating cells. However, according to a study on page 339 of this issue by Shellard *et al.* (*6*), collective cell migration requires formation of cytoskeletal structures that span through adjoining cells at the rear of a cell group to coordinate, orient, and propel the entire group. This mechanism of collective cell migration could be applicable to cancer metastasis and wound healing and might change our understanding of developmental migration.

Shellard *et al.* discovered this unexpected mechanism of collective cell migration by investigating the journey of neural crest cells following a chemotactic gradient of stromal cell–derived factor 1 (SDF1) in developing African clawed frog (*Xenopus laevis*) and zebrafish model systems. Neural crest cells are multipotent embryonic progenitors that emerge from the lateral neural ectoderm (a germ cell layer in developing embryos) during development of all vertebrate embryos. The neural crest is often considered a fourth germ layer (in addition to ectoderm, mesoderm, and endoderm) because of its developmental, evolutionary, and clinical importance. From the neural ectoderm, neural crest cells travel along specific routes to the periphery of the embryo to generate different cell and tissue types, including cartilage, bone, dermis, Schwann cells (that cover neuronal projections), pigment cells, and peripheral autonomic and sensory neurons (*7*). Correct migration of neural crest cells is essential for successful embryo development, and failures in the migration process often lead to developmental pathologies such as Waardenburg syndrome and craniofacial microsomia (*8*).

Shellard *et al.* found that the migration of the neural crest in developing embryos requires tight intercellular bonding combined with the formation of previously unknown supracellular cords, composed of the cytoskeletal proteins actin and myosin, that span through numerous individual cells and generate regular (oscillatory) contractions. Moreover, they found that differences in the oscillation frequency of the cord contractions at the rear of the moving cell group provide directional thrust (see the figure). They called this "a rear-wheel drive" model.

To obtain a holistic understanding of how the rear-wheel drive works, the authors generated a computer simulation, in which the contractility and translocations of neural crest cells at the periphery and within the cluster could direct the displacement of the whole migrating group. The model predicted that during oriented migration, individual cells at the contractile rear end would be propelled forward, which in turn would cause a wave of rear-to-forward displacement of many cells within a moving crowd and thus fuel the movement of the entire group. They also found that the frequency of the contractions of the supracellular actomyosin cord is controlled by the position of the cell group within a gradient of SDF1 concentration and not the absolute local concentration of SDF1.

Some intriguing questions remain. It will be exciting to understand how the supracellular contractile cord is constructed and initiated inside the individual cells involved. Additionally, it will be beneficial to understand how the oscillation frequency of the cord is regulated and how this is coupled to the position of a cell group within a chemoattractant gradient. Importantly, it is essential to reconcile the new rear-wheel drive mechanism with other proposed mechanisms of neural crest migration. For example, contact inhibition of locomotion (in which motility is inhibited by contact with neighboring cells), another characteristic of neural crest migration (*9*), may play unanticipated roles in the fine operation of the rear-wheel drive mechanism. During neural crest migration, the space between migrating cells becomes extremely compact, and the contractions might reduce that space and squeeze the rear cells even further. The resulting translocation of rear cells might lead to contact inhibition of locomotion that will polarize



GRAPHIC: C. BICKEL/*SCIENCE*

[1]Department of Physiology and Pharmacology, Karolinska Institutet, 17177 Stockholm, Sweden. [2]Center for Brain Research, Medical University of Vienna, 1090 Vienna, Austria. Email: igor.adameyko@ki.se

front cells in the forward direction.

How often might this mechanism of collective cell migration occur in nature? Contractile supracellular actomyosin cables and rings have been observed in other developmental settings. For example, they operate in cell-sorting processes (*10*) and in developmental events that shape the ascidian axial skeleton (a notochord) (*11*). However, these supracellular cords do not seem to participate in cell migration. It is possible that supracellular cytoskeletal structures are widely distributed in nature, especially because they may generate biomechanical forces necessary for intercellular coupling that occurs in many different settings, including boundary formation or collective cell migration of healthy and pathological cells. Cancer cells might produce supracellular cytoskeletal cords for driving their metastatic migration (*12*). The rear-wheel drive mechanism might also be involved in wound healing, which requires collective ingression of cells into the open wound (*13*).

"...in this new view of collective cell migration, individual cells team up and form a 'supracell'."

Therefore, in this new view of collective cell migration, individual cells team up and form a "supracell." The contractions at the rear end and resulting flow of cells inside of a moving cluster resemble some viscoelastic materials and gels that are capable of translocating forward by creating a rotating flow or tangential retrograde movement of their surfaces (*14*, *15*). Through this mechanism, cell migration might be more efficient and become more adaptive to a plethora of challenging tissue landscapes. ■

REFERENCES

1. H. M. Olson, A. V. Nechiporuk, *Front. Cell Dev. Biol.* **6**, 83 (2018).
2. D. Stichel *et al.*, *NPJ Syst. Biol. Appl.* **3**, 5 (2017).
3. D. H. Tanaka *et al.*, *Development* **133**, 2167 (2006).
4. D. Dalle Nogare *et al.*, *Development* **141**, 3188 (2014).
5. R. McLennan *et al.*, *Development* **142**, 2014 (2015).
6. A. Shellard *et al.*, *Science* **362**, 339 (2018).
7. M. E. Bronner, N. M. LeDouarin, *Dev. Biol.* **366**, 2 (2012).
8. T. N. Snider, Y. Mishina, *Birth Defects Res. C Embryo Today* **102**, 324 (2014).
9. C. Carmona-Fontaine *et al.*, *Nature* **456**, 957 (2008).
10. L. H. Chang *et al.*, *Development* **138**, 3803 (2011).
11. I. M. Sehring *et al.*, *eLife* **4**, e09206 (2015).
12. J. Plou *et al.*, *Sci. Rep.* **8**, 12723 (2018).
13. L. Li, Y. He, M. Zhao, J. Jiang, *Burns Trauma* **1**, 21 (2013).
14. K. Kruse *et al.*, *Phys. Rev. Lett.* **92**, 078101 (2004).
15. P. R. O'Neill *et al.*, *Dev. Cell* **46**, 9 (2018).

10.1126/science.aav3376

DEVELOPMENT

Toward human egg-like cells in vitro

Human precursor egg-like cells are produced in mouse ovary organoids to study infertility

By **Mark E. Gill**[1] *and* **Antoine H. F. M. Peters**[1,2]

An increasing percentage of couples suffer from infertility, partly related to rising reproductive age. Extensive research efforts seek to understand the causes of idiopathic infertility, such as defects in germ cell (egg and sperm) development (gametogenesis). However, gametogenesis is a long differentiation process that begins in utero in the developing fetus, hampering research. On page 356 of this issue, Yamashiro *et al.* (*1*) describe the generation of human oogonia-like cells (precursor egg-like cells) from male and female human induced pluripotent stem cells (hiPSCs), following an in vitro differentiation approach that had been successfully applied to generate mature mouse eggs in vitro (*2*). Generating human germ cells in vitro is an important advance toward improving human reproduction because it will assist scientists and clinical embryologists to identify critical modes of regulation and to develop clinical treatments that support the formation of healthy gametes.

Yamashiro *et al.* directed hiPSCs, derived from peripheral blood mononuclear cells (PBMCs) (*3*, *4*), toward germ cell fate by exposing them to various signals. The resulting human primordial germ cell–like cells (hPGCLCs) resemble fetal stage germ cells, called primordial germ cells (hPGCs) (*3*). To initiate sexual differentiation, Yamashiro *et al.* combined hPGCLCs with somatic cells from mouse fetal ovaries to form three-dimensional organoid-like aggregates, called xenogeneic reconstituted ovaries (xrOvaries), and cultured them for 17 weeks (see the figure). The organoids underwent morphological changes akin to those occurring during human ovary development in vivo. Moreover, genome-wide gene expression and DNA methylation analyses indicate that the differentiating hPGCLCs resembled human ovarian germ cells—at a stage before their entry into meiosis, a critical phase in egg development—and hence they were called oogonia.

The ability to induce hPGCLCs from PBMCs stems from observations in mammals that germ cells originate from somatic cells in developing embryos exposed to signals [for example, bone morphogenetic proteins (BMPs) and Wnts] emanating from the neighboring cellular environment, a process called epigenesis (*5*, *6*). Epigenesis is prevalent among metazoan phyla (*7*) and can be recapitulated in vitro, driving the fate of mouse and human embryonic stem cells (ESCs) and iPSCs toward that of PGCLCs (*3*, *8*, *9*).

After their specification, mouse and human PGCs migrate to developing gonads and undergo epigenetic reprogramming. This entails removal of DNA methylation throughout the genome and, in females, reactivation of the X chromosome that was repressed before germ cell specification (*5*, *10*). Erasure of DNA methylation is critical for the proper epigenetic regulation of autosomal "genomically imprinted" genes. Remethylation during succeeding egg or sperm development controls monoallelic expression of these genes after fertilization, either from the maternal or paternal genome, which is essential for fetal and placental development.

The gonadal environment is instructive to the fate of germ cells, irrespective of their chromosomal sex. Ovarian somatic cells direct oogonia to enter meiosis (*5*). Meiotic entry constitutes an irreversible change of cell identity, characterized by the programmed generation of many DNA double-strand breaks that initiate the exchange of genetic information between parental genomes. Through the subsequent formation of contacts with somatic granulosa cells, oocytes expand and accumulate many factors and epigenetic changes on maternal chromosomes that are required to support successful embryogenesis (*11*).

The xrOvaries created a somatic environment that is instructive for hPGCLC differentiation. Importantly, differentiating hPGCLCs underwent global erasure of DNA methylation, including at genomi-

[1]*Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.* [2]*Faculty of Sciences, University of Basel, 4056 Basel, Switzerland. Email: antoine.peters@fmi.ch*

cally imprinted loci. However, genes initiating meiotic prophase failed to become expressed after 17 weeks in culture. In vivo, these genes are expressed around week 14 of female gestation (*12*). At 11 weeks of culture, Yamashiro *et al.* found that regulatory regions of key meiotic genes in xrOvaries had not yet undergone derepression through DNA demethylation. In mouse PGCLCs (mPGCLCs), orthologs of such genes also escape DNA demethylation and harbor repressive histone modifications (*5*). They are silenced by Polycomb repressive complex 1 in mPGCs in vivo (*13*). Extended proliferation of mPGCLCs, however, induces demethylation of meiotic genes and leads to their transcriptional activation upon stimulation by exogenous BMP2 and retinoic acid (RA) signaling (*5*). Studies in mESCs identified the TET1 methylcytosine dioxygenase as an activator of meiotic genes, independently of its role in DNA demethylation (*14*). Thus, both the complete removal of repressive chromatin and potent activating signals from the ovarian soma are critical for meiotic gene expression. Mouse ovaries probably do not provide proper signaling for meiotic entry of human oogonia. The addition of exogenous human signaling moieties—or, addition of reprogrammed human granulosa cells—may support more efficient induction of meiosis and oogonia development. The multistep process of meiotic entry likely constitutes a fail-safe mechanism to protect against illegitimate activation of this important program (*13*).

It remains unclear how such technology could be used beyond basic research purposes. Possibly, optimized protocols will enable human cells to complete meiosis and oogenesis in vitro. Parallel efforts may also lead to the in vitro formation of sperm-like cells. In vitro–generated egg- and sperm-like cells could prove valuable for generating livestock with particular traits or for rescuing endangered species. The application may be more efficient and safe than the current assisted reproductive cloning approaches, which result in abnormal calves and lambs suffering from large offspring syndrome (*15*).

However, many fundamental issues must be clarified as to whether and when in vitro–derived gamete-like cells should be used to assist human reproduction. First, because the quality of such cells can only be partly addressed through molecular characterizations, the ultimate functional assay for gamete-like cells will be the generation of "offspring." Correct epigenetic programming during gametogenesis is important for embryonic fitness, representing a critical challenge to in vitro gametogenesis. Second, when in vitro–derived eggs and/or sperm are used, does the "product" have the same ethical and legal status as a naturally conceived embryo? Now is the time to set standards for technical, safety, ethical, and legal implications. ■

REFERENCES AND NOTES

1. C. Yamashiro *et al.*, *Science* **362**, 356 (2018).
2. O. Hikabe *et al.*, *Nature* **539**, 299 (2016).
3. K. Sasaki *et al.*, *Cell Stem Cell* **17**, 178 (2015).
4. S. Yokobayashi *et al.*, *Biol. Reprod.* **96**, 1154 (2017).
5. S. I. Nagaoka, M. Saitou, *Cold Spring Harb. Symp. Quant. Biol.* **82**, 213 (2017).
6. T. Kobayashi, M. A. Surani, *Development* **145**, dev150433 (2018).
7. C. G. Extavour, M. Akam, *Development* **130**, 5869 (2003).
8. K. Hayashi *et al.*, *Cell* **146**, 519 (2011).
9. N. Irie *et al.*, *Cell* **160**, 253 (2015).
10. F. von Meyenn, W. Reik, *Cell* **161**, 1248 (2015).
11. A. Inoue *et al.*, *Nature* **547**, 419 (2017).
12. L. Li *et al.*, *Cell Stem Cell* **20**, 858 (2017).
13. S. Yokobayashi *et al.*, *Nature* **495**, 236 (2013).
14. P. W. S. Hill *et al.*, *Nature* **555**, 392 (2018).
15. J. R. Hill, *Annu. Rev. Anim. Biosci.* **2**, 307 (2014).

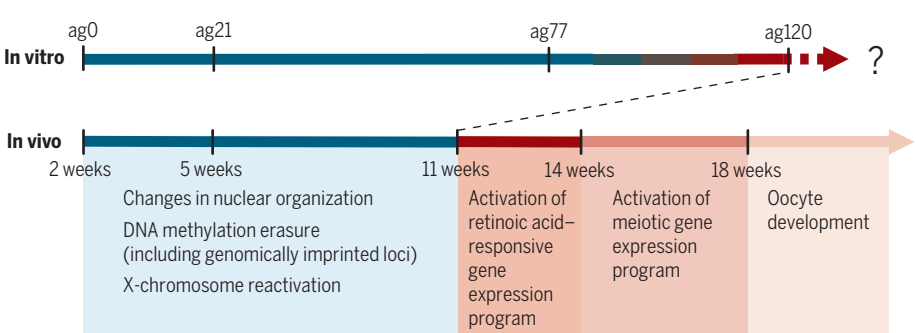
ACKNOWLEDGMENTS

The authors receive funding from the Novartis Research Foundation and the European Research Counsel (ERC 695288–Totipotency).

10.1126/science.aav3479

In vitro germ cell development

Changes in cellular, transcriptional, and epigenetic characteristics during in vivo germ cell development (in weeks of gestation) compared with in vitro germ cell–like development [days after aggregation (ag) of human and mouse cells] (*1*, *12*).



IMMUNOLOGY

A friendly danger

Immune cell receptor DNGR-1 limits inflammatory tissue damage

By **Fabián Salazar** *and* **Gordon D. Brown**

Tissue damage leads to the production of danger signals that regulate inflammatory responses (*1*). These danger signals are detected by an array of cellular receptors on innate immune cells, such as dendritic cells, that enable them to respond to tissue damage appropriately. Dendritic cell NK lectin group receptor-1 (DNGR-1) is a C-type lectin receptor (CLR) that can sense F-actin, which becomes exposed when cells lose plasma membrane integrity (*2*, *3*). On page 351 of this issue, del Fresno *et al.* (*4*) reveal that upon sensing this danger signal, DNGR-1 restricts innate inflammation during both infectious and noninfectious tissue injury, so as to limit excessive immunopathology. Their findings have implications for our understanding of how our immune system regulates responses to tissue damage.

Damage-associated molecular patterns (DAMPs) are self-components that are released or exposed as a result of cell damage or cellular stress. Several classes of DAMPs have been identified, including metabolites, nucleic acids, and proteins, all of which promote inflammation (*5*). Although F-actin functions as a DAMP, it is not directly proinflammatory, and its sensing by the pattern recognition receptor (PRR), DNGR-1, promotes cross-presentation—a process in which dead cell–derived antigens are taken up, processed, and presented through major histocompatibility complex (MHC) class I molecules to cytotoxic T cells to elicit an adaptive immune response (*6*). Unlike other DAMP receptors, expression of DNGR-1 is mostly restricted to one particular immune cell subset, conventional type-1 dendritic cells (cDC1s) in mice and humans (*7*, *8*). Despite its restricted expression, the cross-presentation activity of DNGR-1 is critical for protection against

Medical Research Council Centre for Medical Mycology at the University of Aberdeen, Aberdeen Fungal Group, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK. Email: gordon.brown@abdn.ac.uk

GRAPHIC: N. DESAI/*SCIENCE*

viral infections, including those of herpes simplex and vaccinia virus (*6*, *9*, *10*).

To understand the role of DNGR-1 in innate inflammation, del Fresno *et al.* used a mouse model of sterile inflammation—administering the peptide caerulein to stimulate digestive secretions and induce necrotizing pancreatitis, a disease characterized by cell death, pancreatic edema, and immune cell infiltration. They found that DNGR-1 protected against the development of pancreatitis. Neutrophils, immune cells that normally help combat infection, have a central role in the pathogenesis of this disease, and therapeutic interventions targeting them reduce tissue damage (*11*). Indeed, the protective function of DNGR-1 in these models was due to its ability to restrict neutrophil recruitment into the diseased organ. Despite the previously described role of DNGR-1 in antigen cross-presentation, del Fresno *et al.* provide convincing evidence using a variety of genetically altered mice that this protective response does not involve T (or B) cells and is a direct function of the cDC1s themselves.

Excessive neutrophil recruitment can also cause immunopathology during microbial infection (*12*). Actually, reduction in neutrophil-mediated immunopathology can lead to disease tolerance, a phenomenon in diseases such as tuberculosis whereby the host is able to better resist infection by reducing tissue damage (*13*). To test whether the ability of DNGR-1 to limit neutrophil recruitment also occurred during infection, del Fresno *et al.* used a mouse model of systemic candidiasis in which neutrophil damage in *Candida*-infected kidneys drives considerable immunopathology (*12*). They found that mice lacking DNGR-1 exhibited increased morbidity, mortality, and tissue immunopathology, which correlated with increased neutrophil infiltration. Importantly, tissue damage occurred even after removal of *Candida* with antifungal drugs but could be prevented by removal of the infiltrating neutrophils. These data suggest a crucial role for DNGR-1 in regulating immunopathology after tissue damage, by limiting neutrophil recruitment in both infectious and noninfectious settings. However, it is still unclear whether this mechanism is a universal regulator of neutrophil-mediated immunopathology in all inflammatory contexts.

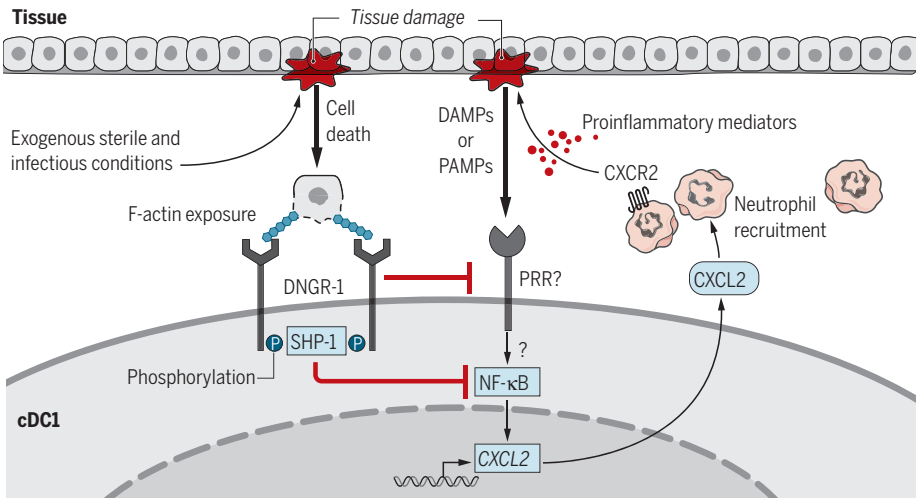
The recognition of DAMPs or pathogen-associated molecular patterns (PAMPs) by PRRs triggers intracellular signaling pathways that normally culminate in either activation or inhibition of cellular functions. Previous research revealed that DNGR-1 binding to F-actin triggered a signaling pathway in cDC1s involving spleen tyrosine kinase (SYK), which promoted cross-presentation (*6*). In the study of del Fresno *et al.*, DNGR-1 was unexpectedly found to induce an inhibitory signaling pathway involving SHP-1 phosphatase and nuclear factor-κB (NF-κB), which repressed inflammatory cytokine production, including a key neutrophil chemoattractant, C-X-C motif chemokine 2 (CXCL2) (see the figure). Indeed, administration of a peptide inhibitor of CXCL2 function (by blocking its receptor CXCR2) reversed the enhanced neutrophil infiltration observed in DNGR-1–deficient mice. However, it is still unclear what drives the inflammatory responses that are being regulated by DNGR-1 after tissue damage in these models. Notably, the functions of DNGR-1 were due to its ability to regulate CXCL2 production in the cDC1s themselves, even though other cells were also producing this chemokine. This result is remarkable because a relatively small subset of cells can exert such a large influence on the overall innate immune response. It is likely, however, that other factors provided by cDC1s also contribute to this effect.

Although there has been considerable progress in our understanding of the mechanisms underlying the proinflammatory responses to tissue damage, much less is known about how these responses are suppressed—knowledge which may have therapeutic application. Given the evolutionary conserved nature of F-actin recognition (*14*), it is likely that DNGR-1 functions in a similar manner in humans, although this remains to be formally established. As many other PRRs, particularly CLRs, are increasingly being found to recognize a variety of endogenous molecules, similar mechanisms involved in controlling damaging immune responses are likely to be more widespread.

Perplexing questions arise from these studies. How can the recognition of F-actin by DNGR-1 drive opposing signals through SYK and SHP-1 and how is this integrated into decisions to regulate innate or adaptive immune responses? Perhaps both happen simultaneously—it is likely that the outcome is influenced by additional factors that remain to be identified (*1*, *15*). DNGR-1 may be acting as a rheostat, responding according to the level of tissue injury and the amount of F-actin exposure. During infection, such mechanisms are thought to regulate tolerance to disease (*13*), and it will be important to determine how or if DNGR-1 influences resistance to other microbes and, particularly, tolerance to commensal microbiota. ■

DNGR-1 controls tissue damage responses

Tissue damage caused by infectious or noninfectious agents induces the production of proinflammatory mediators that recruit neutrophils. DNGR-1 expressed by cDC1 cells recognizes F-actin exposed in dead cells, leading to activation of a SHP-1 pathway. This negatively regulates proinflammatory pathways, which decreases neutrophil infiltration and limits immunopathology.



GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. P. Matzinger, *Science* **296**, 301 (2002).
2. S. Ahrens *et al.*, *Immunity* **36**, 635 (2012).
3. J. G. Zhang *et al.*, *Immunity* **36**, 646 (2012).
4. C. del Fresno *et al.*, *Science* **362**, 351 (2018).
5. E. Vénéreau *et al.*, *Front. Immunol.* **6**, 422 (2015).
6. D. Sancho *et al.*, *Nature* **458**, 899 (2009).
7. L. F. Poulin *et al.*, *Blood* **119**, 6052 (2012).
8. C. Huysamen *et al.*, *J. Biol. Chem.* **283**, 16693 (2008).
9. S. Iborra *et al.*, *J. Clin. Invest.* **122**, 1628 (2012).
10. S. Zelenay *et al.*, *J. Clin. Invest.* **122**, 1615 (2012).
11. Z. W. Yang *et al.*, *J. Cell. Mol. Med.* **19**, 2513 (2015).
12. M. S. Lionakis *et al.*, *PLOS Pathog.* **8**, e1002865 (2012).
13. R. Medzhitov *et al.*, *Science* **335**, 936 (2012).
14. N. Srinivasan *et al.*, *eLife* **5**, e19662 (2016).
15. S. R. Wu *et al.*, *Trends Immunol.* **38**, 231 (2017).

ACKNOWLEDGMENTS

The authors receive funding from the Wellcome Trust, the Medical Research Council Centre for Medical Mycology and the University of Aberdeen, and Arthritis Research UK.

10.1126/science.aav3477

GEOPHYSICS

Earth's soft heart

A modern seismological method raises questions about the properties of Earth's inner core

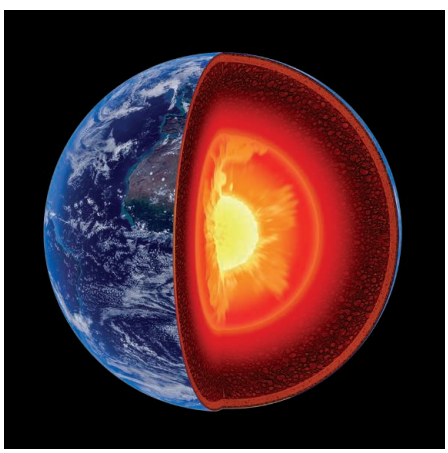
By **Jessica C. E. Irving**

Earth's inner core has proven to be a challenging region for geophysicists to investigate, more than 80 years after its discovery (*1*). It grows slowly from the liquid iron alloy that constitutes the outer core, and its solidification is the result of the planet's cooling over the course of its history. The inner core provides an important part of the energy budget for the geodynamo—the mechanism that generates Earth's magnetic field—as latent heat is released and light elements are preferentially segregated into the fluid outer core. Despite the inner core's importance in the core's energy budget (*2*), its composition and material properties are difficult to ascertain. On page 329 of this issue, Tkalčić and Phạm (*3*) employ a new understanding of method that has been successfully used to study the crust and upper mantle to detect energy that has traveled as shear waves through the inner core. The shear-wave velocity (V_s) reveals a softer inner core composition than previously thought.

Inner core shear bodywaves (called *PKJKP*) provide direct evidence for its solidity, but they have been difficult to detect even after large earthquakes (*4*), making inner core V_s challenging to estimate. The inner core's solidity was therefore established using low-frequency normal mode oscillations of Earth (*5*). Tkalčić and Phạm now obviate the need for imaging *PKJKP* waves directly after a large earthquake. Instead, the authors examined global stacks of Earth's correlation wavefield, calculated from seismograms collected in the hours following large earthquakes, to detect energy that has traveled as shear waves through the inner core.

New theoretical insights (*6*) have explained the genesis of several signals in the seismic correlation wavefield, one of which is akin to *PKJKP* in that it involves interaction between a seismic phase that travels as a shear wave in the inner core with other core phases. This signal in the correlation wavefield is called a *J* phase, and its time delay and amplitude provide constraints on the inner core's V_s and attenuation. Tkalčić and Phạm find that the inner core has a slightly lower V_s than that of the long-standing Preliminary Reference Earth Model (PREM) (*7*) and is more attenuating than in PREM.

This has implications for understanding the composition of the inner core. There is already a challenge in matching the seismologically observed properties of the inner core to values estimated from mineral physics. Pure iron has a V_s that is much higher than that of PREM at inner core conditions (*8*), and many iron alloys suffer from the same velocity discrepancy. The difference between the V_s proposed by Tkalčić and Phạm and that of PREM is smaller than the effect of different alloying light elements on the V_s of iron, but it is clear that every seismic observation requires a V_s that is smaller than expected given the inner core's compressional wave velocity. Mineral physics suggests that premelting phenomena may reduce the V_s to seismological values. Recent work suggests that a ternary mixture of iron, silicon, and carbon in the inner core may reproduce the properties of PREM (*9*). It may also be possible to match the *J* phase observations with a similar composition.



The inner core, Earth's deepest region, may be softer than previously thought.

An alternative to the presence of an alloy with strong premelting effects is the presence of some fraction of melt in the inner core. Melt is often suggested as a mechanism for reducing V_s elsewhere in Earth. This melt may be fluid trapped between solid crystals as the inner core grows. Melt could also increase attenuation of the inner core (*10*), as inferred from the amplitude of the *J* phase's signal. However, Tkalčić and Phạm propose that the inner core has a reduced V_s throughout, requiring that melt pockets have not been eradicated by any form of inner core convection.

Many questions about the shear properties of the inner core remain open. PREM's V_s was strongly informed by the frequency of normal mode oscillations. A detailed assessment of the compatibility of the results of Tkalčić and Phạm with normal mode observations may lead to a better understanding of V_s and other seismic properties in the inner core. At the top of the inner core, V_s can also be inferred from waves reflected at the inner core boundary (ICB). The contrast in seismic properties across the ICB suggests that just below it, V_s is less than 3 km/s (*11*) and may be close to zero in some locations (*12*). Both Tkalčić and Phạm and normal mode studies (*13*) suggest a V_s that is closer to 3.5 km/s, likely reflecting a strong shear wave gradient at the ICB. The *J* phase is likely to be reporting the higher velocity across the bulk of the inner core. The radial and lateral variation of V_s in the inner core remain poorly understood, and there is no firm consensus on shear wave anisotropy in the inner core. The higher attenuation suggested by the amplitude of the *J* phase contrasts with the handful of observations of *PKJKP*, which may suggest that inner core shear wave attenuation is lower than that of PREM (*14*). However, because *PKJKP* detections are rare and may require focusing of seismic waves, it is not unreasonable to expect that the *J* phase results indicate an inner core that is soft, with a low V_s and considerable attenuation of shear waves.

A better understanding of the inner core's shear wave properties can help to ascertain dynamical processes including what, if any, kind of convection is taking place in the inner core and whether it is rotating relative to the mantle. The observation of *J* phases will provide an extra tool to assess the properties of Earth's "soft heart." ■

Department of Geosciences, Princeton University, Princeton, NJ, USA. Email: jirving@princeton.edu

REFERENCES AND NOTES

1. I. Lehmann, *Publ. Bur. Cent. Seism. Int. A* **14**, 87 (1936).
2. P. Olson, *Science* **342**, 431 (2013).
3. H. Tkalčić, T.-S. Phạm, *Science* **362**, 329 (2018).
4. P. M. Shearer *et al.*, *Geophys. J. Int.* **185**, 1379 (2011).
5. A. M. Dziewonski, F. Gilbert, *Nature* **234**, 465 (1971).
6. T.-S. Phạm *et al.*, *Geophys. Res. Lett.* **45**, 3035 (2018).
7. A. M. Dziewonski, D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981).
8. L. Vočadlo, in *Treatise on Geophysics*, G. Schubert, Ed. (Elsevier, ed. 2, 2015), vol. 2, pp. 117–147.
9. Y. Li *et al.*, *Earth Planet. Sci. Lett.* **493**, 118 (2018).
10. S. C. Singh *et al.*, *Science* **287**, 2471 (2000).
11. K. D. Koper, M. Dombrovskaya, *Earth Planet. Sci. Lett.* **237**, 680 (2005).
12. V. F. Cormier, *Geophys. Res. Lett.* **42**, 7459 (2015).
13. A. Deuss, *Earth Planet. Sci. Lett.* **268**, 364 (2008).
14. A. Cao, B. Romanowicz, *Geophys. Res. Lett.* **36**, L09301 (2009).

ACKNOWLEDGMENTS

Supported by NSF grants EAR1644399 and EAR1736046.

10.1126/science.aav2296

BOOKS *et al.*

In Robinson's imagined future, China has colonized the Moon but is mired in earthly drama.

PHOTO: MARK RALSTON/AFP/GETTY IMAGES

SCIENCE FICTION

The familiar politics of the final frontier

Earth's economic woes haunt the first lunar colony in a fictional near future

By **Josh Trapani**

On Kim Stanley Robinson's Moon, gibbons soar like flying squirrels. A billionaire transforms an immense lava tube into a fantasy of classical China. Old friends view a majestic earthrise.

It's the year 2047, and the Moon has been colonized by the world's sole superpower: China. American Fred Fredericks makes his first lunar voyage to deliver a secure communications device to Chang Yazu, chief administrator of the Chinese Lunar Authority. But when they shake hands, Chang drops dead. He's been poisoned, and Fred is accused of murder.

Fred's fate quickly becomes intertwined with that of Chan Qi, the pregnant daughter of a top Chinese official. Qi is herself on the run, her destiny tied in with a nascent global revolution. Together, Fred and Qi travel back and forth between the Moon and China to evade the mysterious forces seeking to capture them.

Red Moon
Kim Stanley Robinson
Orbit, 2018. 464 pp.

Robinson's novels are known for detailed visions of the future and deep supporting research. This pattern holds for *Red Moon*, but although the lunar colony is interesting, it's China that steals the show. Still under the rule of the Communist Party, China in 2047 is a nation of contrasts: cleaner and greener than today's China, while also more urbanized. It is a surveillance state, but one balkanized and weakened by infighting agencies. This future China is an economic powerhouse and also a land of great inequality: One-third of its population suffers discrimination under the *hukou* system, which keeps people tied to the place they were born. Those from poor rural areas face a choice: an impoverished life or existence as a migrant laborer, without health care or protection from abuse.

As Fred flees with Qi, he comes to understand her importance as a leader of these rootless workers. Mass demonstrations in Beijing, ordered by Qi while she and Fred hide on the Moon, quickly spread to the United States and other nations—lesser powers inextricably linked to China through the global financial system—and snowball into a worldwide people's revolt.

If this sounds less like science fiction than political thriller, consider that *The New Yorker* labeled Robinson "one of the most important political writers working in America today" (*1*). His previous work has tackled climate change, artificial intelligence, and economic and governance systems, topics woven into this novel, too. That the market dominates life regardless of whether the system of government is Chinese communism or American democracy is a central underlying theme.

Red Moon is political on multiple levels: from the characters' motivations, to interagency rivalries and international sparring, to—ultimately—global rebellion. But such a focus comes with a downside: The novel lavishes attention on concepts and ideas, while characters and relationships are neglected.

Fred, for instance, is passive and asocial. More than once, I wondered whether he might be on the autism spectrum. But this element of his persona, like so much else about him, remains unexplained and largely unexplored. Qi is unremittingly grouchy. It's unclear how she went from daughter of the elite to leader of the dispossessed, and her pregnancy serves more as a plot device than as a clue to her nature. (Robinson's description of a baby in a space suit, however, forms another haunting image in the book.)

Despite spending much of the story in close quarters, Qi and Fred fail to develop any chemistry. Along with most of the book's other characters, they engage in a great deal of expositional dialogue for seemingly no other purpose than to explore the themes of true interest to the author.

Robinson can draw three-dimensional characters when he chooses to. In *Red Moon*, the most personable by far is Ta Shu, a feng shui–obsessed travel show host whose humor and vulnerability shine through in his love of poetry and his grief over the death of his mother. I wish that Fred and Qi had been given similar depth.

Despite its largely forgettable characters, *Red Moon* portrays striking scenes amid a compelling vision of the future. Then again, perhaps that juxtaposition is part of the novel's lesson: In a churning world of unhappy billions, with growing inequality along with ever more technology and surveillance, it's important not to forget that our individuality is also the source of our shared humanity. ■

The reviewer is an analyst and writer based in Washington, DC, USA. Email: jtrapani@hushmail.com

REFERENCE

1. T. Kreider, *The New Yorker*, 12 December 2013; www.newyorker.com/books/page-turner/our-greatest-political-novelist.

10.1126/science.aav1050

SCIENCE & SOCIETY

Recognizing the role of skilled immigration

A scholar examines the rewards and shortcomings of global talent in the United States

By **Kevin Shih**

What is the world's most precious resource? It can't be held, it can't be mined—it's even hard to clearly define. It exists in every country in the world but also continually moves across borders. Although it can bring wealth and economic prosperity, not a single war has been fought over it, and some nations even take steps to block its entry. The answer is human talent.

William Kerr's *The Gift of Global Talent* reveals that much of America's prosperity can be traced to its long history as a leading destination for skilled immigrants. More than half of college-educated immigrants to nations in the Organisation for Economic Co-operation and Development (OECD) end up in the United States, and immigrant contributions to American science, invention, and entrepreneurship have been remarkable. More than a third of the country's Nobel Prize winners, 40% of the founders of its Fortune 500 companies, and half of its residents with science and engineering doctorates hail from overseas.

However, the key to success isn't as simple as opening borders. Kerr elucidates two important principles that underlie the particular talent necessary to foster innovation and economic growth. First, the influx of talent must create goods and services that can reach customers worldwide. Second, foreign-born talent must be enhanced by the presence of other talented individuals. The clustering of tech entrepreneurs in Silicon Valley, for example, helps enhance everyone's productivity as complex ideas and tacit knowledge are shared and readily accessible. It is this type of talent that the United States has succeeded in attracting from abroad, more effectively than any other nation.

Foreign-born talent in the United States mostly takes the form of scientists and engineers. This is partly because science and engineering knowledge is globally portable and is less inhibited by language or cultural barriers than other forms of knowledge, but it is also due to immigration policy.

Kerr provides an insightful contrast of the strong role played by employers in selecting talent within the H-1B program, comparing it with the admissions policies in other nations. Canada's "points-based" system, for example, awards entry on the basis of broadly defined skills and provides no role for employers; whereas the employer-dominant H-1B visa system provides incentives for firms to invest in finding talent and a clear pathway to jobs for immigrants but is far less transparent than many other systems.

While Kerr spends much time detailing why and how skilled immigrants have benefited the United States, several recent cases paint a different picture. In one such case, the Disney company required American internet technology workers to train their H-1B replacements before being laid off (*1*).

America has long been a desirable destination for skilled immigrants.

Kerr acknowledges that the gains from skilled immigration may also be accompanied by negative consequences for native-born workers and provides a comprehensive but succinct overview of the academic research on this topic. The consensus appears to be that the positive effects of skilled immigrants on American job prospects appear over long horizons, whereas in the short term, the H-1B program may cause worker displacement. In addition, whereas young Americans may benefit from the increased talent from abroad, older workers may never see such gains.

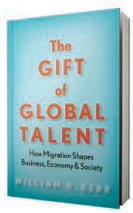
Although not explicitly addressed, certain features of U.S. institutions—such as good governance, political stability, and strong intellectual property laws—are also necessary for skilled immigrants to flourish and push innovative boundaries. Other nations hoping to replicate the U.S. experience by attracting global talent may end up on an entirely different path. Readers wishing for a more global perspective on the gifts and consequences of global talent must look elsewhere: This book maintains a U.S.-centric perspective.

Kerr is masterful at distilling the key findings of hundreds of complex academic studies in just a few words. He tells his story using anecdotes, case studies, and analogies—a tooth fairy payment schedule for his young son, for example, becomes an analogy for exponential technological growth—and his delivery is humorous and captivating.

Kerr concludes with important suggestions for skilled immigration policy that should be required reading for every U.S. legislator. For example, he suggests that lengthening the application period and indexing the number of H-1B visas to an economic index rather than a hard cap can help mitigate bottlenecks and incentivize firms to seek out truly exceptional talent. Redistribution from those who gain the most can protect the American workers who might be displaced.

Despite the gains we've achieved from skilled immigration to the United States, current anti-immigrant sentiment threatens to stop inflows of global talent. There are steps we can and should take to protect American workers that do not require walling ourselves off from the benefits of global talent. ■

The Gift of Global Talent
How Migration Shapes Business, Economy & Society
William R. Kerr
Stanford University Press, 2018. 256 pp.



The reviewer is at the Department of Economics, Queens College, City University of New York, Queens, NY 11367, USA. Email: kevin.shih@qc.cuny.edu

REFERENCE

1. Tribune news services, *Chicago Tribune*, 14 December 2016; www.chicagotribune.com/business/ct-disney-workers-discrimination-lawsuit-20161214-story.html.

10.1126/science.aav2430

LETTERS

The question of whether we should protect endangered species like the woodland caribou is not strictly utilitarian.

PHOTO: GALLO IMAGES/GETTY IMAGES

Edited by **Jennifer Sills**

Ethics and politics of conservation triage

In his News Feature "Should it be saved?" (7 September, p. 962), W. Cornwall uses efforts to save Canada's dwindling woodland caribou populations to illustrate growing acceptance that conservation triage may be needed to achieve efficient allocation of scarce conservation funding. Although Cornwall covers several aspects of the controversial triage debate, he fails to convey the full breadth of its risks.

First, triage is a reactive strategy that applies efficiency criteria to the inadequate resources devoted to conservation (see the Feature's inset, "How triage became a dirty word," p. 965). As such, it represents a last-ditch response to conservation crises that should not have been allowed to arise in the first place. Triage also appears to embrace a rigid utilitarianism, deploying cost-benefit analysis to balance dollar costs against conservation benefits that, in many cases, defy monetary valuation.

This focus on dollar efficiency obscures the serious moral and ethical implications of deciding that we will "allow" certain species to become extinct. The rhetoric of "successes," "failures," and "winners" deployed by some of triage's promoters serves to further undermine these ethical considerations (*1*). Moreover, by imposing a veneer of objectivity over decisions that are inherently messy and value-laden, triage provides cover for self-interested politicians to continue the malign neglect that characterizes conservation policy in many jurisdictions.

Precipitous population declines in the mountain ecotype of the woodland caribou are a direct consequence of decades of such neglect, even as cumulative impacts of resource extraction and recreation have continued to accumulate (*2*). The situation for woodland caribou elsewhere in Canada is not much better, with multiple jurisdictions suspending conservation plans, grandfathering resource extraction, or simply deciding to do nothing (*3*).

Woodland caribou represent the tip of a legislative iceberg of delayed recovery plans, failure to designate critical habitat, and general non-enforcement of Canada's Species at Risk Act (*4*, *5*). Non-enforcement of existing legislation has contributed to population declines or stagnation in most of Canada's at-risk species (*6*). Regarding triage as inevitable, as some researchers have done (*7*), will be interpreted in some quarters as acquiescence to this unacceptable status quo. Such an attitude guarantees that many threatened species will soon be in need of triage themselves when they become endangered in the future.

Andrew Park
Department of Biology, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada.
Email: a.park@uwinnipeg.ca

REFERENCES

1. M. Hebblewhite, *Biol. Conserv.* **206**, 102 (2017).
2. B. Payton, *Canadian Geographic* **May/June 2003**, 54 (2003).
3. A. Pope, "Critical habitat for boreal woodland caribou still unprotected six months after deadline," *Canadian Geographic* (2018); www.canadiangeographic.ca/article/critical-habitat-boreal-woodland-caribou-still-unprotected-six-months-after-deadline.
4. S. C. Bird, K. E. Hodges, *Environ. Sci. Pol.* **71**, 1 (2017).
5. B. Cheadle, "Species At Risk Act: Court finds 'enormous systemic problem' in enforcement," *Huffpost* (2014); www.huffingtonpost.ca/2014/02/14/species-at-risk-act-canada_n_4790196.html.
6. B. Favaro *et al.*, *PLOS One* **9**, e113118 (2014).
7. M. C. Bottrill *et al.*, *Trends Ecol. Evol.* **24**, 183 (2009).

10.1126/science.aav4382

New NSF policy will stifle innovation

As F. Schembri explains in her News story "Biologists irate at NSF's new one-proposal cap" (28 September, https://scim.ag/NSFcap), a new proposal submission policy announced on 23 August by the U.S. National Science Foundation (NSF) Biology Directorate now mandates that investigators can serve as principal investigator (PI) or co-PI on only one proposal per fiscal year to each of the core tracks of the Divisions of Environmental Biology, Integrative Organismal Systems, and Molecular and Cellular Biosciences. The divisions' new program solicitations (*1*–*3*) highlight changes previewed in October 2017 (*4*), including the elimination of submission deadlines and preproposals. However, the restriction on proposal number was not previously announced nor made available for comment.

The rationale for the restriction appeared in a FAQ posted online 3 weeks after the policy announcement (*5*): "Like every other program at NSF that has transitioned to a no-deadline proposal submission process, BIO [Directorate of Biological Sciences] has imposed a limitation on proposal submissions. The goal is to prevent the immediate resubmission of declined proposals...." However, other NSF programs without deadlines prohibit resubmission of a proposal within a year of the original submission date (*6*, *7*), limit PIs to two proposals under consideration for funding at one time (*8*), state that PIs can submit an unlimited number of proposals given that "each proposal must be significantly different from the others"

(*9*), or explicitly state that the number of proposals a PI may submit is unlimited (*8*, *10*). BIO appears to have created a new, more restrictive policy for submission rather than following precedent.

We are deeply concerned that limiting PIs to a single proposal per fiscal year damages biological research. Science is an increasingly collaborative endeavor; many important breakthroughs come from groups of researchers with complementary expertise. For decades, NSF has actively encouraged multi-PI collaboration, but researchers are now forced to limit the diversity of their research. Under the new guidelines, a scientist can serve as "senior personnel" on an unlimited number of proposals each year, but this status does not reflect the time and expertise that scientists bring to collaborations. Furthermore, reviewers, employers, and the community may not see the contribution of senior personnel as on par with co-PIs nor recognize senior personnel as having made substantial intellectual contributions to the work.

An additional serious concern is that early-career investigators, who need to demonstrate their scientific innovation and independence through grant funding, will be particularly disadvantaged. Given the small minority of proposals that are funded, this new limit—a single proposal per year to core programs—will exacerbate pressure on early-career colleagues because they cannot spread their risk of not being funded by submitting collaborative proposals. New researchers may turn to safe, fundable submissions rather than novel and transformative ideas.

We hope that the NSF will reconsider the restriction as a major impediment to an otherwise refreshing change to its current programs.

Heather L. Eisthen,[1]* Kenneth M. Halanych,[2] Darcy B. Kelley,[3] Stephanie A. White,[4] Steven M. Phelps,[5] and 66 additional authors†

[1]Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA. [2]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA. [3]Department of Biological Sciences, Columbia University, New York, NY 10027, USA. [4]Department of Integrative Biology and Physiology, University of California, Los Angeles, CA 90095, USA. [5]Department of Integrative Biology, University of Texas, Austin, TX 78712, USA.
*Corresponding author. Email: eisthen@msu.edu
†The full list of authors is available online.

REFERENCES

1. NSF, "Division of Environmental Biology (core programs) (DEB) program solicitation" (Publication 18-587, NSF, 2018); www.nsf.gov/pubs/2018/nsf18587/nsf18587.htm.
2. NSF, "Division of Integrative Organismal Systems core programs program solicitation" (Publication 18-586, NSF, 2018); www.nsf.gov/pubs/2018/nsf18586/nsf18586.htm.
3. NSF, "Division of Molecular and Cellular Biosciences: Investigator-initiated research projects (MCB) program solicitation" (Publication 18-585, NSF, 2018); www.nsf.gov/pubs/2018/nsf18585/nsf18585.htm.
4. NSF, "Dear colleague letter: Implementation of 'no-deadline,' full-proposal submission process for most programs in the Directorate for Biological Sciences" (Publication 18-011, NSF, 2017); www.nsf.gov/pubs/2018/nsf18011/nsf18011.jsp.
5. NSF, "Frequently asked questions (FAQs) for Directorate of Biological Sciences (BIO) core programs solicitations" (Publication 18-106, NSF, 2018); www.nsf.gov/pubs/2018/nsf18106/nsf18106.jsp.
6. NSF, "Solar and Planetary Research Grants (SPG) program solicitation" (Publication 16-602, NSF, 2016); www.nsf.gov/pubs/2016/nsf16602/nsf16602.htm.
7. NSF, "Geomorphology and Land Use Dynamics program solicitation" (Publication 15-560, NSF, 2015); www.nsf.gov/pubs/2015/nsf15560/nsf15560.htm.
8. NSF, "Arctic research opportunities: Arctic natural sciences; Arctic social sciences; Arctic system science; Arctic observing network program solicitation" (Publication 16-595, NSF, 2016); www.nsf.gov/pubs/2016/nsf16595/nsf16595.htm.
9. NSF, "Frequently asked questions (FAQs) for dear colleague letter (DCL) NSF 17-117: Removal of deadlines for the combustion and fire systems program in ENG/CBET" (Publication 17-118, NSF, 2017); www.nsf.gov/pubs/2017/nsf17118/nsf17118.jsp.
10. NSF, "Frequently asked questions (FAQs) for NSF 17-086, removal of deadlines for the Marine Geology and Geophysics Program in the Division of Ocean Sciences in the Directorate for Geosciences" (Publication 17-087, NSF, 2017); www.nsf.gov/pubs/2017/nsf17087/nsf17087.jsp.

SUPPLEMENTARY MATERIALS

Full list of authors
www.sciencemag.org/content/362/6412/297.2/suppl/DC1

10.1126/science.aav4793

China's fight against soil pollution

Soil pollution caused by human activities in China threatens agricultural productivity, food safety, and human health (*1*, *2*). The Chinese government has taken a series of measures, including the Soil Contamination Prevention and Control Action Plan (*2*), to combat soil pollution (*3*). Recently, the country took another positive step by passing the Law on Soil Pollution Prevention and Control, legislation that was 12 years in the making (*4*).

The new law will help to promote soil conservation in China in several ways (*5*). First, the law clarifies who is responsible for soil pollution and provides a framework for holding those people or entities accountable. Second, targets for soil pollution levels will be put in place, and regular evaluations will determine whether local governments are meeting the targets. Third, China will conduct a nationwide soil condition census at least once every 10 years and will set up several national soil environmental monitoring stations that will share relevant information in real time. Fourth, a list of toxic and hazardous substances will be published. Governments at all levels will use the list to establish management regulations to prevent soil pollution. Finally, China will increase investment in soil pollution control and establish a funding system for soil pollution prevention.

Despite these improvements to soil pollution policies, there is still a long way to go. First, China should strengthen its soil pollution prevention and control by streamlining soil, air, and water pollution efforts through consistent legislation. China should also explicitly state its ecological protection and environmental quality goals, as well as the limits to resource use necessary to meet those goals. In addition, China should set the phased targets of soil pollution prevention to meet the strategic requirements of its ecological civilization construction and rural revitalization. Finally, China's soil pollution has multiple sources, and the type and degree of pollution vary across regions. China must tailor its strategies to account for these differences.

Yang Zhou and Yansui Liu*
Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.
*Corresponding author. Email: liuys@igsnrr.ac.cn

REFERENCES

1. N. Rodríguez-Eugenio, M. McLaughlin, D. Pennock, "Soil pollution: A hidden reality" (Food and Agriculture Organization, 2018).
2. J. Wang, Q. Hu, X. Wang, X. Li, X. J. Yang, *Science* **354**, 562 (2016).
3. R. Chen, A. De Sherbinin, C. Ye, G. Shi, *Science* **344**, 691 (2014).
4. Y. Wei, "The land defense campaign introduces the Law on Soil Pollution Prevention and Control and sets up funding for soil prevention," *The 21st Century Business Herald* (2018); http://epaper.21jingji.com/html/2018-09/06/content_93122.htm [in Chinese].
5. Xinhuanet, "The Law on Soil Pollution Prevention and Control of the P.R. China," Xinhua.net (2018); www.xinhuanet.com/2018-09/01/c_129944871.htm [in Chinese].

10.1126/science.aav4061

TECHNICAL COMMENT ABSTRACTS

Response to Comment on "The plateau of human mortality: Demography of longevity pioneers"

Elisabetta Barbi, Francesco Lagona, Marco Marsili, James W. Vaupel, Kenneth W. Wachter

Beltrán-Sánchez *et al.* based their comments on misleading calculations of the maximum survival age. With realistic numbers of people attaining age 105 and the estimated plateau, the Jeanne Calment record is indeed plausible.

Full text: dx.doi.org/10.1126/science.aav3229

Cite as: E. Barbi *et al.*, *Science* 10.1126/science.aav3229 (2018).

Response to Comment on "The plateau of human mortality: Demography of longevity pioneers"

Elisabetta Barbi[1]*, Francesco Lagona[2], Marco Marsili[3], James W. Vaupel[4,5], Kenneth W. Wachter[6]

[1]Department of Statistics, Sapienza University of Rome, Rome, Italy. [2]Department of Political Sciences, University of Roma Tre, Rome, Italy. [3]Italian National Institute of Statistics (ISTAT), Rome, Italy. [4]Interdisciplinary Centre on Population Change, University of Southern Denmark, Odense, Denmark. [5]Duke University Population Research Institute, Durham, NC, USA. [6]Department of Demography, University of California, Berkeley, CA, USA.

*Corresponding author. Email: elisabetta.barbi@uniroma1.it

Beltrán-Sánchez *et al.* based their comment on misleading calculations of the maximum survival age. With realistic numbers of people attaining age 105 and the estimated plateau, the Jeanne Calment record is indeed plausible.

Beltrán-Sánchez *et al.* (*1*) based their comment on calculations for their quantity T_{max}, projected for cohorts with n members alive at age 105. The authors call T_{max} "maximum survival age," but that is misleading. T_{max} is not an age with low probability of being exceeded. In the model in use, the probability of T_{max} being exceeded is more than 63%. The lifespan of the last survivor out of n is a random quantity. T_{max} is the mode of a distribution with a long upper tail; it is a low-side estimate, not any kind of upper bound.

Specifically, the probability that at least 1 of n further lifespans exceeds x is $1 - [1 - \exp(-mx)]^n$. With $x = T_{max} = (1/m) \log(n)$, the probability is $1 - [1 - (1/n)]^n > 1 - (1/e) = 0.632$. That is, the probability is greater than 63%. To keep chances of longer lifespans below 10%, all ages in the upper part of table 1 of Beltrán-Sánchez *et al.* would need to be shifted upward by about 3.5 years, putting all the bounds for females above 121.

Under the model, percentiles of the distribution of maximum age do depend logarithmically on n. But n grows over time and may well be growing exponentially as cohorts accumulate, cohort sizes grow, and survival to age 105 improves.

Beltrán-Sánchez *et al.* mistakenly take the sample sizes 463 and 3373 from our paper [table 1 of (*2*)], which are period counts at all ages above 105 over 7 years, as if they were values of n, numbers reaching age 105 in a birth cohort. (Also, the upper part of their table 1 uses $m = 0.629$, taken from our table 2, for their calculation for females, not 0.475 as stated in their first paragraph, and not 0.645, the correct value from table 2 of our paper.)

Proper specification of n over time requires study of empirical data. With realistic trends in n and mortality plateaus such as we estimated, Lenart *et al.* (*3*) find that the lifespan of Jeanne Calment is not implausible. It is not surprising that her record still holds; although her record might not be broken for decades, nothing that could be called a limit to lifespan is in sight.

In agreement with us, Beltrán-Sánchez *et al.* are not claiming that there is a fixed limit to lifespan. Their model implies maximum ages increasing into the future at slow rates, just as we expect. Along with advances in survival at younger ages in growing populations, two considerations not featured might make increases in maximum ages faster, although still slow: Hazards beyond age 115 might be decreasing with age, as seen at extreme ages in Mediterranean fruit flies and other model organisms. The cohort trend is toward slightly improving hazards beyond age 105, which we report might be sustained.

Contrast what we are seeing at extreme ages with what a limit to lifespan could look like. It is possible to imagine evolutionary models in which genetic load unrestrained by natural selection at old ages drives hazards to infinity at some finite age in a "Wall of Death" (*4*). That turns out not to be true. It is possible to imagine the processes that are driving up aggregate cohort hazards with age faster and faster through the 50s, 60s, and 70s continuing unabated at extreme ages, reducing chances of survival by further orders of magnitude. That also turns out not to be true.

It is important not to lose sight of the substantive lesson of our study and others: The evolutionary, physiological, and environmental influences that govern lifespans are flexible, not rigid. There is plasticity to senescence. As the study of extreme longevity helps us to understand how evolution can allow flexibility in survival chances, we may be better able to bring genomic knowledge to bear on improving health and survival at the more modest ages to which more of us can aspire.

REFERENCES

1. H. Beltrán-Sánchez, S. N. Austad, C. E. Finch, Comment on "The plateau of human

mortality: Demography of longevity pioneers". *Science* **361**, eaav1200 (2018).

2. E. Barbi, F. Lagona, M. Marsili, J. W. Vaupel, K. W. Wachter, The plateau of human mortality: Demography of longevity pioneers. *Science* **360**, 1459–1461 (2018). doi:10.1126/science.aat3119 Medline

3. A. Lenart, J. M. Aburto, A. Stockmarr, J. W. Vaupel, The human longevity record may hold for decades. arXiv:1809.03732 [q-bio.PE] (11 September 2018).

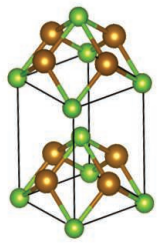
4. K. W. Wachter, S. N. Evans, D. Steinsaltz, The age-specific force of natural selection and biodemographic walls of death. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10141–10146 (2013). doi:10.1073/pnas.1306656110 Medline

19 September 2018; accepted 26 September 2018
Published online 19 October 2018
10.1126/science.aav3229

RESEARCH

Observing Majorana bound states in an iron-based superconductor

Wang et al., *p. 333*



IN *SCIENCE* JOURNALS

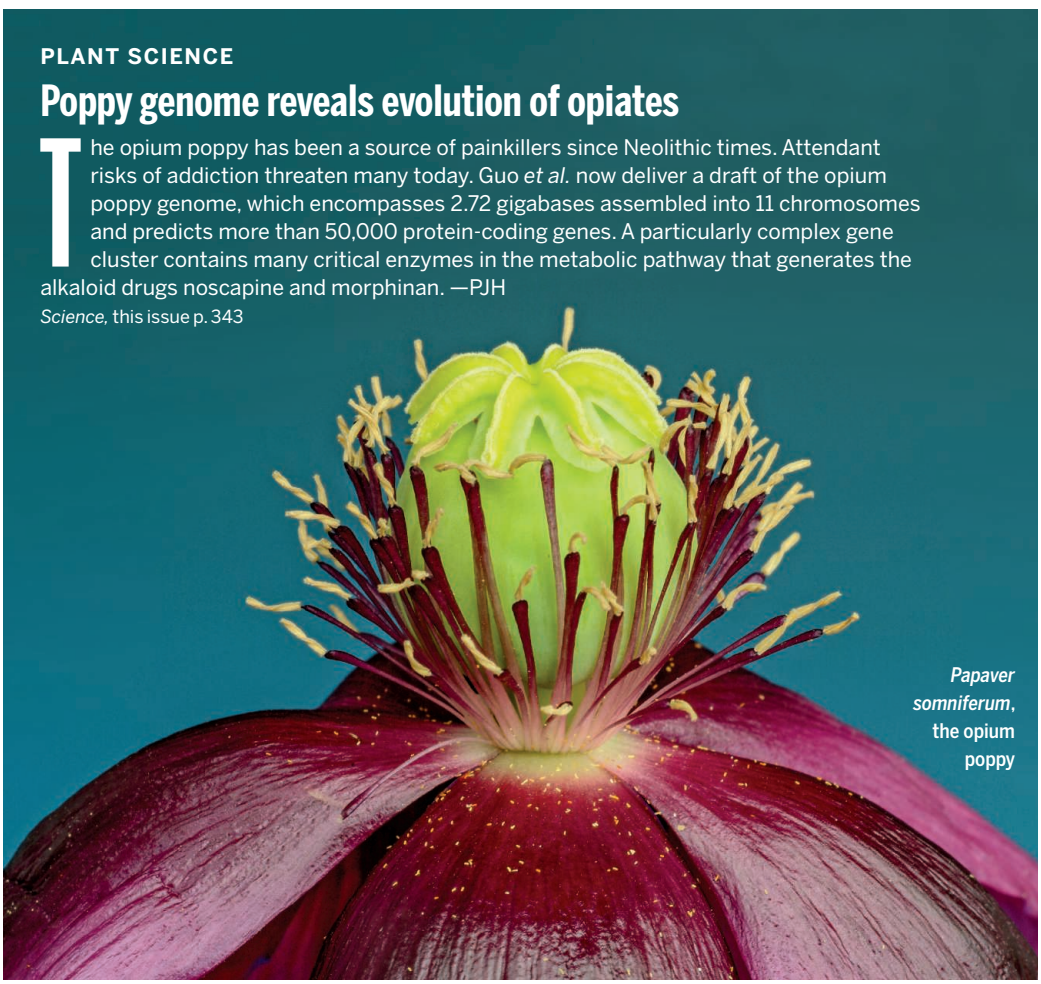
Edited by **Stella Hurtley**

PLANT SCIENCE

Poppy genome reveals evolution of opiates

The opium poppy has been a source of painkillers since Neolithic times. Attendant risks of addiction threaten many today. Guo *et al.* now deliver a draft of the opium poppy genome, which encompasses 2.72 gigabases assembled into 11 chromosomes and predicts more than 50,000 protein-coding genes. A particularly complex gene cluster contains many critical enzymes in the metabolic pathway that generates the alkaloid drugs noscapine and morphinan. —PJH

Science, this issue p. 343



Papaver somniferum, the opium poppy

GRAPHENE

Thickness matters in graphene stacks

If you stack graphene monolayers on top of each other, the number of layers will affect the properties of the material. Intuitively, one would expect that as the stack becomes thicker, the results will converge as the sample starts to resemble graphite. Nam *et al.* measured the conductance of graphene multilayers of increasing thickness. Studying samples up to seven layers thick, they found that in all of them, electronic correlations caused a phase transition at a nonzero critical temperature. However, the critical temperature, as well as the nature of the low-temperature state, depended strongly on the number of layers. This unexpectedly persistent dependence showed no signs of slowing down and will motivate further theoretical and experimental work. —JS

Science, this issue p. 324

GEOPHYSICS

A solid and squishy inner core

Earth's inner core is thought to be solid, which means it should support shear waves. However, the small size of the inner core makes detecting shear waves very difficult. Tkalčić and Pham correlated different types of seismic phases to finally determine the speed of shear waves in Earth's inner core (see the Perspective by Irving). The detection of the waves closes an 80-year quest to find them and confirms a solid, but soft, inner core. —BG

Science, this issue p. 329; see also p. 294

SCIENCE AND SOCIETY

Sharing pharmaceutical research

Increased collaboration will enhance our ability to predict new therapeutic drug candidates. Such data sharing is currently limited by concerns about intellectual property and competing commercial interests. Hie *et al.* introduce an end-to-end pipeline, using modern cryptographic tools, for secure pharmacological collaboration. Multiple entities can thus securely combine their private datasets to collectively obtain more accurate predictions of new drug-target interactions. The computational pipeline is practical, producing results with improved accuracy in a few days over a wide area network on a real dataset with more than a million interactions. —BJ and AMS

Science, this issue p. 347

IMMUNOLOGY

The absence of DNGR-1 is dangerous

Conventional type 1 dendritic cells (cDC1s) can sense tissue damage via DNGR-1, which binds F-actin exposed by necrotic cells. DNGR-1 activation favors cross-presentation, the process by which extracellular antigens are processed and presented to $CD8^+$ T cells via major histocompatibility complex class I molecules. Del Fresno *et al.* studied mice lacking DNGR-1

Published by AAAS

and found that DNGR-1 also has anti-inflammatory effects (see the Perspective by Salazar and Brown). It inhibits the secretion of the chemokine CXCL2 by cDC1s, which, in turn, limits neutrophil recruitment. Thus, DNGR-1 connects cell-death sensing with a mechanism of damage control. —STS

Science, this issue p. 351; see also p. 292

MALARIA

Prenatal *Plasmodium* reactivity

Fetal immunity is generally thought to be skewed toward tolerance. Odorizzi *et al*. used samples from a study in Uganda to determine if placental malaria infection modulated fetal immune responses to malaria. They stimulated cord blood cells in vitro and found that the fetal cells from cases of placental malaria were more reactive to *Plasmodium* antigens. Moreover, clinical follow-up revealed that this increased T cell response correlated with protection from childhood malaria. Thus, protective immune responses in humans can develop even before birth. —LP

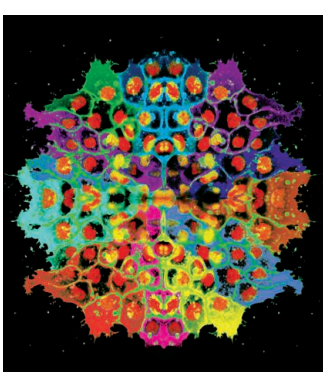
Sci. Transl. Med. **10**, eaat6176 (2018).

NEURODEVELOPMENT

Supracellular cable drives collective cell movement

Neural crest cells migrate far and wide through a vertebrate embryo during development. Shellard *et al*. used *Xenopus* and zebrafish embryos to study how these clumps of mesenchymal cells migrate (see the Perspective by Adameyko). Movement was powered by a supracellular actomyosin cable that contracted around the rear of the clump. Similar supracellular contractility at the front was inhibited by a chemotactic signal. The imbalance in forces caused cells to rearrange so that the whole clump would be propelled forward. —PJH

Science, this issue p. 339; see also p. 290



Supracellular organization during collective migration

CREDITS: (FROM LEFT) SHELLARD *ET AL.*; NASA/JPL-CALTECH/R. HURT (SSC/CALTECH)

PAIN

Channeling metastasis pain with VEGF

Metastatic cancer in the bone is painful. Yang *et al*. found that vascular endothelial growth factor (VEGF) promotes tumor angiogenesis and also contributes to this pain. In a rat model of bone-metastatic breast cancer, tumor-secreted VEGF repressed the expression of the potassium (K^+) channel TRESK. Loss of K^+ current through TRESK increased the excitability of sensory neurons and made the animals hypersensitive to heat and touch near the bone lesion. Blocking this pathway restored channel activity and alleviated pain. —LKF

Sci. Signal. **11**, eaao5150 (2018).

OPTICAL METAMATERIALS

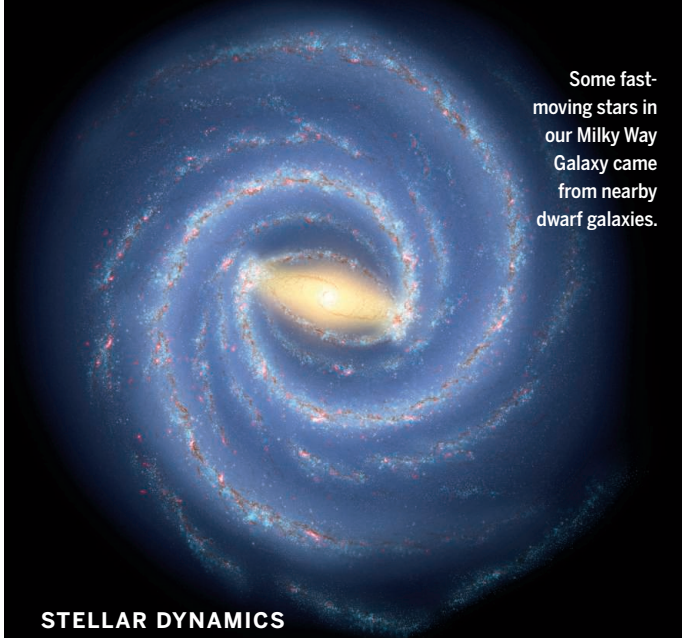
Painting on the cool

Passive radiative cooling materials emit heat. They can reduce the need for air conditioning by providing daytime cooling but are often challenging to apply to rooftops and other building surfaces. Mandal *et al.* fabricated porous poly(vinylidene fluoride-co-hexafluoropropene) to create an excellent radiative cooling material. Better yet, the polymer is easy to paint or spray onto a wide range of surfaces, has good durability, and can even be dyed. This makes it a promising candidate for widespread use as a high-performance passive radiative cooling material. —BG

Science, this issue p. 315

IN OTHER JOURNALS

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



Some fast-moving stars in our Milky Way Galaxy came from nearby dwarf galaxies.

STELLAR DYNAMICS

Fast-moving stars from other galaxies

Dynamical interactions and supernovae can accelerate stars to high velocities, sometimes even fast enough that they are no longer gravitationally bound to their host galaxy and escape from it. Marchetti *et al*. have combined astrometric data with radial velocity measurements to determine the three-dimensional motions of 7 million stars within the Milky Way Galaxy. Within that sample, they identify 20 stars that are not bound to the Galaxy. Only seven of them are moving away from the Milky Way's disc; 13 stars originated elsewhere. The authors postulate that these apparently extragalactic stars may have been ejected or tidally stripped from nearby dwarf galaxies. —KTS

Mon. Not. R. Astron. Soc. 10.1093/mnras/sty2592 (2018).

VIRAL HOST RESPONSE

Stressed out by influenza virus

Viral infection leads to cellular stress. This can act to the host's advantage to curb infection; however, the virus can also subvert stress responses for its own gain. Zhao *et al*. found that influenza A virus (IAV) infection leads to global deregulation of transcription. After IAV infection, RNA polymerase II runs through the transcription termination site of almost all active genes. This down-regulates gene expression by affecting the splicing of some transcripts and delaying the next rounds of transcription initiation. The viral protein NS1 is responsible for this effect and can be modulated by SUMOylation. Although whether this stress response tilts in favor of host survival or facilitates viral replication is unknown, it might help explain the differences in pathogenicity seen in different IAV strains with divergent NS1. —SYM

Nat. Struct. Mol. Biol. **25**, 885 (2018).

GENETIC DISEASE

Impaired constriction

Duchenne muscular dystrophy (DMD) is a hereditary disease caused by mutations in the gene that encodes the protein

Flying squirrels, like this *Petaurista alborufus*, resemble 11.6-million-year-old fossils of *Miopetaurista*.

PALEONTOLOGY

Early tree gliders

Of all the vertebrate groups, mammals possess the greatest variation in form, from giant whales to tiny bumblebee bats. Between these extremes, flying squirrels provide another example of remarkable mammalian adaptation. Flying squirrels comprise a group of 52 species that have evolved to glide from tree to tree by means of large skin flaps between their forelimbs and hindlimbs. There has been much debate about when flying squirrels diverged from their more conventional relatives. Casanovas-Vilar *et al*. describe a flying squirrel fossil from the early Miocene, about 11.6 million years ago, the most complete one found to date. The *Miopetaurista* fossil is remarkably similar to the extant giant flying squirrels in the genus *Petaurista*, even down to the highly adapted wrist bones, indicating a stasis in form of these tropical forest denizens. —SNV

eLife **7**, e39270 (2018).

dystrophin. Dystrophin is also expressed in vascular smooth muscle cells (SMCs). A lack of dystrophin during disease causes damage to both skeletal and cardiac muscle. Pritchard *et al*. used superresolution microscopy to show that clusters of ryanodine receptors, which release Ca^{2+} from the sarcoplasmic reticulum, are larger in SMCs from a mouse model of DMD than in controls. These clusters colocalize with Ca^{2+}-activated K^+ (BK) channels on the channel membrane. The resulting higher activity of the BK channels causes impaired vasoconstriction in the cerebral microvessels of mutant mice. This effect might contribute to DMD-associated cognitive impairment. —VV

Proc. Natl. Acad. Sci. U.S.A. **115**, E9745 (2018).

TISSUE REGENERATION

Golden touch to heal muscles

Recovery from muscle injury requires divergent responses from macrophages. M1 macrophages promote inflammation and proliferation of muscle-resident stem cells (satellite cells). M1 cells then switch to M2 cells that dampen inflammation and promote satellite-cell differentiation. Although the anti-inflammatory cytokine interleukin-4 (IL-4) induces an M1-to-M2 shift, its use in therapy requires repeated infusions. Raimondo and Mooney overcome this hindrance by attaching IL-4 to gold nanoparticles. Fifteen days after such nanoparticles were injected into the injured leg muscle of mice, M2 macrophages and muscle fiber content increased, and the muscle contracted with greater force compared with muscle injected with naked nanoparticles or free IL-4. The nanoparticles sustained IL-4 in the injured area and may be beneficial in treating muscle disorders. —LC

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1806908115 (2018).

MECHANICAL BONDS

Rotaxanes get a hand

Molecules that don't match their mirror image are said to be chiral. In supramolecular rotaxanes, the threading direction of a ring on an axle can produce handedness of this sort even if the two parts themselves are achiral. Jinks *et al.* report a selective method to make either one of the two possible rotaxane mirror images. They rely on the click reaction of an alkyne and azide to assemble the axle within the ring, with a chiral center appended to the azide to bias the process. Symmetrizing that center afterward produces a purely mechanically chiral construct. —JSY

Angew. Chem. Int. Ed. 10.1002/anie.201808990 (2018).

BIOMATERIALS

A healing shot to the gut

Targeted oral delivery of therapeutics for the treatment of chronic bowel disease might alleviate some of the side effects caused by systemic delivery, but this route requires a means to prevent degradation of the drugs within the gastrointestinal tract. Li *et al.* devise a multicomponent drug delivery platform built around porous silicon nanoparticles. Surface functionalization with hyaluronic acid gives the particles a negative charge so that they do not adhere to intact mucosa. A hydrogel based on an active pharmaceutical agent and ascorbyl palmitate is loaded into the pores of the particles and releases the drug in the presence of specific enzymes. Furthermore, the entire construct is coated with hydroxypropyl methylcellulose acetate, which can be tuned to be soluble at a specific pH to provide protected delivery to the required location. —MSL

Biomaterials **185**, 322 (2018).

PHOTO: WILL BURRARD-LUCAS/MINDEN PICTURES

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

QUANTUM INFORMATION

The stages of a quantum internet

As indispensable as the internet has become in our daily lives, it still has many shortcomings, not least of which is that communication can be intercepted and information stolen. If, however, the internet attained the capability of transmitting quantum information—qubits—many of these security concerns would be addressed. Wehner *et al*. review what it will take to achieve this so-called quantum internet and propose stages of development that each correspond to increasingly powerful applications. Although a full-blown quantum internet, with functional quantum computers as nodes connected through quantum communication channels, is still some ways away, the first long-range quantum networks are already being planned. —JS

Science, this issue p. 303

CONSERVATION

A nature-friendly matrix

As the human population has grown, we have taken and modified more and more land, leaving less and less for nonhuman species. This is clearly unsustainable, and the amount of land we protect for nature needs to be increased and preserved. However, this still leaves vast regions of the world unprotected and modified. Such landscapes do not have to be a lost cause. Kremen and Merenlender review how biodiversity-based techniques can be used to manage most human-modified lands as "working landscapes." These can provide for human needs and maintain biodiversity not just for ecosystem services but also for maintenance and persistence of nonhuman species. —SNV

Science, this issue p. 304

ION CHANNELS

Structures of voltage-gated sodium channels

In "excitable" cells, like neurons and muscle cells, a difference in electrical potential is used to transmit signals across the cell membrane. This difference is regulated by opening or closing ion channels in the cell membrane. For example, mutations in human voltage-gated sodium (Na_v) channels are associated with disorders such as chronic pain, epilepsy, and cardiac arrhythmia. Pan *et al*. report the high-resolution structure of a human Na_v channel, and Shen *et al*. report the structures of an insect Na_v channel bound to the toxins that cause pufferfish and shellfish poisoning in humans. Together, the structures give insight into the molecular basis of sodium ion permeation and provide a path toward structure-based drug discovery. —VV

Science, this issue p. 305, p. 306

SOCIAL SCIENCE

Assessing gender differences

What contributes to gender-associated differences in preferences such as the willingness to take risks, patience, altruism, positive and negative reciprocity, and trust? Falk and Hermle studied 80,000 individuals in 76 countries who participated in a Global Preference Survey and compared the data with country-level variables such as gross domestic product and indices of gender inequality. They observed that the more that women have equal opportunities, the more they differ from men in their preferences. —BJ

Science, this issue p. 307

QUANTUM COMPUTING

Quantum outperforms classical

Quantum computers are expected to be better at solving certain computational problems than classical computers. This expectation is based on (well-founded) conjectures in computational complexity theory, but rigorous comparisons between the capabilities of quantum and classical algorithms are difficult to perform. Bravyi *et al*. proved theoretically that whereas the number of "steps" needed by parallel quantum circuits to solve certain linear algebra problems was independent of the problem size, this number grew logarithmically with size for analogous classical circuits (see the Perspective by Montanaro). This so-called quantum advantage stems from the quantum correlations present in quantum circuits that cannot be reproduced in analogous classical circuits. —JS

Science, this issue p. 308; see also p. 289

ORGANIC CHEMISTRY

Carbon nanotubes help nickel work in water

Most synthetic chemistry takes place in hydrocarbon-derived solvents. By contrast, enzymes manage to perform exquisitely selective reactions in water, often by surrounding reactants with hydrophobic pockets. Kitanosono *et al*. show that single-walled carbon nanotubes can similarly render simple nickel catalysts effective in water. Integration of the nickel ions with chiral ligands and surfactants at the nanotube surface produces a highly enantioselective catalyst for nitrone formation from aldoximes and unsaturated ketones. Spectroscopy suggests that the nanotubes enhance electron density at the nickel center as well as provide a hydrophobic milieu. —JSY

Science, this issue p. 311

CHEMICAL SENSING

Transistor sensing in salt solutions

Molecular binding to receptors on the surface of field-effect transistors (FETs) can be sensed through changes in transconductance. However, the saline solutions typically used with biomolecules create an electrical double layer that masks any events that occur within about 1 nanometer from the surface. Nakatsuka *et al*. overcame this limitation by using binding to large, negatively charged DNA stem loop structures that, upon ligand binding, cause conformational changes that can be sensed with an FET, even in solutions with high ionic strength. The authors demonstrate the sensing of charged molecules such as dopamine in artificial cerebrospinal fluid as well as neutral molecules such as glucose and zwitterion molecules like sphingosine-1-phosphate. —PDS

Science, this issue p. 319

TOPOLOGICAL MATTER

An iron home for Majoranas

The surface of the iron-based superconductor $FeTe_{0.55}Se_{0.45}$ has been identified as a potential topological superconductor and is expected to host exotic quasiparticles called the Majorana bound states (MBSs). Wang *et al*. looked for signatures of MBSs in this material by using scanning tunneling spectroscopy on the vortex cores formed by the application of a magnetic field. In addition to conventional states, they observed the characteristic zero-bias peaks associated with MBSs and were able to distinguish between the two, owing to the favorable ratios of energy scales in the system. —JS

Science, this issue p. 333

GERM CELL DEVELOPMENT

Reconstituting a human ovary

Human pluripotent stem cells (hPSCs) have been induced into human primordial germ cell–like cells (hPGCLCs) in vitro, the first step toward human in vitro gametogenesis. Yamashiro *et al*. went a step closer to generating mature gametes by culturing hPSCs with mouse embryonic ovarian somatic cells in xenogeneic reconstituted ovaries (see the Perspective by Gill and Peters). Over a period of 4 months, hPGCLCs underwent hallmark epigenetic reprogramming and differentiated progressively into cells closely resembling human oogonia, an immediate embryonic precursor for human oocytes. This study creates opportunities for human germ cell research and provides a foundation for human in vitro gametogenesis. —BAP

Science, this issue p. 356; see also p. 291

SUSTAINABILITY

Policies to make you happy

Policy decision-making frequently involves economic models to assess their value. However, economic metrics do not tell the full story of how populations experience the policies implemented in their countries. In a Perspective, Graham *et al*. discuss the possible value of introducing well-being metrics—to assess factors influencing life satisfaction and happiness—to policy decision-making. They propose that including well-being considerations in policy-making could ensure future sustainability of the workforce. —GKA

Science, this issue p. 287

ANTIGEN PRESENTATION

Stitching peptides for presentation

Intracellular protein–derived peptides generated by proteasomal degradation are loaded onto major histocompatibility complex (MHC) class I molecules in the endoplasmic reticulum and presented to $CD8^+$ T cells. Although it has been assumed that these peptides are contiguous segments derived from intracellular proteins, recent studies have shown that noncontiguous peptides generated by cis-splicing of two distinct regions of an antigen can be presented by MHC class I molecules. Faridi *et al*. now demonstrate that MHC class I molecules can present peptides that are generated by the splicing together of segments from two distinct proteins—so-called trans-spliced peptides. Precisely how cis- and trans-spliced peptides are generated and how they contribute to T cell selection and expansion remain to be explored. —AB

Sci. Immunol. **3**, eaar3947 (2018).

MAGNETIC RESONANCE

Hyperfine spectra of surface atoms

The interaction of nuclei with nonzero spin with electron spins creates small electronic energy. With a scanning tunneling microscope tip, Willke *et al*. measured these hyperfine interactions for iron and titanium atoms that were manipulated on a magnesium oxide surface. The tip was also used to measure electron paramagnetic resonance spectra. The hyperfine structure of single atoms was sensitive to the binding site of the atom as well as its position relative to other magnetic atoms. —PDS

Science, this issue p. 336

REVIEW SUMMARY

QUANTUM INFORMATION

Quantum internet: A vision for the road ahead

Stephanie Wehner*, David Elkouss, Ronald Hanson

BACKGROUND: The internet has had a revolutionary impact on our world. The vision of a quantum internet is to provide fundamentally new internet technology by enabling quantum communication between any two points on Earth. Such a quantum internet will—in synergy with the "classical" internet that we have today—connect quantum information processors in order to achieve unparalleled capabilities that are provably impossible by using only classical information.

As with any radically new technology, it is hard to predict all uses of the future quantum internet. However, several major applications have already been identified, including secure communication, clock synchronization, extending the baseline of telescopes, secure identification, achieving efficient agreement on distributed data, exponential savings in communication, quantum sensor networks, as well as secure access to remote quantum computers in the cloud.

Central to all these applications is the ability of a quantum internet to transmit quantum bits (qubits) that are fundamentally different than classical bits. Whereas classical bits can take only two values, 0 or 1, qubits can be in a superposition of being 0 and 1 at the same time. Moreover, qubits can be entangled with each other, leading to correlations over large distances that are much stronger than is possible with classical information. Qubits also cannot be copied, and any attempt to do so can be detected. This feature makes qubits well suited for security applications but at the same time makes the transmission of qubits require radically new concepts and technology. Rapid experimental progress in recent years has brought first rudimentary quantum networks within reach, highlighting the timeliness and need for a unified framework for quantum internet researchers.

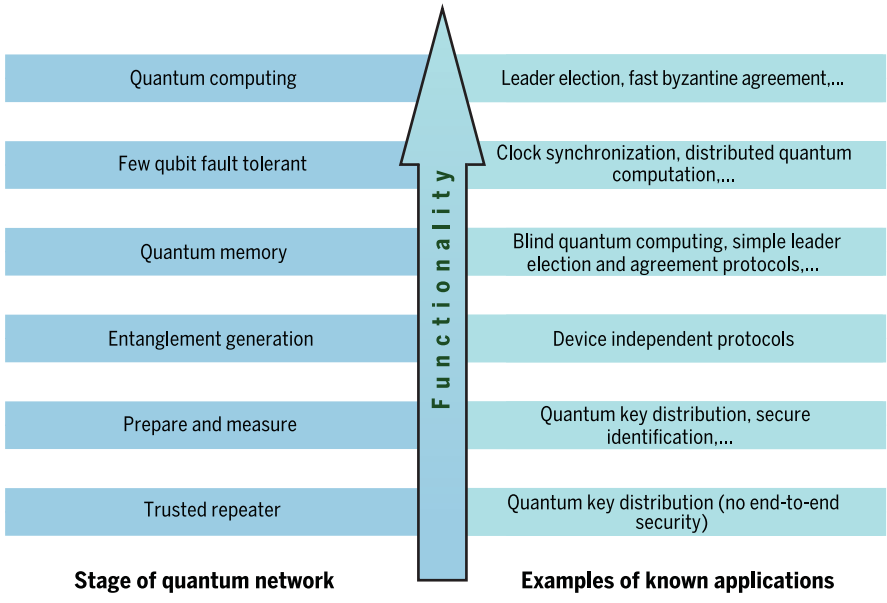
ADVANCES: We define different stages of development toward a full-blown quantum internet. We expect that this classification will be instrumental in guiding and assessing experimental progress as well as stimulating the development of new applications by providing a common language and reference frame for the different scientific and engineering disciplines involved.

More advanced stages are distinguished by a larger amount of functionality, thus supporting ever more sophisticated application protocols. For each stage, we describe some of the application protocols that are already known and that can be realized with the functionality provided in that stage. It is conceivable that a simpler protocol, or better theoretical analysis, may be found in the future that solves the same task but is less demanding in terms of functionality. In parallel to the daunting experimental challenges in making quantum internet a reality, there is thus an opportunity for quantum software developers to design protocols that can realize a task in a stage that can be implemented more easily. We identify relevant parameters for each stage to establish a common language between hardware and software developers. Last, we review technological progress in experimental physics, engineering, and computer science that is required to attain such stages.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aam9288

OUTLOOK: Building and scaling quantum networks is a formidable endeavor, requiring sustained and concerted efforts in physics, computer science, and engineering to succeed. The proposed stages of development will facilitate interdisciplinary communication by summarizing what we may actually want to achieve and providing guidelines both to protocol design and software development as well as hardware implementations through experimental physics and engineering. Although it is hard to predict what the exact components of a future quantum internet will be, it is likely that we will see the birth of the first multinode quantum networks in the next few years. This development brings the exciting opportunity to test all the ideas and functionalities that so far only exist on paper and may indeed be the dawn of a future large-scale quantum internet. ■



Stages in the development of a quantum internet. Each stage is characterized by an increase in functionality at the expense of greater technological difficulty. This Review provides a clear definition of each stage, including benchmarks and examples of known applications, and provides an overview of the technological progress required to attain these stages.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: s.d.c.wehner@tudelft.nl
Cite this article as S. Wehner *et al.*, *Science* **362**, eaam9288 (2018). DOI: 10.1126/science.aam9288

REVIEW

QUANTUM INFORMATION

Quantum internet: A vision for the road ahead

Stephanie Wehner[1]*, David Elkouss[1], Ronald Hanson[1,2]

The internet—a vast network that enables simultaneous long-range classical communication—has had a revolutionary impact on our world. The vision of a quantum internet is to fundamentally enhance internet technology by enabling quantum communication between any two points on Earth. Such a quantum internet may operate in parallel to the internet that we have today and connect quantum processors in order to achieve capabilities that are provably impossible by using only classical means. Here, we propose stages of development toward a full-blown quantum internet and highlight experimental and theoretical progress needed to attain them.

The purpose of a quantum internet is to enable applications that are fundamentally out of reach for the classical internet. A quantum internet could thereby supplement the internet we have today by using quantum communication, but some researchers go further and believe all communication will eventually be done over quantum channels (*1*). The best-known application of a quantum internet is quantum key distribution (QKD), which enables two remote network nodes to establish an encryption key whose security relies only on the laws of quantum mechanics. This is impossible with the classical internet. A quantum internet, however, has many other applications (Fig. 1) that bring advantages that are unattainable with a classical network. Such applications include secure access to remote quantum computers (*2*), more accurate clock synchronization (*3*), and scientific applications such as combining light from distant telescopes to improve observations (*4*). As the development of a quantum internet progresses, other useful applications will likely be discovered in the next decade.

Central to all these applications is that a quantum internet enables us to transmit quantum bits (qubits), which are fundamentally different from classical bits. Classical bits can take only two values, 0 or 1, whereas qubits can be in a superposition of 0 and 1 at the same time. Importantly, qubits cannot be copied, and any attempt to do so can be detected. It is this feature that makes qubits naturally well suited for security applications but at the same time makes transmitting qubits over long distances a truly formidable endeavor. Because qubits cannot be copied or amplified, repetition or signal amplification are ruled out as a means to overcome imperfections, and a radically new technological development—such as quantum repeaters—is needed in order to build a quantum internet (Figs. 2 and 3) (*5*).

We are now at an exciting moment in time, akin to the eve of the classical internet. In late 1969, the first message was sent over the nascent four-node network that was then still referred to as the Advanced Research Projects Agency Network (ARPANET). Recent technological progress (*6*–*9*) now suggests that we may see the first small-scale implementations of quantum networks within the next 5 years.

At first glance, realizing a quantum internet (Fig. 3) may seem even more difficult than building a large-scale quantum computer. After all, we might imagine that in full analogy to the classical internet, the ultimate version of a quantum internet consists of fully fledged quantum computers that can exchange an essentially arbitrary number of qubits. Thankfully, it turns out that many quantum network protocols do not require large quantum computers to be realized; a quantum device with a single qubit at the end point is already sufficient for many applications. What's more, errors in quantum internet protocols can often be dealt with by using classical rather than quantum error correction, imposing fewer demands on the control and quality of the qubits than is the case for a fully fledged quantum computer. The reason why quantum internet protocols can outperform classical communication with such relatively modest resources is because their advantages rely solely on inherently quantum properties such as quantum entanglement, which can be exploited already with very few qubits. By contrast, a quantum computer must feature more qubits than can be simulated on a classical computer in order to offer a computational advantage. Given the challenges posed by the development of a quantum internet, it is useful to reflect on what capabilities are needed to achieve specific quantum applications and what technology is required to realize them.

Here, we propose stages of development toward a full-blown quantum internet. These stages are functionality driven: Central to their definition is not the difficulty of experimentally achieving them but rather the essential question of what level of complexity is needed to actually enable useful applications. Each stage is interesting in its own right and distinguished by a specific quantum functionality that is sufficient to support a certain class of protocols. To illustrate this, for each stage we give examples of known application protocols in which a quantum internet is already known to bring advantages.

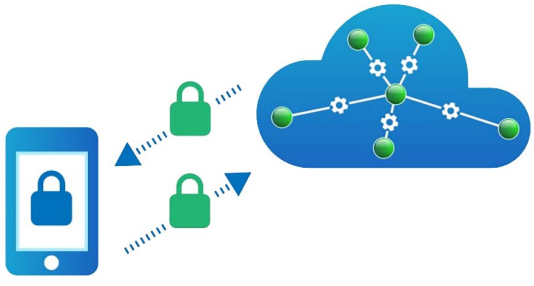


Fig. 1. Applications of a quantum internet. One application of a quantum internet is to allow secure access to remote quantum computers in the cloud (*2*). Specifically, a simple quantum terminal capable of preparing and measuring only single qubits can use a quantum internet to access a remote quantum computer in such a way that the quantum computer can learn nothing about which computation it has performed. Almost all other applications of a quantum internet can be understood from two special features of quantum entanglement. First, if two qubits at different network nodes are entangled with each other, then such entanglement enables stronger than classical correlation and coordination. For example, for any measurement on qubit 1, if we made the same measurement on qubit 2, then we instantaneously obtain the same answer, even though that answer is random and was not determined ahead of time. Very roughly, it is this feature that makes entanglement so well suited for tasks that require coordination. Examples include clock synchronization (*3*), leader election, and achieving consensus about data (*53*), or even using entanglement to help two online bridge players coordinate their actions (*39*). The second feature of quantum entanglement is that it cannot be shared. If two qubits are maximally entangled with each other, then it is impossible by the laws of quantum mechanics for a third qubit to be just as entangled with either of them. This makes entanglement inherently private, bringing great advantages to tasks that require security such as generating encryption keys (*12*) or secure identification (*24*, *25*).

[1]QuTech, Delft University of Technology, Post Office Box 5046, 2600 GA Delft, Netherlands. [2]Kavli Institute of Nanoscience, Delft University of Technology, Post Office Box 5046, 2600 GA Delft, Netherlands.
*Corresponding author. Email: s.d.c.wehner@tudelft.nl

Realizing a quantum internet demands substantial development to realize quantum repeaters as well as end nodes (Figs. 2 and 3). It is clear that in the short term, one may optimize both repeaters and end nodes relatively independently. That is, one can imagine a quantum internet that uses relatively simple end nodes while using repeaters powerful enough to cover larger distances. Similarly, a near-term quantum internet may be optimized for shorter—for example, pan-European—distances, while using much more powerful end nodes capable of realizing a larger set of protocols. Ideally, these designs would ensure forward compatibility to achieve the ultimate goal of a full-blown worldwide quantum internet. Although the quantum repeaters, which enable communication between distant end nodes, need to be able to support the functionality of each stage, an application-centric view makes no other statements regarding their capabilities.

Last, we discuss progress toward implementing a quantum internet, which poses substantial challenges to physics, engineering, and computer science.

Stages of functionality and applications

Let us formulate the functionality-driven stages of quantum internet development. Each successive stage is distinguished by an increasing amount of functionality, at the expense of increasing experimental difficulty. We say that an experimental implementation has reached a certain stage only if the functionality of that stage and all previous stages (Fig. 4) is available to all the end nodes using the network.

Crucial to the distinction between the stages is that the subsequent stage offers a fundamentally new functionality not available in the previous one rather than simply improving parameters or offering "more of the same" by increasing the number of qubits. For the sake of clarity, the stages and tests described below target systems that prepare and transmit qubits, but it is also possible to phrase both in terms of qudits (higher-dimensional quantum systems) or continuous variables. For each stage, we describe some of the application protocols that are already known and that can be realized with the functionality provided in that stage (Table 1). It is conceivable that a simpler protocol, or better theoretical analysis, may be found in the future that solves the same task but is less demanding in terms of functionality. In parallel to the daunting experimental challenges in making quantum internet a reality, there is thus a challenge for quantum software developers to design protocols that can realize a task in a stage that can be implemented more easily. We identify relevant parameters for each stage to establish a common language between hardware and software developers. These parameters can be estimated by using a series of simple tests, allowing us to certify the performance of an experimental implementation in attaining a specific stage, as well as the performance of protocols depending on these parameters.

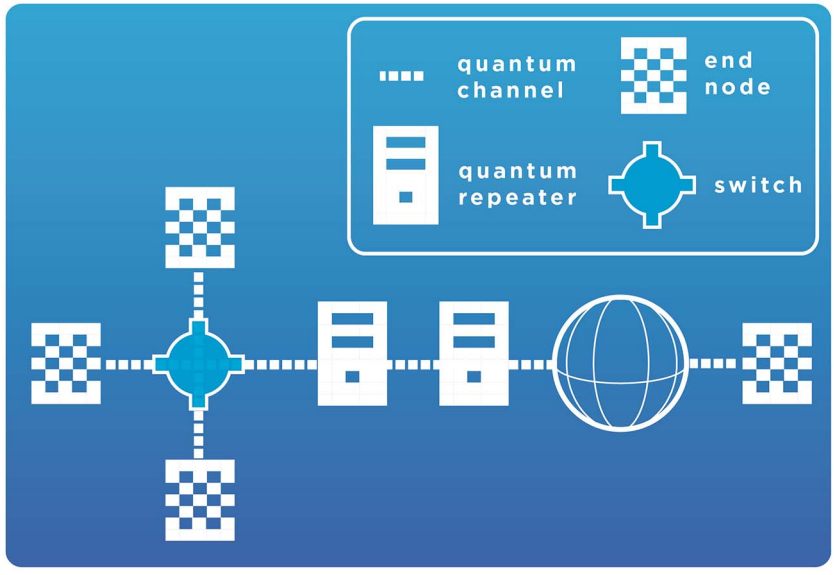


Fig. 2. A quantum internet consists of three essential quantum hardware elements. First, we need a physical connection (quantum channel) that supports the transmission of qubits. Examples are standard telecom fibers because they are presently used to communicate classical light. Second, we need a means to extend these short distances. Quantum channels are inherently lossy. For instance, the transmissivity of fiber optical channels scales exponentially with distance. This scaling has strong implications for applications because for both entanglement and key distribution, the achievable rates can at most be proportional to the transmissivity (*106*, *107*). Hence, in order to reach longer distances, intermediate nodes called quantum repeaters are necessary [(*97*, *108*–*110*), and (*91*, *92*), reviews]. Such a repeater is placed at certain intervals along the optical fiber connection, in theory allowing qubits to be transmitted over arbitrarily long distances. In the future, powerful repeaters may also double as long-distance routers in a quantum network. The final element are the end nodes—that is, the quantum processors connected to the quantum internet. These may range from extremely simple nodes that can only prepare and measure single qubits to large-scale quantum computers. End nodes may themselves act as quantum repeaters, although this is not a requirement. A quantum internet is not meant to replace classical communication but rather to supplement it with quantum communication. We hence assume all nodes can communicate classically—for example, over the classical internet—in order to exchange control information.

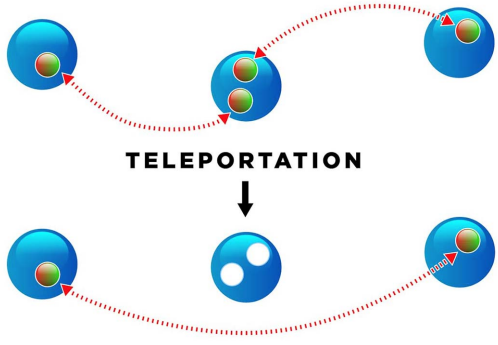


Fig. 3. Quantum repeaters work in a fundamentally different way from classical repeaters. Quantum repeaters are used to transmit quantum information over long distances. In its simplest form, a quantum repeater works by first generating entanglement (dashed line) between the repeater (middle) and each of the end nodes (left and right) individually. Intuitively, this can be done because the distance of each end point to the repeater is still sufficiently small to allow direct entanglement generation by transmitting photons over telecom fiber. Subsequently, the repeater teleports one of the qubits entangled with node 1 onto node 2. This procedure is known as entanglement swapping and allows the creation of entanglement over distances at which direct transmission is infeasible. After establishing long-distance entanglement, a data qubit may now be sent by using quantum teleportation.

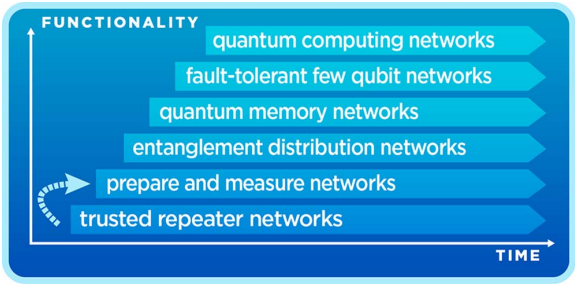


Fig. 4. Stages of quantum internet development. A specific implementation of a quantum internet may, like for a classical network, be optimized for distance, functionality, or both. The term network commonly refers to a situation that goes beyond point-to-point communication; the objective of a network is to provide any end nodes (connected to the network) with the means to exchange data, making three end nodes the smallest instance of a true network. Outside the laboratory, only trusted repeater networks (first stage) have been realized in metropolitan areas (*62*–*65*). Two single far-away end nodes (*68*) have also been connected via satellite.

So far, most application protocols have only been analyzed for perfect parameters. As such, the exact requirements of many application protocols on these parameters have not yet been determined and deserve future investigation. Although functionality-driven stages make demands on the communication links and quantum repeaters, it will not be important in this section how these links are realized; they may be realized by direct transmission in fiber, by being relayed by any kind of quantum repeater, or even by means of teleportation using preshared entanglement. What matters is that these links can be used to generate the necessary quantum states for a specific stage.

Trusted repeater networks

The first stage differs substantially from the others in the sense that it does not allow the end-to-end transmission of qubits. Nevertheless, from a technological perspective, trusted repeater networks can form an interesting stepping stone toward a quantum internet, spurring infrastructure deployment and engineering developments; depending on the underlying technology, trusted repeaters (*10*) can be upgraded to true quantum repeaters later on.

Specifically, a trusted repeater network (sometimes called a trusted node network) has at least two end nodes and a sequence of short distance links that connect nearby intermediary repeater nodes. Each pair of adjacent nodes uses QKD (*11*–*13*) to exchange encryption keys. These pairwise keys allow the end nodes to generate their own key, provided that all intermediary nodes are trusted (*14*). A first step toward upgrading such networks could be measurement device–independent QKD (*15*–*17*), which is a QKD protocol that is secure even with untrusted measurement devices that can be implemented with standard optical components and sources (*17*); this protocol already incorporates some useful ingredients for later stages, such as two-photon Bell measurements.

Prepare and measure networks

This stage is the first to offer end-to-end quantum functionality. It enables end-to-end QKD without the need to trust intermediary repeater nodes and already allows a host of protocols for other interesting tasks. Informally, this stage allows any node to prepare a one-qubit state and transmit the resulting state to any other node, which then measures it (definition is provided in Table 1). Transmission and measurement are allowed to be post-selected; that is, a signal that the qubit is lost may be generated instead. For instance, the receiving node is allowed to ignore nondetection events and conclude that such qubits are lost. If the sender can prepare an entangled state of two qubits, then this stage also includes the special case in which the sender transmits the first and second qubit to two different nodes in the network (or to another node and itself). Such entanglement distribution is then also post-selected.

Such a post-selected prepare-and-measure functionality is not equivalent to transmitting arbitrary qubits across the network (*18*). The task of transmitting arbitrary qubits demands the ability to transfer an unknown state $|\Psi\rangle$ (which the sender does not know how to prepare) deterministically to the receiver—that is, no post-selection on detection events is allowed.

The classical reader may wonder what is the use of transmitting qubits at all if there is a procedure for the sender to prepare the state $|\Psi\rangle$. After all, we might imagine that the sender simply sends classical instructions for this procedure to the receiver, who then prepares the qubit itself. The difference between such a classical protocol and sending different quantum states $|\Psi\rangle$ directly is that in the latter case, an eavesdropper, or indeed the receiver, cannot make a copy of $|\Psi\rangle$ without disturbing the quantum state. This means that attempts to gain information from $|\Psi\rangle$ by an eavesdropper may be detected, enabling QKD.

Application protocols

This stage is already sufficient to realize protocols for many interesting cryptographic tasks, as long as the probability of loss (p) and the inaccuracies in transmission (ε_T) and measurement (ε_M) (Table 1) are sufficiently low. The most famous of such tasks is QKD, which provides a solution to the task of generating a secure encryption key between two distant end nodes (Alice and Bob) (*11*–*13*). QKD is secure even if the eavesdropper trying to learn the key has access to an arbitrarily large quantum computer with which to attack the protocol, and remains secure at any point in the future, even if such a quantum computer becomes available later on. This is provably impossible when using classical communication. The BB84 QKD (*11*) protocol can be realized by using only single-qubit preparations and measurements tolerating some amount of post-selection p (*19*). For known protocols in this stage, $\varepsilon_T + \varepsilon_M \leq 0.11$ is sufficient and can be estimated by testing for only a small number of states (*20*). In practice, single-qubit preparation can be replaced with attenuated laser pulses, using also decoy-state BB84 to guarantee security (*21*). QKD is commercially available at short distances by using standard telecom fibers (*22*), and a variety of protocols are known [(*23*), survey].

Another class of protocols in this stage is in the domain of two-party cryptography. Here, there is no eavesdropper, but rather Alice and Bob themselves do not trust each other. An example of such a task is secure identification, in which Alice (a potentially impersonating user) may wish to identify herself to Bob (a potentially malicious server or automated teller machine) without revealing her authentication credentials (*24*, *25*). It is known that even by using quantum communication, such tasks cannot be implemented securely without imposing assumptions on the power of the adversary (*26*–*28*). Classical protocols rely on computational assumptions, whose security against an attacker who holds a quantum computer is unclear. Nevertheless, it is possible to achieve provable security for all such relevant tasks by sending more qubits than the adversary can store easily within a short time frame, which is known as the bounded (*29*) or more generally noisy-storage model (*30*, *31*). This assumption only needs to hold during the execution of the protocol, and security is preserved into the future even if the adversary later obtains a better quantum memory. There exist protocols for which it is sufficient to prepare and measure single qubits, in which the sufficient values of p, ε_M, ε_T (Table 1) depend on the storage assumption (*32*).

Other known protocols in this stage include position verification (*33*); weakened forms of two-party cryptographic tasks that can form building blocks, such as imperfect bit commitments (*34*); and coin-flipping (*35*). Here, the requirements in terms of p, ε_M, and ε_T have not been analyzed yet; no task exists for which a full set of necessary and sufficient conditions on these parameters is known.

Entanglement distribution networks

The third stage allows the end-to-end creation of quantum entanglement in a deterministic or heralded fashion, as well as local measurements. The end nodes require no quantum memory for this stage (Table 1).

The term "deterministic entanglement generation" refers to the fact that the process succeeds with (near) unit probability. Heralding is a slightly weaker form of deterministic entanglement generation in which we signal the successful generation of entanglement with an event that is independent of the (immediate) measurement of the

entangled qubits themselves. Here, the generation of entanglement is deterministic, conditioned on such a successful heralding signal. Specifically, this prohibits post-selecting on detection events when measuring the entangled qubits. We remark that this stage also includes networks that allow the generation of multipartite entangled states, followed by immediate measurements, but no memory. However, the generation of multipartite entanglement is not required to attain this stage.

Application protocols

The main advance over the previous stage is that this stage allows the realization of device-independent protocols, in which the quantum devices are largely untrusted. Specifically, the concept of device independence (*36*, *37*) models the end nodes as black boxes, to which we can give classical instructions to perform specific measurements and receive the resulting measurement outcomes. No guarantees are given about the actual quantum state or measurements performed by the device, where the device may even be constructed by the adversary. The classical software used to control such quantum devices is trusted, and it is assumed that the quantum device merely exhibits input/output behavior. In particular, devices can record their inputs and outputs (*38*) but cannot transmit the key back to the adversary. The coordination allowed by entanglement now also in principle allows players to "cheat" an online bridge game (*39*).

Low errors in preparation (ε_P) and measurement (ε_M) as $\varepsilon_P + \varepsilon_M \leq 0.057$ (Table 1) are sufficient to ensure the implementability of device-independent QKD (*36*), in which necessary and sufficient conditions for the parameters to implement general tasks in this stage are unknown.

Quantum memory networks

The fourth stage is distinguished by the capability of the end nodes to have local memory while simultaneously allowing universal local control (Table 1). This allows the implementation of much more complex protocols that require temporary storage of a quantum state during further quantum or classical communication. Examples include protocols for solving distributed systems tasks. This stage also implies the ability to perform entanglement distillation and generate multipartite entangled states from bipartite entanglement by exploiting the ability for local memory and control. A crucial difference between this stage and the previous one is that we are now able to transfer

Table 1. Formal definitions of the stages, parameters for protocol design, and classification of known protocols. Higher stages include all functionality available at the previous ones. It is an open question to determine necessary and sufficient conditions for these parameters to realize general protocols. In the future, quantum network programmers may be able to find protocols for the same tasks that can be realized with lower stages of a quantum internet. It is an interesting open question what minimum stage is required in order to realize a specific task.

Stage	Additional functionality	Parameters	Example protocols
Prepare and measure	For any two end nodes i, j, any one qubit state $\|\Psi\rangle$ and any one qubit projective measurement M, there exists a way for i to prepare $\|\Psi\rangle$, transfer it to j, so that either (i) j performs measurement M on $\|\Psi\rangle$ or (ii) j concludes the qubit was lost.	Distances ε_T and ε_M from the ideal transmission and measurement operations (Box 1). Probability p that the state is not lost.	QKD, Two-party cryptography, position verification, imperfect coin flipping
Entanglement distribution	For any two end nodes i, j, (i) the network allows the heralded creation of a maximally entangled state $\|\Phi_{ij}\rangle$ and (ii) nodes i and j can deterministically perform any single-qubit measurements M_i and M_j.	Distances ε_P from the ideal preparation, and ε_M from the idealized measurement (Box 1).	Device independence for QKD and other protocols in the prepare and measure stage
Quantum memory	For any two end nodes i, j, the network allows the execution entanglement generation and the following additional tasks in any order: (i) preparation of a one qubit ancilla state $\|\Psi\rangle$ by end node i or j, (ii) measurements of any subset of the qubits at node, and (iii) application of an arbitrary unitary U at node. Storage of the qubits for a minimum time $k \cdot C_m \cdot t$, where t is defined as the time that is required to generate one Einstein–Podolsky–Rosen (EPR) pair and send a classical message from node i to j maximized over all pairs of nodes, and C_m is the time that it takes for the execution of a depth m quantum circuit at the end node.	Number of rounds k, circuit depth m, number of physical qubits q. For each of the operations, an estimate ε_j from the ideal operation (Box 1).	Blind quantum computing (using remote quantum servers), improved coin flipping, anonymous quantum transmissions, extending baseline of telescopes, secret sharing, simple leader election and agreement protocols, and time-limited clock synchronization
Few-qubit fault-tolerant	Fault-tolerant execution of a universal gate set on q logical qubits, where $q \geq 1$ is small enough such that the local processor can efficiently be simulated on a classical computer.	Number of logical qubits q	Clock synchronization and distributed quantum computation
Quantum computing	q is larger than the number of qubits that can effectively be simulated on a classical computer.	Number of logical qubits q	Leader election, fast byzantine agreement, and weak coin flipping with arbitrarily small bias

unknown qubits from one network node to another—for example, by performing deterministic teleportation. This capability is not guaranteed in the previous stage: Technology that can be used to deterministically relay qubits over long distances by means of large-scale quantum error correction implies the technological capability of realizing a good local quantum memory. We emphasize that a quantum memory network does not require operations to be performed with an accuracy that would be above threshold for fault-tolerant computation.

An important parameter in application protocols is the number of communication rounds k (Table 1), the number of times information is sent back and forth between two end nodes during the course of the protocol. In order to realize useful application protocols, the storage time t thus needs to be compared with the communication time in the network instead of an absolute time. This means that networks of nodes that are far apart do in fact need to exhibit longer memory times in order to attain this stage, and the quality of the memory is time dependent. That this time t is related to the maximum time that it takes any two nodes to communicate is because a stage is attained only if the functionality is available to any two nodes in the network, even the two that are farthest apart.

Application protocols

The availability of quantum memories and the deterministic transmission of qubits opens up many new protocols in this stage. We start with cryptographic tasks: To allow clients to make use of these computers securely—that is, without revealing the nature or outcome of their computation—it is possible to perform secure assisted quantum computation (*40*), or blind quantum computation (*2*, *41*). Here, a simple quantum device capable of preparing and measuring single qubits is sufficient to perform a computation on a large-scale quantum computer so that the quantum computer cannot gain information about the program and result. That we need one large-scale quantum computer does not imply that a quantum computing network (the highest stage) is required to run such protocols; we only need a quantum internet that allows a client to communicate with the computing server. A network attains a specific stage only if the functionality is available to all nodes.

Other cryptographic tasks in this domain are tools such as protocols for the sharing of classical (*42*) or quantum (*43*) secrets, including verifiable secret-sharing schemes (*44*) and anonymous transmissions (*45*). Evidently, the number of qubits determines the size of the secrets or qubits transmitted, but no fault tolerance is in principle required.

This stage also opens the door to interesting applications outside the domain of cryptography. For example, proposals exist for exploiting long-distance entanglement to extend the baseline of telescopes (*4*), for basic forms of leader election (*46*), and for improving the synchronization of clocks (*3*). Depending on the demands made on such synchronization, the proposed protocols could be realized with quantum memory or few-qubit fault-tolerant networks.

Necessary and sufficient parameter requirements for solving the above mentioned tasks are not yet known in general. It is also conceivable that an improved analysis considering whether deterministic qubit delivery is really necessary, or whether maybe post-selected transmission of qubits is enough, can push some of the protocols above to a lower stage. Initial results for blind quantum computation indicates that this might indeed be the case (*47*).

Few-qubit fault-tolerant networks

The next stage differs by demanding that the local operations can be performed fault-tolerantly, which is considerably more challenging. Fault tolerance is not necessary for many known quantum internet protocols, but fault-tolerant operations being available would allow the execution of local quantum computation of high circuit depth as well as an (in theory) arbitrary extension of storage times to execute protocols with an arbitrary number of rounds of communication.

The term "few qubits" here refers to the fact that the number of qubits available is still small enough so that the end nodes themselves can be simulated effectively on a classical computer. This does not imply that the entire network can be simulated efficiently or that there would exist equivalent classical protocols; the effects of entanglement cannot generally be replicated classically.

Here, we are only interested in the performance of the fault-tolerant scheme, not how it is realized. Fault tolerance implies that all error parameters (Table 1) of a quantum memory network can be made negligible by adding more resources. As a guideline to relevant experimental parameters, we refer to works in distributed quantum computing (*48*).

Application protocols

Having access to fault-tolerant gates allows higher-accuracy clock synchronization (*3*) and protocols that require many rounds of communication and high circuit depth to be useful. This includes distributed quantum computing as well as applications for full-scale quantum computing networks, restricted to few qubits. This could be of great practical interest, especially for applications in the domain of distributed systems, but as with the implementation of quantum algorithms on quantum computers, the power of having only a limited number of qubits at our disposal is an important subject of investigation.

Quantum computing networks

The final stage consists of quantum computers that can arbitrarily exchange quantum communication. In some sense, it breaks with our paradigm that the next stage is not "more of the same." However, in this case, we really do gain a new ability: finding solutions to computational problems that can no longer be found efficiently on classical computers.

Application protocols

It is clear that this ultimate stage of a quantum internet allows in principle all protocols to be realized. Small-scale versions of the protocols below can also be realized in the few-qubit fault-tolerant stage, and further development may yield more sophisticated protocols and analysis that places them in lower stages.

First, we again focus on cryptography. In this stage, it is possible to perform coin flipping with an arbitrarily small bias (*49*, *50*). We can also solve genuinely quantum tasks, such as secure multiparty quantum computation, which forms an extension of classical secure function evaluation to the quantum regime. Classically, this means that node j holds an input string x_j, and all

Box 1. Performance of quantum internet protocols.

A general quantum internet protocol is composed of a series of operations consisting of state preparation, transmission, unitary operations, and measurements. In reality, each of these operations is noisy, so instead of executing a sequence of ℓ ideal operations $\mathcal{J} = \mathcal{J}_\ell \circ \ldots \circ \mathcal{J}_1$, we are executing the real (noisy) protocol $\mathcal{R} = \mathcal{R}_\ell \circ \ldots \circ \mathcal{R}_1$. To assess the performance of the real protocol execution, it is sufficient to estimate the diamond norm distance (*20*)

$$D_\diamond(\mathcal{R}, \mathcal{J}) = \max_{\rho_{SE}} D[\mathcal{R} \otimes \mathrm{id}_E(\rho_{SE}), \mathcal{J} \otimes \mathrm{id}_E(\rho_{SE})]$$

where $D(\tau, \sigma)$ is the well-known trace distance (*18*) that determines how well two states τ and σ can be distinguished by any physical process, and S denotes the system that the protocol acts on which may be part of a larger system SE. Because $D_\diamond$ is (unlike the fidelity) a metric, it is straightforward to show that having estimated individual errors $\|\mathcal{R}_j - \mathcal{J}_j\|_\diamond \leq \varepsilon$ allows an estimate of the overall error as

$$D_\diamond(\mathcal{R}, \mathcal{J}) \leq \ell \cdot \varepsilon$$

For unitary operations and projective measurements, the diamond norm distance is directly related to the average gate fidelity (*111*). If the ideal operation $\mathcal{J}(\rho) = \Phi$ simply aims to prepare a state Φ, and the real operation prepares $\mathcal{R}(\rho) = \tilde{\Phi}$, then the diamond norm distance satisfies $D_\diamond(\mathcal{R}, \mathcal{J}) \leq \sqrt{1 - F(\Phi, \tilde{\Phi})}$, where F is the fidelity. Evidently, the end-user—who desires to run application protocols—should be able to perform tests that give confidence for any possible operation instead of having to test the exact unitaries and measurements in any conceivable protocol.

n nodes jointly want to compute $y = f(x_1, ..., x_n)$. The goal is that malicious nodes cannot infer anything more about the inputs x_j of the honest nodes than they can by observing the output y. An example of such a problem is secure voting, in which $x_j \in \{0, 1\}$ corresponds to the choice one of two possible candidates, and f is the majority function. The quantum version of this primitive (*51*) allows each party to hold a quantum state $|\Psi_j\rangle$ as input, and the parties jointly wish to compute a quantum operation U.

Next, we focus on distributed systems, which are formed when several computing devices are connected, sometimes colloquially referred to as a cloud. Many challenges arise in the coordination and control of such systems that may be less familiar to a physicist. As a very simple example, consider a bank transaction being recorded redundantly on several backup servers. If one or more of the backup servers fail during the update, then they may later show inconsistent data (for example, $1 million versus $0). Tool protocols for achieving consensus between processors are widely deployed in practice—for example, in Google's Chubby system (*52*). Outside the domain of the internet itself, examples include the reliability in smart grids, flight control systems, and sensor arrays.

Although this area is presently much less developed in the quantum domain (*53*), several protocols are known that show that a quantum internet has great potential for solving the problems in distributed systems much more efficiently than what is possible classically. Very intuitively, the reason why quantum communication could help solve these problems is that entanglement allows coordination among distant processors that greatly surpasses what is possible classically. It is this that yields advantages for distributed systems tasks such as consensus and agreement. One of the most striking examples of a quantum advantage in distributed systems can be found for the task of byzantine agreement. Here, the goal is to allow n processors to agree on a common bit, while some fraction of them may be faulty. The term "byzantine" refers to the very demanding model of arbitrarily correlated faults, in which the faulty processors essentially collude to thwart the protocol. In (*54*), it is shown that in some regimes, there exists a quantum protocol to solve this task by using only a constant number of rounds of quantum communication, while the amount of classical communication scales as $O(\sqrt{n}\log n)$, where n is the number of processors. The protocol given in (*54*) requires many qubits, thus demanding the final stage of a quantum internet. The objective of leader election is to elect a distinct leader from a number of distributed processors, which is an important tool, for example, for deciding which processor gets to use a particular resource. This task is particularly challenging in an anonymous network, in which no node has an identifier. In this setting, there is no exact classical algorithm for leader election for general network topologies, whereas quantumly, leader election is possible (*55*). The protocol proposed in (*55*) requires each end node to process a number of qubits that scales with the number of processors (end nodes). To be used in networks of reasonable size, we thus require a quantum computing network. A number of other leader-election protocols have been proposed in a variety of models (*56*, *57*).

Last, this stage allows distributed computational tasks to be solved by transmitting in some cases even exponentially fewer (*58*) qubits than classical bits. A notable example is fingerprinting (*59*). However, these protocols generally require a large number of qubits at each end node to achieve a substantial advantage. Specific variants of such protocols with energy constraints can also be realized at lower stages (*60*). Last, the presence of entanglement also brings new security issues for existing classical protocols (*61*), requiring new insights and analysis.

Implementation status and challenges

The current experimental status of long-distance quantum networks is at the lowest stage—trusted-repeater networks—with several commercial systems for QKD on the market. The first extended trusted repeater networks have already been implemented over metropolitan distances (*62*–*65*), and a long-distance implementation has recently been completed (*66*). The hardware required at the lowest stage (mainly light sources, optical links, and detectors) has been described in detail in previous literature (*14*, *23*). Realizing the first stage with end-to-end quantum functionality—prepare-and-measure networks—over long distances demands the use of quantum repeaters to bridge long distances via intermediate qubit storage or error correction, as well as routers to forward the quantum state to the desired node. Several recent experiments have demonstrated elements belonging to this and higher stages at short distances, suggesting that higher-functionality networks are within reach. To put these experiments into the right perspective, we briefly summarize the main requirements for three types of quantum internet hardware.

Photonic communication channels

Photonic channels establish quantum links between the distant repeater stations and between the end nodes. Two types of photonic channels can be distinguished: free-space channels [potentially via satellites (*67*, *68*)] and fiber-based channels. Each has its own advantages and disadvantages, and a future quantum internet—similar to the current classical internet—may use a combination of them. We require these channels to exhibit minimal photon loss and decoherence. The effect of photon loss on fidelity can in general be dealt with by photon-heralding protocols, but photon loss unavoidably affects the communication rate across the network. For photons in the telecom frequency bands, loss in fibers can be as low as 0.2 dB/km. Decoherence can in general be overcome through entanglement distillation (*69*–*71*), which requires additional levels of qubit processing. Last, the bandwidth of the channels is of practical importance; multiplexing in frequency, time, spatial, and/or polarization degrees of freedom allows for increases of the communication rates.

End nodes

For the quantum internet to reach its full potential, the end nodes need to meet the following requirements.

(i) Robust storage of quantum states during the time needed to establish entanglement between end nodes. This robustness must persist under quantum operations performed on the end node.

(ii) High-fidelity processing of quantum information within the node. For the more advanced tasks, multiple qubits will be required, making the end nodes similar to small-scale quantum computers.

(iii) Compatibility with photonic communication hardware: efficient interface to light at the relevant wavelength (telecom bands for fiber-based networks).

Several experimental platforms are currently being pursued for the end nodes. Each of these combines well-controlled matter-based qubits with a quantum optical interface via internal electronic transitions. The generation of photon-mediated entanglement between distant matter qubits has been achieved with trapped ions (*72*), atoms (*73*, *74*), nitrogen-vacancy (N-V) centers in diamond (*75*), and semiconductor quantum dots (*76*, *77*) over distances up to 1.3 km (*78*). By using measurement-based schemes with heralding, high-fidelity entangled states could be created in these experiments, even though substantial photon loss was present. The major challenge in extending these point-to-point entangled links into true networks is the robust storage of quantum states. The intrinsic coherence times of most above-mentioned platforms are very long (for instance, more than a second for ions and N-V centers). However, cross-talk caused by unwanted couplings or imperfect individual addressability can severely affect the coherence of a memory qubit under operations on another qubit in the same node (*79*, *80*).

A promising approach is to use different types of qubits within a node. For instance, trapping different species of ions allows for individual addressing of the ions via their different electronic transition frequencies (*81*–*83*). In a similar fashion, carbon-13 nuclear spins near a diamond N-V center provide a robust register of memory qubits that do not interact with the laser control fields on the N-V electron spin (*84*). In a very recent experiment, such hybrid network nodes enabled the generation of two remote entangled states on which entanglement distillation could then be performed (*85*). If several of such robust memories can be successfully integrated into a multi-qubit network node, the highest stages of the quantum internet may come into reach.

Another challenge for most of the above systems is that these do not intrinsically couple to light in the telecom band. To fulfill requirement (iii), wavelength conversion at the single-photon level can be used. Pioneering experiments using nonlinear optics (*86*, *87*) have already demonstrated

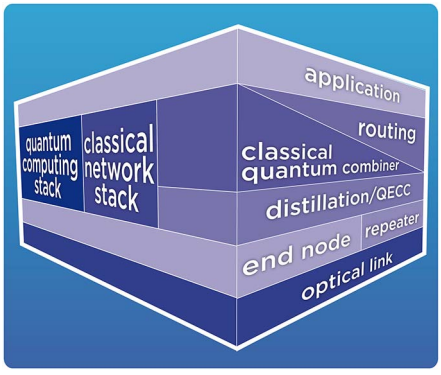


Fig. 5. Possible elements of a future quantum network stack.

the feasibility of such conversion; the current challenge is to realize a robust and high-efficiency (say, >50%) converter that exhibits a high signal-to-noise ratio (say, >100).

As an alternative to the above systems with intrinsic optical interface, the end nodes could be formed from a quantum processor with qubit frequencies in the microwave domain, such as a superconducting qubit circuit, in combination with a microwave-to-optical conversion process. The physics of such a conversion—for instance, by use of mechanical resonators (*88*, *89*) or atomic transitions (*90*)—is currently being investigated in many laboratories.

Quantum repeater requirements

Quantum repeater stations need to improve the rate of photonic qubit transfer. The requirements for quantum repeaters are similar to but less strict than for the end nodes. In particular, depending on the exact architecture [(*91*), review], the storage of quantum states may only be required for the time needed to establish entanglement between the nearest active nodes; this storage time can deviate substantially from that required for the end nodes. Also, the qubit processing capabilities required are limited, and therefore systems different from the ones above can be considered. As a prime example, an ensemble of atoms and ions either in gas phase or in a solid can be used as an on-demand quantum memory (*92*). If the memory can herald the arrival of a photon and store the photon's quantum state, photon loss can be efficiently mitigated. Storage and on-demand retrieval have already been achieved (*93*–*96*), although efficiencies are still to be improved. Such memories also allow for multiplexing within a single device. Furthermore, they are compatible with current-day down-conversion sources for entangled photon pairs. Current challenges are to combine heralding and on-demand high-efficiency retrieval with long coherence times.

A radically different approach to quantum repeaters has emerged in recent years in which the quantum state of interest is encoded in multiple photons so that error correction performed at the repeater stations can erase errors caused by photon loss and decoherence during transmission (*97*–*100*). The main advantage of such a scheme is that the classical two-way communication of standard repeater schemes (necessary to convey the heralding signal of whether or not the photons arrived at the stations) becomes obsolete. The communication rates of these schemes are therefore potentially much higher. However, the experimental demands seem daunting at present; for encoding the qubit, the near-deterministic generation of a many-photon cluster state is required, which is far beyond the state of the art (*101*). Furthermore, because these schemes require quantum error correction, they will only work if the error thresholds associated with the desired transmission qualities are met, thus placing more stringent requirements on the control and readout fidelities within the repeater nodes. That being said, theory research (*102*) in this direction is likely to yield more insights, and experimental progress may bring such schemes closer to reality in the future.

Last, the end nodes that are currently being developed may also function themselves as repeaters.

Network stack requirements

In order to enable widespread use and application development, it is essential to develop methods that allow quantum protocols to connect to the underlying hardware implementation transparently and to make fast and reactive decisions for generating entanglement in the network in order to mitigate limited qubit lifetimes (Fig. 5). Classically, this is achieved by a series of layered protocols such as the Transmission Control Protocol/Internet Protocol (TCP/IP) (*103*) that provide an abstraction that ultimately allows application protocols to exchange data between two end nodes without having to know any details on how this connection is actually realized. No such network stack presently exists for a quantum internet, and only some basic elements have been noted (*104*). As a trivial example on why a new stack is required for a quantum network, the first novel feature is a mapping between classical control information (header) and the underlying qubits. By contrast, classically a header and data may be nicely combined in one piece of data to be transmitted. Another example is the use of error detection at the link layer of the classical network stack that does not easily translate to a realistic quantum network. Clearly, error detection can theoretically be realized by using quantum error-correcting codes, but this method may be prohibitively expensive in practice, and other methods (*105*) may be more suitable. These are just two simple examples of the challenges involved in designing such a network stack, calling for substantial development.

Although it is hard to predict what the exact physical components of a future quantum internet will be, it is likely that we will see the birth of the first multinode quantum networks in the next few years. This development brings the exciting opportunity to test all the ideas and functionalities that so far only exist on paper and may indeed be the dawn of a future large-scale quantum internet.

REFERENCES AND NOTES

1. D. Castelvecchi, The quantum internet has arrived (and it hasn't). Nature **554**, 289–292 (2018).
2. A. Broadbent, J. Fitzsimons, E. Kashefi, *50th Annual IEEE Symposium on Foundations of Computer Science, 2009.* (IEEE, 2009), pp. 517–526.
3. P. Kómár *et al.*, A quantum network of clocks. *Nat. Phys.* **10**, 582–587 (2014). doi: 10.1038/nphys3000
4. D. Gottesman, T. Jennewein, S. Croke, Longer-baseline telescopes using quantum repeaters. *Phys. Rev. Lett.* **109**, 070503 (2012). doi: 10.1103/PhysRevLett.109.070503; pmid: 23006349
5. H. J. Kimble, The quantum internet. *Nature* **453**, 1023–1030 (2008). doi: 10.1038/nature07127; pmid: 18563153
6. A. I. Lvovsky, B. C. Sanders, W. Tittel, Optical quantum memory. *Nat. Photonics* **3**, 706–714 (2009). doi: 10.1038/nphoton.2009.231
7. T. Northup, R. Blatt, Quantum information transfer using photons. *Nat. Photonics* **8**, 356–363 (2014). doi: 10.1038/nphoton.2014.53
8. D. D. Awschalom, R. Hanson, J. Wrachtrup, B. B. Zhou, Quantum technologies with optically interfaced solid-state spins. *Nat. Photonics* **12**, 516–527 (2018). doi: 10.1038/s41566-018-0232-2
9. A. Reiserer, G. Rempe, Cavity-based quantum networks with single atoms and optical photons. *Rev. Mod. Phys.* **87**, 1379–1418 (2015). doi: 10.1103/RevModPhys.87.1379
10. L. Salvail *et al.*, Security of trusted repeater quantum key distribution networks. *J. Comput. Secur.* **18**, 61–87 (2010). doi: 10.3233/JCS-2010-0373
11. C. H. Bennett, G. Brassard, *International Conference on Computer System and Signal Processing, IEEE, 1984* (1984), pp. 175–179.
12. A. K. Ekert, Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991). doi: 10.1103/PhysRevLett.67.661; pmid: 10044956
13. S. Wiesner, Conjugate coding. *ACM SIGACT News* **15**, 78–88 (1983). doi: 10.1145/1008908.1008920
14. V. Scarani *et al.*, The security of practical quantum key distribution. *Rev. Mod. Phys.* **81**, 1301–1350 (2009). doi: 10.1103/RevModPhys.81.1301
15. E. Biham, B. Huttner, T. Mor, Quantum cryptographic network based on quantum memories. *Phys. Rev. A* **54**, 2651–2658 (1996). doi: 10.1103/PhysRevA.54.2651; pmid: 9913773
16. S. L. Braunstein, S. Pirandola, Side-channel-free quantum key distribution. *Phys. Rev. Lett.* **108**, 130502 (2012). doi: 10.1103/PhysRevLett.108.130502; pmid: 22540685
17. H.-K. Lo, M. Curty, B. Qi, Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **108**, 130503 (2012). doi: 10.1103/PhysRevLett.108.130503; pmid: 22540686
18. M. M. Wilde, *Quantum Information Theory* (Cambridge Univ. Press, 2013).
19. C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, H. M. Wiseman, One-sided device-independent quantum key distribution: Security, feasibility, and the connection with steering. *Phys. Rev. A* **85**, 010301 (2012). doi: 10.1103/PhysRevA.85.010301
20. A. Gilchrist, N. K. Langford, M. A. Nielsen, Distance measures to compare real and ideal quantum processes. *Phys. Rev. A* **71**, 062310 (2005). doi: 10.1103/PhysRevA.71.062310
21. H.-K. Lo, X. Ma, K. Chen, Decoy state quantum key distribution. *Phys. Rev. Lett.* **94**, 230504 (2005). doi: 10.1103/PhysRevLett.94.230504; pmid: 16090452
22. A. Extance, *Fibre Systems* (2017); www.fibre-systems.com/feature/quantum-security.
23. E. Diamanti, H.-K. Lo, B. Qi, Z. Yuan, Practical challenges in quantum key distribution. *NPJ Quantum Information* **2**, 16025 (2016). doi: 10.1038/npjqi.2016.25
24. I. Damgård, S. Fehr, L. Salvail, C. Schaffner, Secure identification and QKD in the bounded-quantum-storage model. *Theor. Comput. Sci.* **560**, 12 (2014). doi: 10.1016/j.tcs.2014.09.014
25. F. Dupuis, O. Fawzi, S. Wehner, Entanglement sampling and applications. *IEEE Trans. Inf. Theory* **61**, 1093–1112 (2014). doi: 10.1109/TIT.2014.2371464
26. D. Mayers, Unconditionally secure quantum bit commitment is impossible. *Phys. Rev. Lett.* **78**, 3414–3417 (1997). doi: 10.1103/PhysRevLett.78.3414

27. H.-K. Lo, H. F. Chau, Is quantum bit commitment really possible? *Phys. Rev. Lett.* **78**, 3410–3413 (1997). doi: 10.1103/PhysRevLett.78.3410
28. H.-K. Lo, Insecurity of quantum secure computations. *Phys. Rev. A* **56**, 1154–1162 (1997). doi: 10.1103/PhysRevA.56.1154
29. I. B. DamgÅrd, S. Fehr, L. Salvail, C. Schaffner, Cryptography in the bounded-quantum-storage model. *SIAM J. Comput.* **37**, 1865 (2000). doi: 10.1137/060651343
30. S. Wehner, C. Schaffner, B. M. Terhal, Cryptography from noisy storage. *Phys. Rev. Lett.* **100**, 220502 (2008). doi: 10.1103/PhysRevLett.100.220502; pmid: 18643410
31. R. Konig, S. Wehner, J. Wullschleger, Unconditional security from noisy quantum storage. *IEEE Trans. Inf. Theory* **58**, 1962 (2012). doi: 10.1109/TIT.2011.2177772
32. N. H. Y. Ng, S. K. Joshi, C. C. Ming, C. Kurtsiefer, S. Wehner, Experimental implementation of bit commitment in the noisy-storage model. *Nat. Commun.* **3**, 1326 (2012). doi: 10.1038/ncomms2268; pmid: 23271659
33. J. Ribeiro, F. Grosshans, arXiv:1504.07171 [quant-ph] (2015).
34. A. Chailloux, I. Kerenidis, *Proceedings of the 52th Annual Symposium on Foundations of Computer Science* 10.1109/FOCS.2011.42 (2011).
35. D. Aharonov, A. Ta-Shma, U. V. Vazirani, A. C. Yao, *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing* 10.1145/335305.335404 (2000).
36. A. Acín *et al*., Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **98**, 230501 (2007). doi: 10.1103/PhysRevLett.98.230501; pmid: 17677888
37. D. Mayers, A. Yao, *Proceedings of the 39th Annual Symposium on Foundations of Computer Science, 1998* (IEEE, 1998), pp. 503–509.
38. J. Barrett, R. Colbeck, A. Kent, Memory attacks on device-independent quantum cryptography. *Phys. Rev. Lett.* **110**, 010503 (2013). doi: 10.1103/PhysRevLett.110.010503; pmid: 23383767
39. S. Muhammad *et al*., Quantum bidding in bridge. *Phys. Rev. X* **4**, 021047 (2014). doi: 10.1103/PhysRevX.4.021047
40. A. M. Childs, *Quantum Inf. Comput.* **5**, 456 (2005).
41. D. Aharanov, M. Ben-Or, E. Eban, *Proceedings of Innovations in Computer Science* (2008), pp. 453–469.
42. M. Hillery, V. Bužek, A. Berthiaume, Quantum secret sharing. *Phys. Rev. A* **59**, 1829–1834 (1999). doi: 10.1103/PhysRevA.59.1829
43. R. Cleve, D. Gottesman, H.-K. Lo, How to share a quantum secret. *Phys. Rev. Lett.* **83**, 648–651 (1999). doi: 10.1103/PhysRevLett.83.648
44. C. Crépeau, D. Gottesman, A. Smith, *Proceedings of EUROCRYPT* (2005), pp 285–301.
45. M. Christandl, S. Wehner, Proceedings of *ASIACRYPT* (2005), pp 217–235.
46. A. Ambainis, H. Buhrman, Y. Dodis, H. Röhrig, *Proceedings of IEEE Complexity* 10.1109/CCC.2004.1313848 (2004).
47. S. Barz *et al*., Demonstration of blind quantum computing. *Science* **335**, 303–308 (2012). doi: 10.1126/science.1214707; pmid: 22267806
48. N. H. Nickerson, J. F. Fitzsimons, S. C. Benjamin, Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. *Phys. Rev. X* **4**, 041041 (2014). doi: 10.1103/PhysRevX.4.041041
49. C. Mochon, arXiv:0711.4114 [quant-ph] (2007).
50. A. Chailloux, I. Kerenidis, *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on* (IEEE, 2009), pp. 527–533.
51. C. Crépeau, D. Gottesman, A. Smith, *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC '02 (ACM, 2002), pp. 643–652.
52. M. Burrows, *Proceedings of the 7th symposium on Operating systems design and implementation* (USENIX Association, 2006), pp. 335–350.
53. V. S. Denchev, G. Pandurangan, Distributed quantum computing. *ACM SIGACT News* **39**, 77 (2008). doi: 10.1145/1412700.1412718
54. M. Ben-Or, A. Hassidim, *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing* (2005), pp. 481–485.
55. S. Tani, H. Kobayashi, K. Matsumoto, *Proceedings of STACS: 22nd Annual Symposium on Theoretical Aspects of Computer Science* (2005), pp. 581–592.
56. M. Ganz, Quantum leader election. *Quantum Inform. Process.* **16**, 73 (2017). doi: 10.1007/s11128-017-1528-8
57. N. Aharon, J. Silman, Quantum dice rolling: A multi-outcome generalization of quantum coin flipping. *New J. Phys.* **12**, 033027 (2010). doi: 10.1088/1367-2630/12/3/033027
58. H. Buhrman, R. Cleve, S. Massar, R. de Wolf, Nonlocality and communication complexity. *Rev. Mod. Phys.* **82**, 665–698 (2010). doi: 10.1103/RevModPhys.82.665
59. H. Buhrman, R. Cleve, J. Watrous, R. de Wolf, Quantum fingerprinting. *Phys. Rev. Lett.* **87**, 167902 (2001). doi: 10.1103/PhysRevLett.87.167902; pmid: 11690244
60. J. M. Arrazola, N. Lütkenhaus, Quantum fingerprinting with coherent states and a constant mean number of photons. *Phys. Rev. A* **89**, 062305 (2014). doi: 10.1103/PhysRevA.89.062305
61. C. Crépeau, L. Salvail, J.-R. Simard, A. Tapp, *Proceedings of ASIACRYPT* (2011), pp. 407–430.
62. M. Peev *et al*., The SECOQC quantum key distribution network in Vienna. *New J. Phys.* **11**, 075001 (2009). doi: 10.1088/1367-2630/11/7/075001
63. M. Sasaki *et al*., Field test of quantum key distribution in the Tokyo QKD Network. *Opt. Express* **19**, 10387–10409 (2011). doi: 10.1364/OE.19.010387; pmid: 21643295
64. D. Stucki *et al*., Long-term performance of the SwissQuantum quantum key distribution network in a field environment. *New J. Phys.* **13**, 123001 (2011). doi: 10.1088/1367-2630/13/12/123001
65. S. Wang *et al*., Field and long-term demonstration of a wide area quantum key distribution network. *Opt. Express* **22**, 21739–21756 (2014). doi: 10.1364/OE.22.021739; pmid: 25321550
66. R. Courtland, *IEEE Spectr.* **53**, 11 (2016).
67. G. Vallone *et al*., Experimental satellite quantum communications. *Phys. Rev. Lett.* **115**, 040502 (2015). doi: 10.1103/PhysRevLett.115.040502; pmid: 26252672
68. J. Yin *et al*., Satellite-based entanglement distribution over 1200 kilometers. *Science* **356**, 1140–1144 (2017). doi: 10.1126/science.aan3211; pmid: 28619937
69. C. H. Bennett *et al*., Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996). doi: 10.1103/PhysRevLett.76.722; pmid: 10061534
70. D. Deutsch *et al*., Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.* **77**, 2818–2821 (1996). doi: 10.1103/PhysRevLett.77.2818; pmid: 10062053
71. W. Dür, H.-J. Briegel, J. Cirac, P. Zoller, Quantum repeaters based on entanglement purification. *Phys. Rev. A* **59**, 169–181 (1999). doi: 10.1103/PhysRevA.59.169
72. D. L. Moehring *et al*., Entanglement of single-atom quantum bits at a distance. *Nature* **449**, 68–71 (2007). doi: 10.1038/nature06118; pmid: 17805290
73. S. Ritter *et al*., An elementary quantum network of single atoms in optical cavities. *Nature* **484**, 195–200 (2012). doi: 10.1038/nature11023; pmid: 22498625
74. J. Hofmann *et al*., Heralded entanglement between widely separated atoms. *Science* **337**, 72–75 (2012). doi: 10.1126/science.1221856; pmid: 22767924
75. H. Bernien *et al*., Heralded entanglement between solid-state qubits separated by three metres. *Nature* **497**, 86–90 (2013). doi: 10.1038/nature12016; pmid: 23615617
76. A. Delteil *et al*., Generation of heralded entanglement between distant hole spins. *Nat. Phys.* **12**, 218–223 (2016). doi: 10.1038/nphys3605
77. R. Stockill *et al*., Phase-tuned entangled state generation between distant spin qubits. *Phys. Rev. Lett.* **119**, 010503 (2017). doi: 10.1103/PhysRevLett.119.010503; pmid: 28731764
78. B. Hensen *et al*., Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015). doi: 10.1038/nature15759; pmid: 26503041
79. D. Hucul *et al*., Modular entanglement of atomic qubits using photons and phonons. *Nat. Phys.* **11**, 37–42 (2015). doi: 10.1038/nphys3150
80. W. Pfaff *et al*., Quantum information. Unconditional quantum teleportation between distant solid-state quantum bits. *Science* **345**, 532–535 (2014). doi: 10.1126/science.1253512; pmid: 25082696
81. C. J. Ballance *et al*., Hybrid quantum logic and a test of Bell's inequality using two different atomic isotopes. *Nature* **528**, 384–386 (2015). doi: 10.1038/nature16184; pmid: 26672554
82. T. R. Tan *et al*., Multi-element logic gates for trapped-ion qubits. *Nature* **528**, 380–383 (2015). doi: 10.1038/nature16186; pmid: 26672553
83. I. V. Inlek, C. Crocker, M. Lichtman, K. Sosnova, C. Monroe, Multispecies trapped-ion node for quantum networking. *Phys. Rev. Lett.* **118**, 250502 (2017). doi: 10.1103/PhysRevLett.118.250502; pmid: 28696766
84. A. Reiserer *et al*., Robust quantum-network memory using decoherence-protected subspaces of nuclear spins. *Phys. Rev. X* **6**, 021040 (2016). doi: 10.1103/PhysRevX.6.021040
85. N. Kalb *et al*., Entanglement distillation between solid-state quantum network nodes. *Science* **356**, 928–932 (2017). doi: 10.1126/science.aan0070; pmid: 28572386
86. S. Tanzilli *et al*., A photonic quantum information interface. *Nature* **437**, 116–120 (2005). doi: 10.1038/nature04009; pmid: 16136138
87. S. Zaske *et al*., Visible-to-telecom quantum frequency conversion of light from a single quantum emitter. *Phys. Rev. Lett.* **109**, 147404 (2012). doi: 10.1103/PhysRevLett.109.147404; pmid: 23083285
88. R. W. Andrews *et al*., Bidirectional and efficient conversion between microwave and optical light. *Nat. Phys.* **10**, 321–326 (2014). doi: 10.1038/nphys2911
89. J. Bochmann, A. Vainsencher, D. Awschalom, A. N. Cleland, Nanomechanical coupling between microwave and optical photons. *Nat. Phys.* **9**, 712–716 (2013). doi: 10.1038/nphys2748
90. S. Probst *et al*., Anisotropic rare-earth spin ensemble strongly coupled to a superconducting resonator. *Phys. Rev. Lett.* **110**, 157001 (2013). doi: 10.1103/PhysRevLett.110.157001; pmid: 25167299
91. W. J. Munro, K. Azuma, K. Tamaki, K. Nemoto, Inside quantum repeaters. *IEEE J. Sel. Top. Quantum Electron.* **21**, 78 (2015). doi: 10.1109/JSTQE.2015.2392076
92. N. Sangouard, C. Simon, H. De Riedmatten, N. Gisin, Quantum repeaters based on atomic ensembles and linear optics. *Rev. Mod. Phys.* **83**, 33–80 (2011). doi: 10.1103/RevModPhys.83.33
93. N. Kalb, A. Reiserer, S. Ritter, G. Rempe, Heralded storage of a photonic quantum bit in a single atom. *Phys. Rev. Lett.* **114**, 220501 (2015). doi: 10.1103/PhysRevLett.114.220501; pmid: 26196608
94. C. Kurz *et al*., Experimental protocol for high-fidelity heralded photon-to-atom quantum state transfer. *Nat. Commun.* **5**, 5527 (2014). doi: 10.1038/ncomms6527; pmid: 25413900
95. H. Tanji, S. Ghosh, J. Simon, B. Bloom, V. Vuletić, Heralded single-magnon quantum memory for photon polarization States. *Phys. Rev. Lett.* **103**, 043601 (2009). doi: 10.1103/PhysRevLett.103.043601; pmid: 19659349
96. A. Delteil, Z. Sun, S. Fält, A. Imamoğlu, Realization of a cascaded quantum system: Heralded absorption of a single photon qubit by a single-electron charged quantum dot. *Phys. Rev. Lett.* **118**, 177401 (2017). doi: 10.1103/PhysRevLett.118.177401; pmid: 28498703
97. K. Azuma, K. Tamaki, H.-K. Lo, All-photonic quantum repeaters. *Nat. Commun.* **6**, 6787 (2015). doi: 10.1038/ncomms7787; pmid: 25873153
98. M. Pant, H. Krovi, D. Englund, S. Guha, Rate-distance tradeoff and resource costs for all-optical quantum repeaters. *Phys. Rev. A* **95**, 012304 (2017). doi: 10.1103/PhysRevA.95.012304
99. S. Muralidharan, J. Kim, N. Lütkenhaus, M. D. Lukin, L. Jiang, Ultrafast and fault-tolerant quantum communication across long distances. *Phys. Rev. Lett.* **112**, 250501 (2014). doi: 10.1103/PhysRevLett.112.250501; pmid: 25014798
100. W. J. Munro, A. M. Stephens, S. J. Devitt, K. A. Harrison, K. Nemoto, Quantum communication without the necessity of quantum memories. *Nat. Photonics* **6**, 777–781 (2012). doi: 10.1038/nphoton.2012.243
101. I. Schwartz *et al*., Deterministic generation of a cluster state of entangled photons. *Science* **354**, 434–437 (2016). doi: 10.1126/science.aah4758; pmid: 27608669
102. T. Rudolph, Why I am optimistic about the silicon-photonic route to quantum computing. *APL Photonics* **2**, 030901 (2017). doi: 10.1063/1.4976737
103. V. G. Cerf, R. E. Kahn, A protocol for packet network intercommunication. *IEEE Trans. Commun.* **22**, 637–648 (1974). doi: 10.1109/TCOM.1974.1092259
104. R. Van Meter, J. Touch, Designing quantum repeater networks. *IEEE Commun. Mag.* **51**, 64 (2013). doi: 10.1109/MCOM.2013.6576340
105. C. Pfister, M. A. Rol, A. Mantri, M. Tomamichel, S. Wehner, Capacity estimation and verification of quantum channels with arbitrarily correlated errors. *Nat. Commun.* **9**, 27 (2018). doi: 10.1038/s41467-017-00961-2; pmid: 29295975
106. M. Takeoka, S. Guha, M. M. Wilde, Fundamental rate-loss tradeoff for optical quantum key distribution. *Nat. Commun.* **5**, 5235 (2014). doi: 10.1038/ncomms6235; pmid: 25341406

107. S. Pirandola, R. Laurenza, C. Ottaviani, L. Banchi, Fundamental limits of repeaterless quantum communications. *Nat. Commun.* **8**, 15043 (2017). doi: 10.1038/ncomms15043; pmid: 28443624
108. L.-M. Duan, M. D. Lukin, J. I. Cirac, P. Zoller, Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001). doi: 10.1038/35106500; pmid: 11719796
109. C. Simon *et al*., Quantum repeaters with photon pair sources and multimode memories. *Phys. Rev. Lett.* **98**, 190503 (2007). doi: 10.1103/PhysRevLett.98.190503; pmid: 17677612
110. L. Jiang *et al*., Quantum repeater with encoding. *Phys. Rev. A* **79**, 032325 (2009). doi: 10.1103/PhysRevA.79.032325
111. J. J. Wallman, S. T. Flammia, Randomized benchmarking with confidence. *New J. Phys.* **16**, 103032 (2014). doi: 10.1088/1367-2630/16/10/103032

ACKNOWLEDGMENTS

We thank J. Borregaard, E. Diamanti, D. Englund, R. A. Friedman, T. Northup, I. Kerenidis, W. Tittel, and all members of the Quantum Internet Alliance collaboration for feedback on earlier versions of this document. **Funding:** D.E. and S.W. are funded by an ERC Starting Grant (S.W.) and an NWO VIDI Grant (S.W.). R.H. is funded by an ERC Consolidator Grant and an NWO VICI Grant. R.H. and S.W. are also supported by the NWO Zwaartekracht Grant QSC. D.E., R.H., and S.W. are also funded by the EU H2020 FETFLAG Quantum Internet Alliance. **Competing interests:** There are no competing interests.

10.1126/science.aam9288

REVIEW SUMMARY

CONSERVATION

Landscapes that work for biodiversity and people

C. Kremen* and A. M. Merenlender

BACKGROUND: Biodiversity is under siege, with greatly enhanced rates of local and global extinction and the decline of once-abundant species. Current rates of human-induced climate change and land use forecast the Anthropocene as one of the most devastating epochs for life on earth. How do we handle the Anthropocene's triple challenge of preventing biodiversity loss, mitigating and adapting to climate change, and sustainably providing resources for a growing human population? The answer is in how we manage Earth's "working lands"; that is, farms, forests, and rangelands. These lands must be managed both to complement the biodiversity conservation goals of protected areas and to maintain the diverse communities of organisms, from microbes to mammals, that contribute to producing food, materials, clean water, and healthy soils; sequestering greenhouse gases; and buffering extreme weather events, functions that are essential for all life on Earth.

ADVANCES: Protected areas are the cornerstone of biodiversity conservation. Although the total area of protected regions needs to be increased, parks will nonetheless continue to lose species if these areas are isolated from one another by inhospitable land uses and are faced with a rapidly changing climate. Further, many species, such as those that migrate, remain unprotected as they occupy lands outside of parks for all or portions of their life cycles. Lastly, protected-area effectiveness is greatly influenced by surrounding land management. "Working lands conservation" aims to support biodiversity while providing goods and services for humanity over the long term, assuring sustainability and resilience. By managing lands surrounding parks favorably, working lands can buffer protected areas from threats and connect them to one another. This approach complements protected areas by providing accessory habitats and resources for some species while facilitating dispersal and climate change adaptation for others. Further, by maintaining the biodiversity that supplies critical ecosystem services within working lands, these approaches ensure that the production of food, fiber, fuel, and timber can be sustained over the long run and be more resilient to extreme events, such as floods, droughts, hurricanes, and pest and disease outbreaks, which are becoming more frequent with climate change. A variety of biodiversity-based land management techniques can be used in working lands, including agroforestry, silvopasture, diversified farming, and ecosystem-based forest management, to ensure sustainable production of food and fiber.

OUTLOOK: The underlying principle of biodiversity-based management of working lands has been practiced since ancient times. Today, these systems have largely been replaced by unsustainable resource extraction, rather than serving as models that could be adapted to modern conditions. Although various regulatory, voluntary, and financial tools exist to promote sustainable land management, many barriers prevent individuals, communities, and corporations from adopting biodiversity-based practices, including deeply entrenched policy and market conditions that favor industrialized or extractive models of land use. Thus, uptake of these approaches has been patchy and slow and is not yet sufficient to create change at the temporal and spatial scales needed to face the triple Anthropocene threat.

Biodiversity-based land management practices are knowledge- rather than technology-intensive. They are well adapted to empower local communities to manage their natural resources. One of the most exciting emerging trends is community-driven initiatives to manage working landscapes for conservation and sustainability. By linking up through grassroots organizations, social movements, and public-private partnerships, these initiatives can scale up to create collective impact and can demand changes in government policies to facilitate the conservation of working lands. Scientists and conservation practitioners can support these initiatives by engaging with the public, listening to alternative ways of knowing, and cocreating landscapes that work for biodiversity and people. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau6020

TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ckremen@berkeley.edu
Cite this article as C. Kremen and A. M. Merenlender, *Science* **362**, eaau6020 (2018). DOI: 10.1126/science.aau6020

Strawberry production in Central Coast, California. On the left, a homogeneous landscape of strawberry monoculture, including organic fields, supports fewer wild species then a diversified, organic farm (right) in the same region, which includes a small field of strawberry, surrounded by orchards, hedgerows, diverse vegetable crops, and natural habitats. The monoculture landscape creates barriers to wildlife dispersal, whereas the diversified landscape is more permeable.

PHOTO: C. KREMEN

REVIEW

CONSERVATION

Landscapes that work for biodiversity and people

C. Kremen* and A. M. Merenlender

How can we manage farmlands, forests, and rangelands to respond to the triple challenge of the Anthropocene—biodiversity loss, climate change, and unsustainable land use? When managed by using biodiversity-based techniques such as agroforestry, silvopasture, diversified farming, and ecosystem-based forest management, these socioeconomic systems can help maintain biodiversity and provide habitat connectivity, thereby complementing protected areas and providing greater resilience to climate change. Simultaneously, the use of these management techniques can improve yields and profitability more sustainably, enhancing livelihoods and food security. This approach to "working lands conservation" can create landscapes that work for nature and people. However, many socioeconomic challenges impede the uptake of biodiversity-based land management practices. Although improving voluntary incentives, market instruments, environmental regulations, and governance is essential to support working lands conservation, it is community action, social movements, and broad coalitions among citizens, businesses, nonprofits, and government agencies that have the power to transform how we manage land and protect the environment.

Biodiversity, the product of 3.8 billion years of evolution, is under siege. Not only are both marine and terrestrial species experiencing accelerated rates of local and global extinction (*1*–*3*), but even common species are declining (*2*, *4*, *5*). This alarming situation has prompted a strong call for increasing the number (*6*, *7*) and effectiveness (*8*) of protected areas, the principal method for combatting species loss. Though such protections are essential, we cannot rely on protected areas alone to preserve species. As protected areas become increasingly isolated because of habitat loss and degradation, much research has revealed that they will lose species over time (*9*). Further, many critical threats to species do not respect protected-area boundaries (*10*), including climate change, which both exacerbates species losses (*11*) and threatens to alter the biomes of many currently protected regions entirely (*12*).

More hopefully, recent studies show that some human-dominated landscapes can support much more biodiversity than previously recognized (*13*–*17*), suggesting a complementary path forward. Specifically, when these areas, generally referred to as the "matrix," represent a high-quality mosaic of land uses, they can play a critical role in sustaining biodiversity, both in situ and by promoting species dispersal among protected areas and remnant habitats and along migratory routes (Fig. 1) (*15*, *18*, *19*). Of course, human survival also depends on the long-term capacity of this matrix of "working lands," including rangelands, forests, and farms, to produce food, water, fiber, fuel, and forest products. All too often, however, these goods are produced at severe environmental cost, including habitat degradation, toxic contamination, and depletion of water quantity and quality, leading to ecological collapse, local extinctions, and the creation of unproductive wastelands (*20*, *21*). We argue that, instead, working lands can be used to support high levels of biodiversity while satisfying human needs in a sustainable way. Because rangelands, forests, and cultivated lands collectively occupy ~80% of terrestrial area (*21*), the potential for conservation in such lands is enormous.

Critical ecosystem functions and services are provided by a suite of diverse organisms, from microbes to mammals, and thus maintenance of these organisms is necessary for long-term and sustainable productivity of working lands (*22*, *23*). Hence, managing the matrix to maintain biodiversity is not only necessary for species conservation but also essential for sustainable production. Biodiversity-based production systems, including agroecological farming or ecosystem-based forest management, are often perceived as unproductive, an incorrect viewpoint that impedes the public investment needed to develop and promote these methods. Here, we describe managing the matrix jointly and sustainably for biodiversity and people through "working lands conservation" and ask what strategies can be used to strengthen and scale up this approach as rapidly as possible to help combat the triple Anthropocene threats of biodiversity loss, climate change, and unsustainable land use.

Working lands conservation defined

Although the term "working lands conservation" is already used in policy statements and in guidance for conservation programs [e.g., (*24*)], the concept has yet to be formally defined and risks being misapplied. We define it at the landscape scale (Box 1).

To avoid mass extinction and ecosystem collapse, we must integrate biodiversity conservation into the landscapes we use and not simply relegate nature to a limited number of protected areas that are doomed if left as isolated habitat islands within biological deserts. Working lands can provide food, breeding sites, and shelter for a myriad of species while maintaining abiotic conditions, including temperature, light, wind, water, fire, and other disturbance processes, within required ranges. They can facilitate functional connectivity—that is, the movement of organisms across the landscape and among habitat patches that promotes population persistence by allowing for gene flow, recolonization, and adaptation to climate and other global changes (*25*, *26*).

To support humanity sustainably, a working landscape must be productive and maintain the ecosystem services, such as pollination, pest control, and nutrient cycling, that underlie that production. Maintaining these services requires supporting the underlying populations of service-providing organisms. Within each service, a greater diversity of service providers often enhances the level and/or quality of services and reduces uncertainty in service delivery (*22*), because different species respond differentially to environmental change (*27*, *28*). Maintaining connectivity is also important, both to support flows of ecosystem service providers and/or materials (e.g., pollination requires animal vectors to move pollen between flowers; water purification requires water to flow through vegetation) (*29*) and to enhance metacommunity persistence of service-providing organisms to sustain ecosystem functions and services over space and time (*22*, *30*).

Box 1. Definition of working lands conservation.

Definition: Conservation in working landscapes maintains biodiversity, provides goods and services for humanity, and supports the abiotic conditions necessary for sustainability and resilience.

These socioecological systems both support biodiversity by providing critical resources and rely on biodiversity (specifically, ecosystem service providers) for sustainable production of food, water, fiber, fuel, and forest products. These landscapes also enhance connectivity to promote the movement of organisms, natural processes, and ecosystem services.

Working lands conservation emphasizes the critical role of managing the matrix for species conservation to complement protected areas.

Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA 94720, USA.
*Corresponding author. Email: ckremen@berkeley.edu

Fig. 1. Rebuilding connectivity in the matrix by using silvopasture. Photo of Finca La Luisa showing several types of silvopastoral systems, including regenerating secondary tropical dry forest trees with grass understory (yellow) and rows of planted *Eucalyptus* trees interspersed with nitrogen-fixing *Leucaena leucocephala* fodder shrubs and forage grasses (blue). These systems were established on former monoculture agricultural lands to restore compacted, degraded soils; the red area shows early stages of tropical dry forest regeneration prior to grass seeding for silvopasture. Silvopastures produce more cattle sustainably on less land, buffer ranchers from economic losses due to climate extremes, and create landscape connectivity to other forest fragments (orange) in the Cesar river valley, Colombia.

PHOTO: J. J. LOPERA/CIPAV; ADAPTED BY NIRJA DESAI/*SCIENCE*

Ensuring the sustainability of production requires balancing across provisioning, regulating, and supporting services; in other words, seeking multifunctionality and stability rather than maximal production. For example, conventional (chemically intensive) monoculture agriculture produces high yields but often at the expense of water quality, climate regulation, and soil health (Fig. 2A) (*20*) and can suffer production collapse in response to periodic extreme weather, pests, and diseases (*31*–*33*). Although transforming to a more sustainable system may reduce average yields somewhat [e.g., (*34*)], by relying on ecosystem services produced on the farm and in the surrounding landscape, a sustainable system is both multifunctional and more resilient to change (*20*, *31*) (Fig. 2C).

Working landscapes often comprise heterogeneous patch types, including novel communities made up of mixtures of native and nonnative species, as well as remnants of natural or semi-natural habitats whose composition is more similar to that of a historical ecological community (*35*). Although management goals likely will differ among patch types, both individual patches and the whole landscape should be managed for sustainability. For example, patches whose communities are far from historical could be managed principally for crops (a provisioning service) by using sustainable agricultural practices to minimize negative effects on biodiversity and ecosystem services on and off site. Remnant patches could be retained as stepping-stone habitats to support species dispersal and provide regulating services such as pollination (*29*, *31*). Maintaining mosaic landscapes composed of different patch types provides opportunities to maximize diversity, resilience, and multifunctionality. Radar diagrams reveal likely trade-offs and sustainability within and across patches (Fig. 2B), as well as multifunctionality at the landscape scale (Fig. 2C).

Conservation in working landscapes draws upon several related concepts. Integrated landscape management initiatives seek to simultaneously improve food production, biodiversity or ecosystem conservation, and rural livelihoods and are being implemented by governments and nongovernmental organizations in Latin America and Africa (*36*). The ecosystem stewardship concept focuses on the need to sustain Earth's capacity to provide ecosystem services and support socioecological resilience under conditions of uncertainty and change (*27*). The socioecological production landscape of the Japan Satoyama Satoumi Assessment refers to dynamic landscape mosaics that have been shaped over time by the interactions between people and nature in ways that jointly support biodiversity and human well-being (*37*). These concepts also emphasize critical social components, such as involving multiple stakeholders at the landscape scale, community participation, intersectoral coordination, flexible and adaptive governance systems, social learning, and adaptive management, which are necessary for successful conservation of working landscapes.

The underlying principle of maintaining ecological diversity inherent to these approaches has been practiced since ancient times. Some of these management systems, such as indigenous use of fire, weeding, pruning, and the seed dispersal that shaped Californian ecosystems (*38*), no longer exist in their original form, whereas others, such as regional pastoral and high-mountain farming systems in Europe (*39*), persist in some areas. By creating highly simplified and intensified production systems (*21*, *40*), from corn and soy in U.S. midwestern states to palm oil plantations in southeast Asia and vineyards in Chile, we have abandoned this critical sustainability principle across much of Earth's cultivated landscapes. However, it is a fallacy that such systems will ultimately spare more land for nature conservation or feed the world indefinitely; rather, we need to find ways to allow biodiversity-based production methods to figure much more prominently in local, regional, and global markets (*16*).

Working lands conservation as a complement to protected areas

Given the dire situation facing many species and the expectation of further species losses and shifts in ecosystem composition due to climate change (*2*, *4*, *11*), ceasing further habitat conversion completely and protecting large regions of Earth effectively are critical necessities for conservation (*6*–*8*), although just how much should be protected is highly debated (*41*). [By "protected area," we refer to parks whose primary function is to conserve biodiversity and wilderness (International Union for Conservation of Nature and Natural Resources categories I to IV, constituting 6.75% of terrestrial area) (*42*), in contrast to areas blending conservation and livelihood objectives (categories V to VI, constituting 8.65%).] However, the protected-area strategy alone will not be successful without complementary working lands conservation in the surrounding landscapes. First, even the largest protected areas will lose species over the long term (*9*) unless surrounding landscapes can be managed to provide connectivity among parks. Further, less than 10% of protected areas are expected to represent current climatic conditions within 100 years, increasing the criticality of matrix connectivity to permit species to follow their suitable climates (*12*). Lastly, effectiveness in controlling threats, such as invasive species, encroachment, poaching, and other impacts on protected lands, also critically depends on the surrounding matrix (*10*). Thus, to stem the tide of biodiversity loss, we must expand beyond protected areas, using working lands conservation both to buffer and to reduce the threats that cross park boundaries and to create accessory habitats for both movement and persistence.

Working lands conservation is a key linchpin for combatting the triple Anthropocene challenge of biodiversity loss, climate change, and unsustainable land use. A large-scale example is the Mesoamerican Biological Corridor project, which has fostered a multistakeholder participatory process to enhance connectivity on cultivated, range, and forest working lands to link

more than 650 protected areas in the region (*43*). A concurrent goal is to use sustainable agriculture and forestry techniques to promote livelihoods and enhance resilience to climate change (*36*). Protected areas are vital in this region because many species are restricted to forest; however, most reserves are small and isolated. In combination with steep elevational and latitudinal gradients in the region, this isolation makes species inhabiting reserves particularly vulnerable to climate change. The Mesoamerican Biological Corridor project recognizes the role that working lands can play to restore critical connectivity by increasing tree diversity and cover through live fences, agroforestry, silvopasture, forest fallows, home gardens, and protection or restoration of riparian forests and forest fragments (*43*). These forest elements, which include both ribbonlike and patch structures, support a large number of neotropical birds, insects, mammals, and plants (*17*, *44*); enhance the movement of birds and bats across the landscape (*45*–*47*); and thus contribute to conservation, even of vulnerable wildlife (*17*, *47*, *48*). Forest elements also promote sustainable land use and contribute to local livelihoods by supporting ecosystem services. For example, evidence suggests that an economically devastating invasive pest, the coffee berry borer, is reduced by the integration of forest elements within coffee landscapes, which both limits the borer's ability to colonize new coffee fields (*49*) and promotes bird species that prey on the borer (*50*). Reduced economic losses due to pest control from birds are similar in magnitude to average per capita income in the region and are strongly related to forest cover (*50*). Adopting sustainable agricultural techniques and enhancing tree cover simultaneously creates more flexible and resilient production systems that allow farmers and ranchers to adapt to extreme conditions prompted by climate change (*33*, *51*). Although some critics decry the effectiveness of the Mesoamerican Biological Corridor project, it may be too early to judge. Quite a few integrated landscape initiatives are concentrated in the region, in association with biological corridors (*36*). However, many began relatively recently, and we know from the few scientific studies that exist that developing an effective multistakeholder participatory process takes substantial time (*36*, *43*, *52*). In one case that is more advanced (the San Juan–La Selva Biological Corridor in Costa Rica), some success has been achieved in arresting deforestation and encouraging tree planting, forest regeneration, and connectivity through a government-run payments for ecosystem services program, as well as other grassroots initiatives (*43*, *53*).

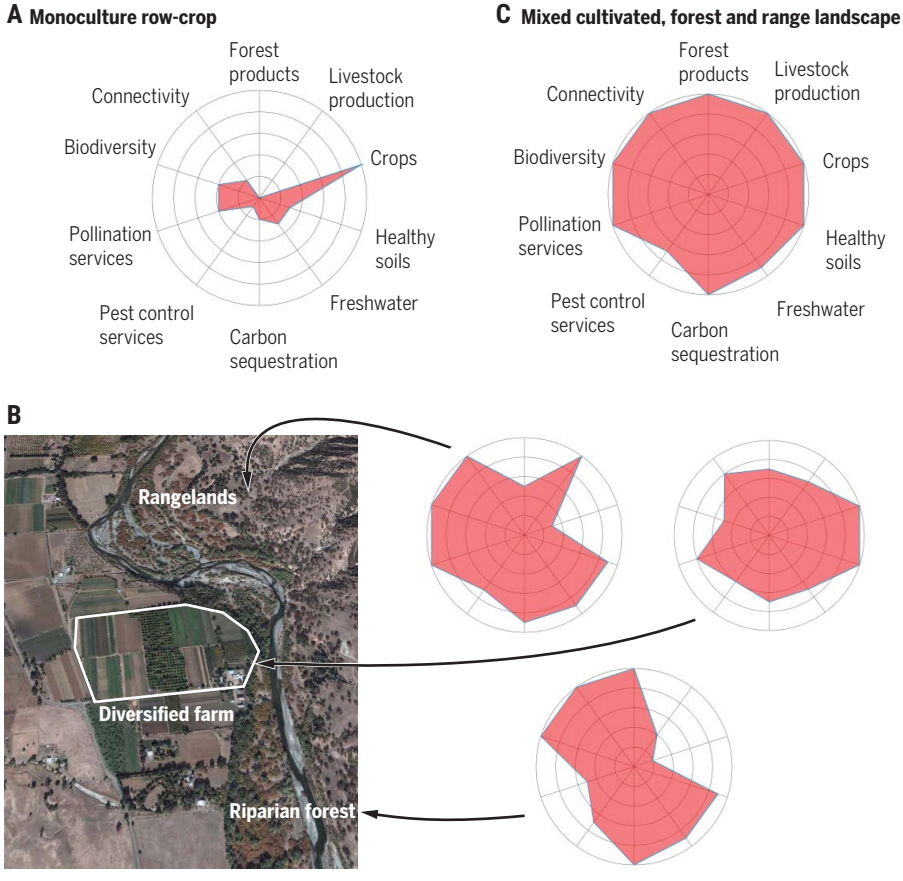


ILLUSTRATION: NIRJA DESAI/*SCIENCE* BASED ON C. KREMEN AND A.M. MERENLENDER; PHOTO: NATIONAL AERIAL IMAGERY PROGRAM. ADAPTED BY NIRJA DESAI/*SCIENCE*

Fig. 2. Ecosystem service trade-offs with land management. Radar diagrams display how different land uses affect various ecosystem services and biodiversity. (**A**) Monoculture row cropping contributes to food production at the expense of other ecosystem services and biodiversity. (**B**) In a working landscape managed for conservation, patch types differ in the services they provide, but each patch type should display a relatively even array of services, minimizing trade-offs. (**C**) Across patches, the services provided for the working landscape in (B) are multifunctional.

Mechanisms for promoting working lands conservation

The challenge of shifting from managing working lands solely for profit to conservation of working lands is not insignificant, but there are clear paths toward larger-scale integration of this approach. These strategies include various regulatory, voluntary, incentive, market-based, or governance instruments (table S1), which vary in their applicability to private, communal, or state-owned lands and the extent to which they support biodiversity conservation versus livelihoods or economies (Fig. 3A). Each approach has challenges, especially around reconciling conservation and socioeconomic objectives (table S1) (*42*, *54*). Collectively, problems associated with regulatory and incentive programs can include inter alia lack of permanence or compliance, complex implementation, unintended economic consequences, low adoption rates, high monitoring costs, and little evaluation of effectiveness against goals (table S1).

Further, there is often the risk that the biodiversity conserved through these actions is not equivalent to that which was lost because of economically driven land conversion. Instruments for private lands may result in piecemeal land management actions that have little positive effect on biodiversity at the landscape scale; promising public-private initiatives to overcome this defect include corridor planning (*43*, *55*) (Box 2 and Fig. 4) and landscape-level mitigation (table S1). For example, landowners required to set aside forest on their properties under Brazil's forest code may develop these lands in exchange for mitigating lands elsewhere within the same biome that provide greater conservation value (*56*). Managing the matrix to promote biodiversity could also exacerbate human-wildlife conflict; however, the recovery of carnivore populations within human-dominated areas in Europe provides a hopeful and inspiring example for how landscapes can be shared between wildlife and people (*14*) (Box 3). These instruments can exacerbate the unequal distribution of benefits and costs within and across communities (table S1). For example, trading development rights on forestlands in exchange for permitting high-density urban development elsewhere can provide open spaces for working lands conservation. However, such trades could exacerbate the lack of access to open space already experienced by low-income urban households. Thus, the effects of conservation measures on social equity and environmental justice should also be considered (*57*). A final concern is that there is often a trade-off between the rigor of environmental standards or restrictions enforced and the likelihood of adoption (table S1); incentive schemes that are flexible, provide obvious

benefits, target likely adopters, fit the sociocultural context, foster enabling market and regulatory environments, and provide technical assistance may boost adoption (*58*). For example, payments for conserving or restoring forests in Costa Rica are based on area, whereas transaction costs are the same regardless of size, disincentivizing smaller landowners from participating in the payments for ecosystem services scheme. Encouraging smallholders to participate would require adjusting the costs of participation so that these landowners could also realize net gains (*53*). Although numerous changes are required, careful attention to the construction of these programs could increase their success.

Further, several current trends favor working lands conservation approaches. First, new policy instruments [such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation)] operating across a range of scales, from individual private landholdings to large-scale community-based or government-funded initiatives, are being developed to incentivize conservation on working lands. Second, the number and variety of institutions involved in working lands conservation are increasing, and such institutions include both public-private partnerships and nongovernmental conservation organizations that formerly focused primarily on protected areas (*36*, *59*, *60*). Third, these institutions can take advantage of recent increases in both public and private "investments for conservation" (investments designed to cogenerate financial returns and conservation benefits) (*60*). Such investments include projects in sustainable food and fiber production, water quality and quantity projects, and outright habitat conservation (in the latter, financial returns are based on changing land values or carbon stocks). Fourth, outside of these investments, an increasing number of companies have committed to greening their supply chains by reducing the environmental impacts at the source, processing, delivery, and end-of-life management of the product (*61*). Although supply chain greening requires much better monitoring, accountability, and inclusion of biodiversity conservation as an explicit goal (*61*, *62*), it could ultimately contribute to conservation in working landscapes, particularly given the vast economic power represented within corporations (*61*). A final trend is the creation of voluntary, community-driven programs (Box 2) in which local communities participate in the conservation of working landscapes to gain increased access to information and expertise, build interpersonal connections, and obtain both personal benefits and public recognition for practicing sustainable methods (*63*).

We argue that this latter trend of community-based actions and the innovations, networks, and social movements that sometimes emerge from them present the most exciting opportunity to turn the tide against the triple Anthropocene threat [see also (*64*)]. Communities seeking solutions for socioecological resilience frequently rely on working lands conservation approaches. For example, Sustainable Solutions restores mangrove forests in Sri Lanka and India through youth-based community engagement to build shoreline resilience to cyclones while enhancing livelihoods from fisheries dependent on mangrove ecosystems.

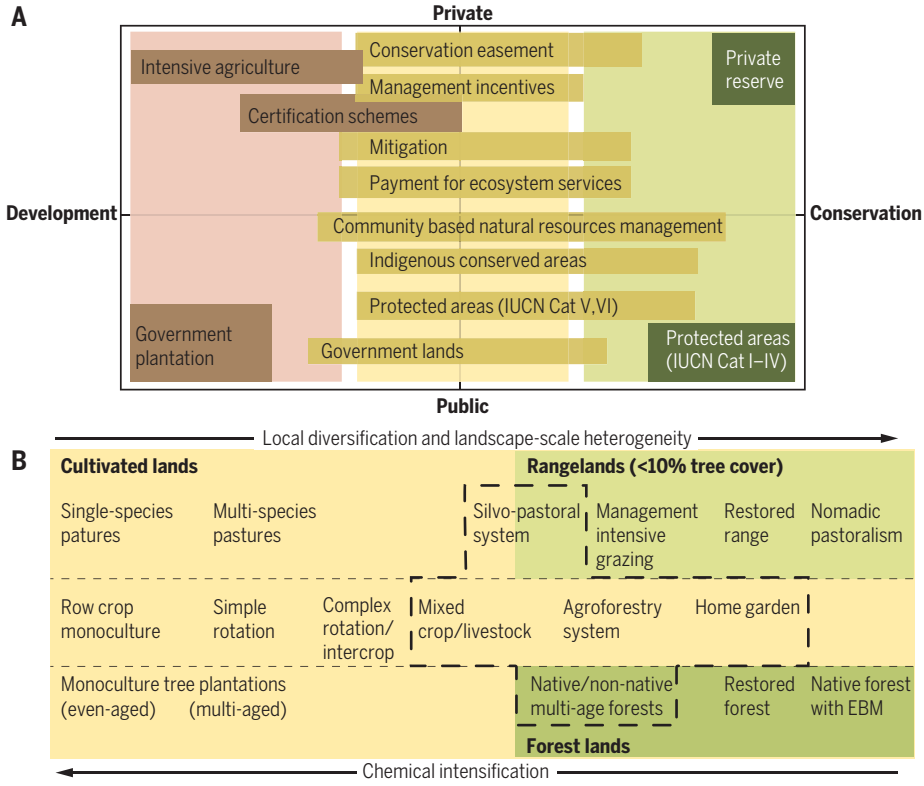


Fig. 3. Approaches for conservation of working lands occupy the space (yellow) between highly developed (brown) and highly conserved (green) land uses. (**A**) An array of tools are available for working lands conservation, for private, communal, or public lands (see table S2 for more detail and examples). IUCN Cat. International Union for Conservation of Nature and Natural Resources categories. (**B**) Forms of management for forage, crops, and tree products from cultivated lands (yellow), rangelands (light green), and forests (dark green), arrayed roughly along a management gradient of diversification (left to right) or chemical intensification (right to left). Cultivated lands include all planted systems. Dashed lines indicate overlapping concepts. EBM, ecosystem-based management.

Box 2. Community stewardship: The case of Landcare Australia.

The Landcare movement is a well-documented community stewardship effort begun in the mid-1980s to conserve biodiversity and sustain agriculture in Australia, resulting in more than 5000 Landcare and Coastcare groups. More than 20 countries have since adopted the model. In Australia, this model combines substantial government investment with landowner and community engagement. For example, Landcare groups across eastern Australia contribute to the delivery of the Great Eastern Ranges (GER) Initiative (*105*), alongside public land management authorities, conservation organizations, research institutions, and traditional owners groups. The GER is one of Australia's largest public-private partnerships to conserve biodiversity in the face of climate change (Fig. 4) as part of Australia's National Wildlife Corridors Plan. Landcare groups along the corridor undertake restoration and management activities, along with community building and engagement. In the Queanbeyan Landcare group, 25 landholders signed up to increase the foraging habitat for the glossy black cockatoo (*Calyptorhynchus lathami*) through the restoration of 10,000 she-oaks (*Allocasuarina* sp.) in production lands along three river catchments. The social networks and learning spaces created are promising ways of encouraging conservation commitment among land managers. However, far more landowners must become engaged to restore connectivity at the scale desired.

Further, local initiatives can link together to form larger networks with the help of boundary organizations to form social movements that can advance environmental policies, improve sustainable behaviors, and demand supply chain

ILLUSTRATION: NIRJA DESAI/*SCIENCE* BASED ON C. KREMEN AND A.M MERENLENDER

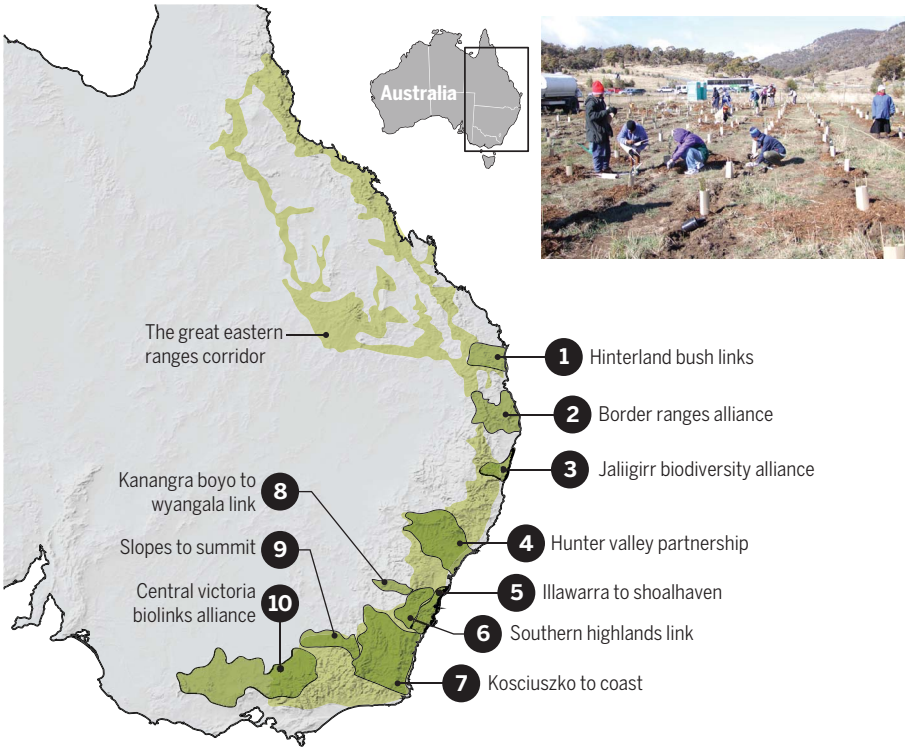


Fig. 4. The GER Corridor Initiative, Australia. The light green outline represents the plan to protect and restore more than 3,600 km^2 as a climate corridor. The numbered, dark green shapes denote regional alliances of conservation and natural resource management organizations, including Landcare communities (Box 2). In the photo, members of the Molonglo Catchment Group Landcare community conduct restoration.

Box 3. Carnivore conservation in shared landscapes.

Maintaining populations of large carnivores ranks among the greatest of conservation challenges. These area-demanding species require larger territories than most protected areas possess, potentially necessitating costly translocations to ensure gene flow and maintain populations. Further, these species conflict with people in surrounding matrices through predation on livestock or, occasionally, maiming or killing of humans. Nonetheless, in Europe, most large carnivore populations are stable or expanding. One-third of the area of mainland Europe hosts at least one permanent population of its four large carnivore species, persisting alongside moderate human densities and largely outside of protected areas. The success of carnivore conservation in Europe is attributed to well-enforced, coordinated legislative protection, improvements in habitat and ungulate prey base, and rural depopulation. Importantly, ranchers have found ways to live with carnivores by using carnivore-proofed electric fences and re-invigorating traditional livestock-guarding practices using shepherds and dogs (*14*). Similarly, in a cultivated region in India, large carnivore species (the leopard and striped hyena) persist with few conflicts despite high human densities (300 people/km^2) and the lack of wild prey (*106*), suggesting the potential that exists for carnivore conservation in shared landscapes.

accountability (*64*). For example, the withdrawal of the United States from the Paris Agreement at the 21st Conference of Parties (COP21) and delays in regulation of emissions by other nations galvanized a series of on-the-ground climate actions from civil society, businesses, nonprofits, and subnational government. The Global Action Climate Summit of 2018 instigated by California governor Jerry Brown illustrates a new stage of this growing social movement. Its Land and Ocean Stewardship "30 × 30" challenge brings together more than 100 organizations focused on managing forests, farmlands, and oceans to provide 30% of the climate change solution by 2030, rather than waiting on agreements among nation states that continue to fall short of the necessary carbon reduction targets. The land management techniques being developed locally to mitigate and adapt to climate change are generally consistent with the conservation of working lands approach [e.g., (*65*)].

The benefits of local land conservation can also be scaled up and made more effective if they are carried out within a landscape or regional conservation program organized by a state or nonprofit agency (*58*). Innovative social and institutional arrangements for working lands conservation may emerge, such as The Nature Conservancy's BirdReturns program in California. Through a reverse auction, the program finds and pays farmers willing to alter water management to create "pop-up" wetlands to provide habitats for shorebirds during their northward migration, selecting sites that optimize the conservation benefits relative to payments (*15*).

Management techniques for conserving working lands

Cultivated lands

Cultivated lands make up 12% of the terrestrial ice-free surface (*66*) and comprise row and forage crops, seeded pastures, vineyards and orchards, mixed crop and livestock systems, and tree crops and plantations (Fig. 3B). Cultivated lands are often highly simplified ecologically; thus, they rely extensively on chemical fertilizers and pesticides to replace ecosystem services formerly generated within or around agroecosystems (*31*), often creating negative consequences for the environment and human health (Fig. 2A) (*21*), including continued large-scale forest conversion in some areas of the biodiverse tropics (*62*). Instead, diversified farming systems using agroecological management practices operate by fostering biophysical conditions and ecological interactions favorable to crop production (*31*, *67*, *68*), producing a more balanced (sustainable) distribution of ecosystem services (Fig. 2B). Evidence also suggests that they minimize many of the negative environmental consequences associated with simplified farming (*31*) (Fig. 5). Further, these techniques can maintain crop yields and profitability; create new market opportunities; enhance food security, nutrition, and livelihoods; and contribute substantially to the global food supply, particularly under a changing climate (table S2). Because they rely on relatively low-cost, low-technology, knowledge-based methods (*69*), agroecological diversification techniques can be made accessible to the majority of farmers. [Small-scale farms with <5 ha make up 94% of farms worldwide (*40*) and produce more than half of world food crops (*70*).] These farming methods use open-pollinated seed varieties that can be saved and cultivars that are locally adapted; thus, they are less dependent on purchased seeds and other inputs that can lead to poverty traps (*71*). Multiple grassroots organizations and social movements support learning, sharing, and adaptation of agroecological knowledge and seeds through farmer-to-farmer networks under participatory governance (*64*). Diversified, agroecological practices are therefore farming methods that are highly compatible with working lands conservation, although potentially more applicable to certain farming systems. Large-scale

MAPS: GREAT EASTERN RANGES; PHOTO: ANDREW CAMPBELL; ADAPTED BY NIRJA DESAI/*SCIENCE*

commercial farmers that have invested heavily in the machinery associated with chemically intensive agriculture may not readily switch to agroecological techniques (*68*, *72*); however, the use of some agroecological techniques can be compatible with existing infrastructure and can lead to reduced agrochemical use at similar or even enhanced profits [e.g.,(*73*)].

A concern is that the use of "wildlife-friendly" agroecological practices will require more land to be farmed to produce the same amount of food, promoting deforestation and harming biodiversity (*74*). However, a number of diversified, agroecological farming methods maintain or increase yields (table S2) (*32*, *50*, *73*, *75*–*78*). For example, techniques such as intercropping, cover cropping, and crop rotation may promote crop yields through a variety of ecological mechanisms (*23*), including complementarity of water and nutrient use (e.g., different crops access different soil layers for water and nutrient uptake), facilitation of nutrient uptake [e.g., intercropped faba bean acidifies the soil, mobilizing phosphorus that is taken up by rice (*79*)], reduction of pests and diseases [e.g., pests and diseases spread more slowly in spatially or temporally heterogeneous crop systems, and such systems also support predator populations that keep pests in check (*80*, *81*)], and enhancement of soil biota and fertility (*82*). By improving soil structure and stability, which then enhances water infiltration and retention, these techniques also stabilize yields against annual environmental fluctuations and more catastrophic disturbances such as droughts and hurricanes (*32*, *33*).

Beyond providing resources and habitats for agrobiodiversity, specific techniques such as agroforestry and the use of silvopasture, hedgerows, flower strips, live fences, and riparian buffers may also enhance the connectivity of landscapes and promote the dispersal of various wildlife species (*16*, *47*, *83*). Although these structural features are known to increase the occurrence of a wide variety of organisms within agricultural landscapes (*43*, *84*), how they affect the dispersal potential of organisms within diversified agricultural lands is poorly understood. Nonetheless, ambitious, large-scale connectivity projects, such as the Mesoamerican Biological Corridor project (*43*), the silvopastoral and rotational grazing project in the Santa Catarina Atlantic Forest (*55*), various linkages in Australia (Box 2), and the restoration of the migratory pathway of the monarch butterfly (*Danaus plexippus*) in the U.S. midwestern states (*85*), are under way for agricultural lands. In the latter case, although a daunting amount of restoration would be required to support the butterfly, it could simultaneously enhance soybean pollination, improve water quality, protect other biodiversity, and increase agricultural profitability (Fig. 5 and table S2) (*86*, *87*).

Although entrenched policies and the extreme concentration of agrifood industries favor industrialized supply chains and make transformation to diversified, agroecological systems difficult (*68*, *72*), reasons for optimism exist. Global grassroots movements such as La Via Campesina have provided technical, social, and material support to farmers for the spread of agroecology, confronted industrial agribusiness, and fought to influence national and global policies (*64*). Alternative agrifood systems and local and regional initiatives that provide support for diversified, agroecological systems are emerging (*64*, *69*). International initiatives supporting agroecology include the United Nations Right to Food program, which embraces it as a key element for enhancing food security globally (*88*), and programs of the Food and Agriculture Organization, which has held global and regional conferences on agroecology and included it in Farmer Field Schools since 2014 (*68*).

Rangelands and forests

Forests in the boreal, temperate, and tropical regions make up ~30% of Earth's area (*89*), whereas rangelands, which are defined as having <10% tree cover and include grasslands, desert shrublands, savanna woodlands, alpine meadows, and areas of tundra grasses and shrubs, constitute ~44% (*90*). Grazed by wild and domestic animals, they vary greatly in productivity. Both natural forests and rangelands have been lost or degraded over the past several hundred years by the increased extent and intensity of human use, including timber harvest, grazing, and conversion to agriculture. Forests continue to be lost and degraded at an alarming rate (*62*), although forest regrowth due to rural depopulation is also occurring in some areas (*20*). A recent global analysis of sources of tree cover losses showed that industrial agriculture for commodity crops is responsible for the permanent conversion of 5 million ha of forest per year (27% of losses, concentrated primarily in portions of Latin America and Southeast Asia), whereas shifting agriculture (primarily in Africa) and forestry (primarily in North America and Europe) cause forest disturbance or degradation over an equivalent land

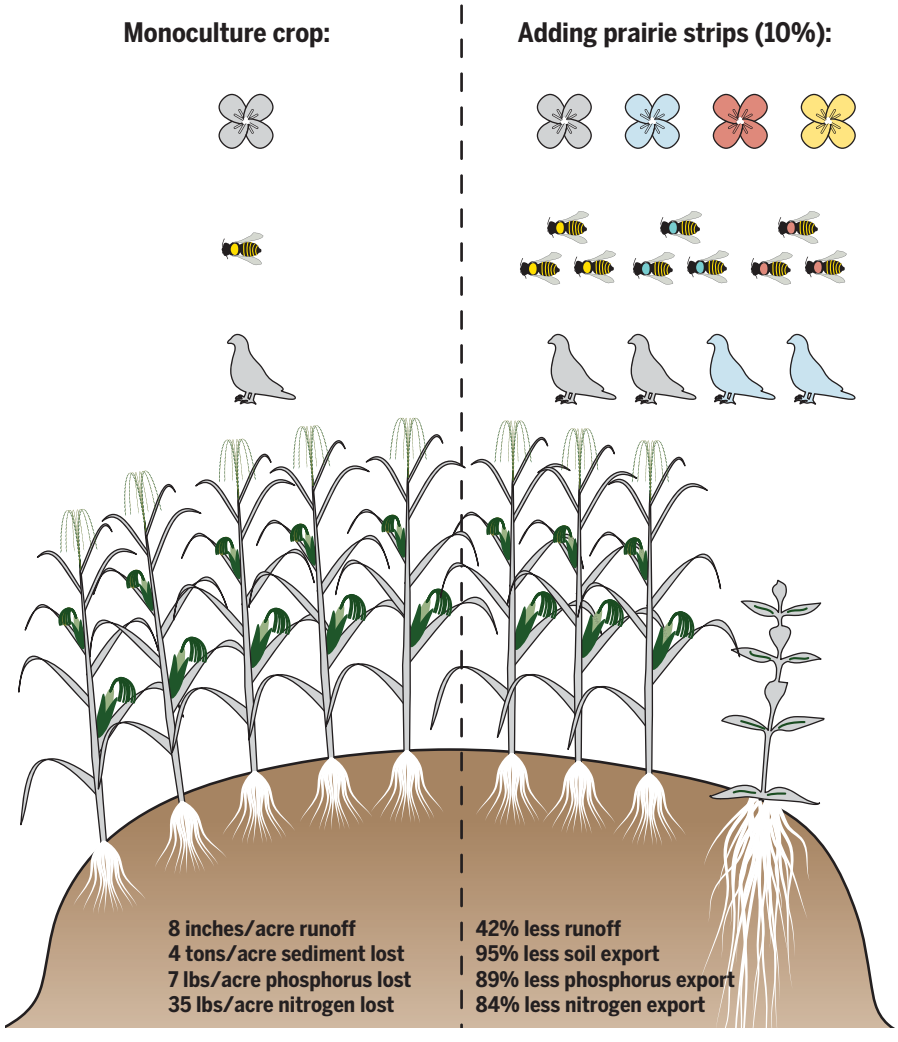


Fig. 5. Diversification practices can increase biodiversity. The integration of prairie strips into a corn-soy rotation exemplifies how diversification within working lands can substantially increase plant, pollinator, and bird species richness and abundance by two- to fourfold (as indicated by colors and numbers of icons, respectively) while minimizing externalities and enhancing other ecosystem services, such as pollination for the soy crop (table S2) (*86*).

ILLUSTRATION: NIRJA DESAI/*SCIENCE* BASED ON STRIPS RESEARCH TEAM AND LEOPOLD CENTER FOR SUSTAINABLE AGRICULTURE

area, followed by regrowth (62). It is critical, therefore, to cease permanent conversion of forests for commodity cropping and to apply restorative management approaches in working forests and rangelands.

Since 1990, many nations have created enabling policies and legislation for sustainable forest management (89). Of the 54% of global forests considered "permanent" (that is, expected to retain forest cover in the long term), 99% of these 2.17 billion ha are covered by such policies, a necessary but not sufficient condition for sustainable management. Indicators of sustainable management also show positive temporal trends, but over smaller areas. For example, forest certification (table S1) covered 430 million ha by 2014 (89), but largely within boreal and temperate regions, where land-clearing rates are less acute than those in the tropics.

An array of restorative forest and rangeland management options exist that are compatible with the conservation of working lands (Fig. 3B and table S2). For forests, the adoption of ecosystem-based management approaches has led to the integration of a greater variety of tree species and age and size classes, including old growth and dead and downed trees, and the incorporation of natural disturbance regimes to support more diverse ecological communities (91). This uneven-aged management style maintains similarities between natural and managed forests, contrasting with even-aged management from clear-cutting. Evidence from silvicultural trials and natural forests suggests that greater tree diversity also enhances wood yield quantity and stability (23). In keeping with the ecosystem stewardship concept (27), ecosystem-based management also emphasizes collaborative decentralized control and adaptive management, as well as landscape planning and the designation of corridors to promote wildlife (92). However, stakeholders may reject harvesting practices that negatively affect financial returns in the short term. Environmental outcomes suffered when stakeholders had stronger oversight of the process than a regulatory authority with political backing (93), supporting the need for public-private partnerships to achieve biodiversity conservation objectives.

In rangelands, compatible management practices are exemplified by the dehesa and montado traditional pastoral systems in oak savannas of Spain and Portugal, respectively. The oak trees (*Quercus rotundifolia* and *Q. suber*) are pruned to increase the production of acorns to feed to pigs and other livestock grown for high-value meat products; other sustainably harvested products include fuelwood and cork from oaks (94). These ecosystems also support endangered species and high plant and animal diversity relative to other seminatural habitats in Europe. However, grazing, browsing, and trampling can limit oak regeneration; thus, pasture areas need periodic temporary protection from livestock to promote oak recruitment and sustainable use (95). In Colombia, many ranchers are restoring degraded agricultural lands by using various silvopastoral techniques, which also enhance connectivity in these landscapes (Fig. 1).

Freshwater ecosystems

Maintaining stream flows and hydrologic connectivity is essential for conserving freshwater biodiversity and ecosystems. Because of changes in stream flows, estimates suggest that up to 75% of freshwater fish species are headed for local extinction by 2070 (96). Fresh water also limits the production of many natural resources, and its quantity and quality are in turn affected by landscape management. Appropriate management techniques can promote groundwater recharge and stream flow in working landscapes (table S2) (31, 86), of increasing importance under drier futures with more extreme precipitation events (97). Flood plains and associated riparian zones are particularly critical to conserve in working landscapes, because they disproportionately support biodiversity and ecosystem processes compared with other landscape elements (98). Riparian corridors also provide cooler and moister microclimates than surrounding areas and often span elevational and climatic gradients that may permit species to follow their climate envelopes (99).

Recommendations and concluding thoughts

Managing the working lands matrix for biodiversity needs to become a mainstream component of public and private conservation efforts, complementing the more traditional (and essential) focus on increasing the extent and effectiveness of protected areas (16). These restorative, working lands conservation approaches (table S2) should be applied to the large land area that is already used for farming, forestry, and ranching. At the same time, we critically need policies to prevent further conversion and degradation of wilderness and relatively intact ecosystems (62).

To scale up working lands conservation, increased support is needed for the voluntary, policy, and market instruments described in table S1. However, further adaptation and learning is needed to improve their efficacy, both at the project level and through evidence-based syntheses [e.g., (100)], and to increase adoption rates by considering an array of social factors (58). Further, these measures must be complemented by community-driven conservation initiatives, which, by involving young and old in stewardship, communication, citizen science, and education, can create a shared vision and innovative practices that result in collective impact. Scientists can support community-driven conservation and help advance environmental social movements by engaging the public, listening to alternative ways of knowing, and cocreating conservation, management, and policy alternatives. Especially important is to create alliances with existing community actions and social movements that share common ground, such as climate or local food movements.

Ultimately, our efforts to protect biodiversity and sustain resources must be accompanied by measures to reduce human population and consumption while increasing equitable access to resources to achieve sustainability. Opportunities to stabilize population and consumption exist. For example, through concerted government investment in voluntary family planning programs, enormous progress in reducing total fertility rates has been made even in poor countries [e.g., (101)], leading to smaller families living better. Globally, a large unmet need for family planning still exists (101); further investment could help stabilize the global population at 6 billion people by 2100, instead of the 9 to 12 billion projected without intervention (102, 103). To reduce consumption, critical targets include reducing food waste and meat consumption (104) and seeking efficiencies in energy and water use that can accompany urbanization (102). Even with well-structured policies, these changes toward lower human population and consumption would take time; thus, concerns exist that humanity will destroy biodiversity and natural resources before achieving a more sustainable human population (102). Conservation in working landscapes can help maintain all species, including people, as we strive to achieve a planet where a smaller human population lives better and more equitably with and because of wild nature.

REFERENCES AND NOTES

1. J. L. Payne, A. M. Bush, N. A. Heim, M. L. Knope, D. J. McCauley, Ecological selectivity of the emerging mass extinction in the oceans. *Science* **353**, 1284–1286 (2016). doi: 10.1126/science.aaf2416; pmid: 27629258
2. G. Ceballos, P. R. Ehrlich, R. Dirzo, Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6089–E6096 (2017). pmid: 28696295
3. G. Ceballos *et al.*, Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015). doi: 10.1126/sciadv.1400253; pmid: 26601195
4. C. A. Hallmann *et al.*, More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE* **12**, e0185809 (2017). doi: 10.1371/journal.pone.0185809; pmid: 29045418
5. T. Newbold *et al.*, Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015). doi: 10.1038/nature14324; pmid: 25832402
6. E. O. Wilson, *Half Earth* (Liveright Publishing Company, 2016).
7. P. R. Elsen, W. B. Monahan, A. M. Merenlender, Global patterns of protection of elevational gradients in mountain ranges. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6004–6009 (2018). doi: 10.1073/pnas.1720141115; pmid: 29784825
8. R. M. Pringle, Upgrading protected areas to conserve wild biodiversity. *Nature* **546**, 91–99 (2017). doi: 10.1038/nature22902; pmid: 28569807
9. J. M. Halley, N. Monokrousos, A. D. Mazaris, W. D. Newmark, D. Vokou, Dynamics of extinction debt across five taxonomic groups. *Nat. Commun.* **7**, 12283 (2016). doi: 10.1038/ncomms12283; pmid: 27452815
10. W. F. Laurance *et al.*, Averting biodiversity collapse in tropical forest protected areas. *Nature* **489**, 290–294 (2012). doi: 10.1038/nature11318; pmid: 22832582
11. M. C. Urban, Climate change. Accelerating extinction risk from climate change. *Science* **348**, 571–573 (2015). doi: 10.1126/science.aaa4984; pmid: 25931559
12. S. R. Loarie *et al.*, The velocity of climate change. *Nature* **462**, 1052–1055 (2009). doi: 10.1038/nature08649; pmid: 20033047
13. L. O. Frishkoff *et al.*, Loss of avian phylogenetic diversity in neotropical agricultural systems. *Science* **345**, 1343–1346 (2014). doi: 10.1126/science.1254610; pmid: 25214627
14. G. Chapron *et al.*, Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science* **346**, 1517–1519 (2014). doi: 10.1126/science.1257553; pmid: 25525247

15. M. D. Reynolds *et al*., Dynamic conservation for migratory species. *Sci. Adv.* **3**, e1700707 (2017). doi: 10.1126/sciadv.1700707; pmid: 28845449
16. C. Kremen, Reframing the land-sparing/land-sharing debate for biodiversity conservation. *Ann. N.Y. Acad. Sci.* **1355**, 52–76 (2015). doi: 10.1111/nyas.12845; pmid: 26213864
17. C. D. Mendenhall, A. Shields-Estrada, A. J. Krishnaswami, G. C. Daily, Quantifying and sustaining biodiversity in tropical agricultural landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14544–14551 (2016). doi: 10.1073/pnas.1604981113; pmid: 27791070
18. C. D. Mendenhall, D. S. Karp, C. F. J. Meyer, E. A. Hadly, G. C. Daily, Predicting biodiversity change and averting collapse in agricultural landscapes. *Nature* **509**, 213–217 (2014). doi: 10.1038/nature13139; pmid: 24739971
19. C. M. Kennedy, E. H. C. Grant, M. C. Neel, W. F. Fagan, P. P. Marra, Landscape matrix mediates occupancy dynamics of Neotropical avian insectivores. *Ecol. Appl.* **21**, 1837–1850 (2011). doi: 10.1890/10-1044.1; pmid: 21830722
20. J. A. Foley *et al*., Global consequences of land use. *Science* **309**, 570–574 (2005). doi: 10.1126/science.1111772; pmid: 16040698
21. N. Ramankutty *et al*., Trends in global agricultural land use: Implications for environmental health and food security. *Annu. Rev. Plant Biol.* **69**, 789–815 (2018). doi: 10.1146/annurev-arplant-042817-040256; pmid: 29489395
22. F. Isbell *et al*., Linking the influence and dependence of people on biodiversity across scales. *Nature* **546**, 65–72 (2017). doi: 10.1038/nature22899; pmid: 28569811
23. F. Isbell *et al*., Benefits of increasing plant diversity in sustainable agroecosystems. *J. Ecol.* **105**, 871–879 (2017). doi: 10.1111/1365-2745.12789
24. D. J. Eastburn, A. T. O'Geen, K. W. Tate, L. M. Roche, Multiple ecosystem services in a working landscape. *PLOS ONE* **12**, e0166595 (2017). doi: 10.1371/journal.pone.0166595; pmid: 28301475
25. D. A. Driscoll, S. C. Banks, P. S. Barton, D. B. Lindenmayer, A. L. Smith, Conceptual domain of the matrix in fragmented landscapes. *Trends Ecol. Evol.* **28**, 605–613 (2013). doi: 10.1016/j.tree.2013.06.010; pmid: 23883740
26. W. F. Fagan, E. E. Holmes, Quantifying the extinction vortex. *Ecol. Lett.* **9**, 51–60 (2006). pmid: 16958868
27. F. S. Chapin 3rd *et al*., Ecosystem stewardship: Sustainability strategies for a rapidly changing planet. *Trends Ecol. Evol.* **25**, 241–249 (2010). doi: 10.1016/j.tree.2009.10.008; pmid: 19923035
28. F. Isbell *et al*., High plant diversity is needed to maintain ecosystem services. *Nature* **477**, 199–202 (2011). doi: 10.1038/nature10282; pmid: 21832994
29. M. G. E. Mitchell, E. M. Bennett, A. Gonzalez, Linking landscape connectivity and ecosystem service provision: Current knowledge and research gaps. *Ecosystems* **16**, 894–908 (2013). doi: 10.1007/s10021-013-9647-2
30. M. Loreau, N. Mouquet, A. Gonzalez, Biodiversity as spatial insurance in heterogeneous landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12765–12770 (2003). doi: 10.1073/pnas.2235465100; pmid: 14569008
31. C. Kremen, A. Miles, Ecosystem services in biologically diversified versus conventional farming systems: Benefits, externalities, and trade-offs. *Ecol. Soc.* **17**, 40 (2012). doi: 10.5751/ES-05035-170440
32. A. C. M. Gaudin *et al*., Increasing crop diversity mitigates weather variations and improves yield stability. *PLOS ONE* **10**, e0113261 (2015). doi: 10.1371/journal.pone.0113261; pmid: 25658914
33. S. M. Philpott, B. B. Lin, S. Jha, S. J. Brines, A multi-scale assessment of hurricane impacts on agricultural landscapes based on land use and topographic features. *Agric. Ecosyst. Environ.* **128**, 12–20 (2008). doi: 10.1016/j.agee.2008.04.016
34. L. C. Ponisio *et al*., Diversification practices reduce organic to conventional yield gap. *Proc. R. Soc. London Ser. B* **282**, 20141396 (2015). doi: 10.1098/rspb.2014.1396; pmid: 25621333
35. R. J. Hobbs *et al*., Managing the whole landscape: Historical, hybrid, and novel ecosystems. *Front. Ecol. Environ.* **12**, 557–564 (2014). doi: 10.1890/130300
36. N. Estrada-Carmona, A. K. Hart, F. A. J. DeClerck, C. A. Harvey, J. C. Milder, Integrated landscape management for agriculture, rural livelihoods, and ecosystem conservation: An assessment of experience from Latin America and the Caribbean. *Landsc. Urban Plan.* **129**, 1–11 (2014). doi: 10.1016/j.landurbplan.2014.05.001
37. H. Gu, S. M. Subramanian, Drivers of change in socio-ecological production landscapes: Implications for better management. *Ecol. Soc.* **19**, 41 (2014). doi: 10.5751/ES-06283-190141
38. M. K. Anderson, *Tending the Wild: Native American Knowledge and the Management of California's Natural Resources* (Univ. of California Press, 2005).
39. E. M. Bignal, D. I. McCracken, The nature conservation value of European traditional farming systems. *Environ. Rev.* **8**, 149–171 (2000). doi: 10.1139/a00-009
40. S. K. Lowder, J. Skoet, T. Raney, The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.* **87**, 16–29 (2016). doi: 10.1016/j.worlddev.2015.10.041
41. C. L. Gray *et al*., Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nat. Commun.* **7**, 12306 (2016). doi: 10.1038/ncomms12306; pmid: 27465407
42. C. L. Shafer, Cautionary thoughts on IUCN protected area management categories V-VI. *Global Ecol. Conserv.* **3**, 331–348 (2015). doi: 10.1016/j.gecco.2014.12.007
43. F. A. J. DeClerck *et al*., Biodiversity conservation in human-modified landscapes of Mesoamerica: Past, present and future. *Biol. Conserv.* **143**, 2301–2313 (2010). doi: 10.1016/j.biocon.2010.03.026
44. C. A. Harvey *et al*., Patterns of animal diversity in different forms of tree cover in agricultural landscapes. *Ecol. Appl.* **16**, 1986–1999 (2006). doi: 10.1890/1051-0761(2006)016[1986:POADID]2.0.CO;2; pmid: 17069389
45. Ç. H. Şekercioğlu *et al*., Tropical countryside riparian corridors provide critical habitat and connectivity for seed-dispersing forest birds in a fragmented landscape. *J. Ornithol.* **156** (Suppl. 1), 343–353 (2015). doi: 10.1007/s10336-015-1299-x
46. C. A. Harvey *et al*., Contribution of live fences to the ecological integrity of agricultural landscapes. *Agric. Ecosyst. Environ.* **111**, 200–230 (2005). doi: 10.1016/j.agee.2005.06.011
47. A. Medina, C. A. Harvey, D. S. Merlo, S. Vílchez, B. Hernández, Bat diversity and movement in an agricultural landscape in Matiguás, Nicaragua. *Biotropica* **39**, 120–128 (2007). doi: 10.1111/j.1744-7429.2006.00240.x
48. K. Williams-Guillen, C. McCann, J. C. Martinez Sanchez, F. Koontz, Resource availability and habitat use by mantled howling monkeys in a Nicaraguan coffee plantation: Can agroforests serve as core habitat for a forest mammal? *Anim. Conserv.* **9**, 331–338 (2006). doi: 10.1111/j.1469-1795.2006.00042.x
49. J. Avelino, A. Romero-Gurdián, H. F. Cruz-Cuellar, F. A. J. Declerck, Landscape context and scale differentially impact coffee leaf rust, coffee berry borer, and coffee root-knot nematodes. *Ecol. Appl.* **22**, 584–596 (2012). doi: 10.1890/11-0869.1; pmid: 22611856
50. D. S. Karp *et al*., Forest bolsters bird abundance, pest control and coffee yield. *Ecol. Lett.* **16**, 1339–1347 (2013). doi: 10.1111/ele.12173; pmid: 23981013
51. C. A. Harvey *et al*., Climate-smart landscapes: Opportunities and challenges for integrating adaptation and mitigation in tropical agriculture. *Conserv. Lett.* **7**, 77–90 (2014). doi: 10.1111/conl.12066
52. A. T. H. Keeley *et al*., Making habitat connectivity a reality. *Conserv. Biol.* (2018). doi: 10.1111/cobi.13158 pmid: 29920775
53. W. C. Morse *et al*., Consequences of environmental service payments for forest retention and recruitment in a Costa Rican biological corridor. *Ecol. Soc.* **14**, 23 (2009). doi: 10.5751/ES-02688-140123
54. J. Owley, D. Takacs, "Flexible Conservation in Uncertain Times," in *Contemporary Issues in Climate Change Law and Policy: Essays Inspired by the IPCC* (UC Hastings Research Paper 180, Univ. at Buffalo School of Law, 2016), pp. 65–102; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2700658.
55. J. Alvez, A. L. Schmitt Filho, J. Farley, G. Alarcon, A. C. Fantini, The potential for agroecosystems to restore ecological corridors and sustain farmer livelihoods: Evidence from Brazil. *Ecol. Restor.* **30**, 288–290 (2012). doi: 10.3368/er.30.4.288
56. C. M. Kennedy *et al*., Bigger is better: Improved nature conservation and economic returns from landscape-level mitigation. *Sci. Adv.* **2**, e1501021 (2016). doi: 10.1126/sciadv.1501021; pmid: 27419225
57. I. Scoones, P. Newell, M. Leach, in *The Politics of Green Transformations*, I. Scoones, M. Leach, P. Newell, Eds. (Earthscan from Routledge, 2015), pp. 1–24.
58. M. B. Mascia, M. Mills, When conservation goes viral: The diffusion of innovative biodiversity conservation policies and practices. *Conserv. Lett.* **11**, e12442 (2018). doi: 10.1111/conl.12442
59. D. F. Doak, V. J. Bakker, B. E. Goldstein, B. Hale, What is the future of conservation? *Trends Ecol. Evol.* **29**, 77–81 (2014). doi: 10.1016/j.tree.2013.10.013; pmid: 24332874
60. K. Hamrick, "State of private investment in conservation 2016: A landscape assessment of an emerging market" (Forest Trends, 2016).
61. R. Chaplin-Kramer *et al*., Ecosystem service information to benefit sustainability standards for commodity supply chains. *Ann. N.Y. Acad. Sci.* **1355**, 77–97 (2015). doi: 10.1111/nyas.12961; pmid: 26555859
62. P. G. Curtis, C. M. Slay, N. L. Harris, A. Tyukavina, M. C. Hansen, Classifying drivers of global forest loss. *Science* **361**, 1108–1111 (2018). doi: 10.1126/science.aau3445; pmid: 30213911
63. D. Pannell *et al*., Understanding and promoting adoption of conservation practices by rural landholders. *Aust. J. Exp. Agric.* **46**, 1407–1424 (2006). doi: 10.1071/EA05037
64. M. Leach, I. Scoones, in *The Politics of Green Transformations*, I. Scoones, P. Newell, M. Leach, Eds. (Earthscan from Routledge, 2015), pp. 119–133.
65. D. R. Cameron, D. C. Marvin, J. M. Remucal, M. C. Passero, Ecosystem management and land conservation can substantially contribute to California's climate mitigation goals. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12833–12838 (2017). doi: 10.1073/pnas.1707811114; pmid: 29133408
66. N. Ramankutty, A. T. Evan, C. Monfreda, J. A. Foley, Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochem. Cycles* **22**, GB1003 (2008). doi: 10.1029/2007GB002952
67. M. A. Altieri, The ecological role of biodiversity in agroecosystems. *Agric. Ecosyst. Environ.* **74**, 19–31 (1999). doi: 10.1016/S0167-8809(99)00028-6
68. International Panel of Experts on Sustainable Food Systems (IPES-Food), "From uniformity to diversity: A paradigm shift from industrial agriculture to diversified agroecological systems" (IPES-Food, 2016); http://www.ipes-food.org/images/Reports/UniformityToDiversity_FullReport.pdf.
69. C. Kremen, A. Iles, C. M. Bacon, Diversified farming systems: An agro-ecological, systems-based alternative to modern industrial agriculture. *Ecol. Soc.* **17**, 44 (2012). doi: 10.5751/ES-05103-170444
70. V. Ricciardi, N. Ramankutty, Z. Mehrabi, L. Jarvis, B. Chookolingo, How much of the world's food do smallholders produce? *Global Food Sec.* **17**, 64–72 (2018). doi: 10.1016/j.gfs.2018.05.002
71. M. Montenegro, Banking on wild relatives to feed the world. *Gastron. J. Crit. Food Stud.* **16**, 1–8 (2016). doi: 10.1525/gfc.2016.16.1.1
72. A. Iles, R. Marsh, Nurturing diversified farming systems in industrialized countries: How public policy can contribute. *Ecol. Soc.* **17**, 42 (2012). doi: 10.5751/ES-05041-170442
73. A. S. Davis, J. D. Hill, C. A. Chase, A. M. Johanns, M. Liebman, Increasing cropping system diversity balances productivity, profitability and environmental health. *PLOS ONE* **7**, e47149 (2012). doi: 10.1371/journal.pone.0047149; pmid: 23071739
74. B. Phalan, M. Onial, A. Balmford, R. E. Green, Reconciling food production and biodiversity conservation: Land sharing and land sparing compared. *Science* **333**, 1289–1291 (2011). doi: 10.1126/science.1208742; pmid: 21885781
75. K. Garbach *et al*., Examining multi-functionality for crop yield and ecosystem services in five systems of agroecological intensification. *Int. J. Agric. Sustain.* **15**, 11–28 (2017). doi: 10.1080/14735903.2016.1174810
76. J. Pretty, Z. P. Bharucha, Sustainable intensification in agricultural systems. *Ann. Bot.* **114**, 1571–1596 (2014). doi: 10.1093/aob/mcu205; pmid: 25351192
77. G. M. Gurr *et al*., Multi-country evidence that crop diversification promotes ecological intensification of agriculture. *Nat. Plants* **2**, 16014 (2016). doi: 10.1038/nplants.2016.14; pmid: 27249349
78. R. F. Pywell *et al*., Wildlife-friendly farming increases crop yield: Evidence for ecological intensification. *Proc. R. Soc. London Ser. B* **282**, 20151740 (2015). doi: 10.1098/rspb.2015.1740; pmid: 26423846

79. L. Li *et al*., Diversity enhances agricultural productivity via rhizosphere phosphorus facilitation on phosphorus-deficient soils. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11192–11196 (2007). doi: 10.1073/pnas.0704591104; pmid: 17592130
80. D. K. Letourneau *et al*., Does plant diversity benefit agroecosystems? A synthetic review. *Ecol. Appl.* **21**, 9–21 (2011). doi: 10.1890/09-2026.1; pmid: 21516884
81. Y. Zhu *et al*., Genetic diversity and disease control in rice. *Nature* **406**, 718–722 (2000). doi: 10.1038/35021046; pmid: 10963595
82. S. F. Bender, C. Wagg, M. G. A. van der Heijden, An underground revolution: Biodiversity and soil ecological engineering for agricultural sustainability. *Trends Ecol. Evol.* **31**, 440–452 (2016). doi: 10.1016/j.tree.2016.02.016; pmid: 26993667
83. I. Perfecto, J. Vandermeer, A. Wright, *Nature's Matrix: Linking Agriculture, Conservation and Food Sovereignty* (Earthscan, 2009).
84. J. A. Hilty, A. M. Merenlender, Use of riparian corridors and vineyards by mammalian predators in northern California. *Conserv. Biol.* **18**, 126–135 (2004). doi: 10.1111/j.1523-1739.2004.00225.x
85. W. E. Thogmartin *et al*., Restoring monarch butterfly habitat in the Midwestern US: 'All hands on deck.' *Environ. Res. Lett.* **12**, 074005 (2017). doi: 10.1088/1748-9326/aa7637
86. L. A. Schulte *et al*., Prairie strips improve biodiversity and the delivery of multiple ecosystem services from corn-soybean croplands. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11247–11252 (2017). doi: 10.1073/pnas.1620229114; pmid: 28973922
87. E. Brandes *et al*., Subfield profitability analysis reveals an economic case for cropland diversification. *Environ. Res. Lett.* **11**, 014009 (2016). doi: 10.1088/1748-9326/11/1/014009
88. O. De Schutter, "Agroecology and the right to food," report presented at the 16th Session of the United Nations Human Rights Council [A/HRC/16/49], Geneva, Switzerland, 8 March 2011.
89. K. G. MacDicken *et al*., Global progress toward sustainable forest management. *For. Ecol. Manage.* **352**, 47–56 (2015). doi: 10.1016/j.foreco.2015.02.005
90. H. G. Lund, Accounting for the world's rangelands. *Soc. Range Manage.* **29**, 3–10 (2007).
91. S. Gauthier, M.-A. Vaillancourt, D. Kneeshaw, P. Drapeau, L. De Grandpré, Y. Claveau, D. Paré, in *Ecosystem Management in the Boreal Forest*, S. Gauthier, M.-A. Vaillancourt, A. Leduc, L. De Grandpré, D. Kneeshaw, H. Morin, P. Drapeau, Y. Bergeron, Eds. (Univ. du Québec, 2009), pp. 13–38.
92. M. Mangel *et al*., Principles for the conservation of wild living resources. *Ecol. Appl.* **6**, 338–362 (1996). doi: 10.2307/2269369
93. J. A. Layzer, *Natural Experiments: Ecosystem-Based Management and the Environment* (MIT, 2008).
94. P. Campos, L. Huntsinger, J. L. Oviedo, P. F. Starrs, M. Diaz, R. B. Standiford, G. Montero, Eds., *Mediterranean Oak Woodland Working Landscapes: Dehesas of Spain and Ranchlands of California* (Springer Science+Business Media, 2013).
95. J. A. Ramírez, M. Díaz, The role of temporal shrub encroachment for the maintenance of Spanish holm oak Quercus ilex dehesas. *For. Ecol. Manage.* **255**, 1976–1983 (2008). doi: 10.1016/j.foreco.2007.12.019
96. M. A. Xenopoulos *et al*., Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biol.* **11**, 1557–1564 (2005). doi: 10.1111/j.1365-2486.2005.001008.x
97. S. D. Polade, A. Gershunov, D. R. Cayan, M. D. Dettinger, D. W. Pierce, Precipitation in a warming world: Assessing projected hydro-climate changes in California and other Mediterranean climate regions. *Sci. Rep.* **7**, 10783 (2017). doi: 10.1038/s41598-017-11285-y; pmid: 28883636
98. F. R. Hauer *et al*., Gravel-bed river floodplains are the ecological nexus of glaciated mountain landscapes. *Sci. Adv.* **2**, e1600026 (2016). doi: 10.1126/sciadv.1600026; pmid: 27386570
99. A. T. Keeley *et al*., New concepts, models, and assessments of climate-wise connectivity. *Environ. Res. Lett.* **13**, 073002 (2018). doi: 10.1088/1748-9326/aacb85
100. W. Sutherland, L. V. Dicks, N. Ockendon, R. Smith, Eds., *What Works in Conservation* (Open Book, ed. 2, 2017), vol. 2.
101. J. Bongaarts, S. W. Sinding, A response to critics of family planning programs. *Int. Perspect. Sex. Reprod. Health* **35**, 39–44 (2009). doi: 10.1363/3503909; pmid: 19465347
102. E. W. Sanderson, J. Walston, J. G. Robinson, From bottleneck to breakthrough: Urbanization and the future of biodiversity conservation. *Bioscience* **68**, 412–426 (2018). doi: 10.1093/biosci/biy039; pmid: 29867252
103. C. J. Bradshaw, B. W. Brook, Human population reduction is not a quick fix for environmental problems. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16610–16615 (2014). doi: 10.1073/pnas.1410465111; pmid: 25349398
104. J. Poore, T. Nemecek, Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992 (2018). doi: 10.1126/science.aaq0216; pmid: 29853680
105. Great Eastern Ranges, www.ger.org.au.
106. V. Athreya, M. Odden, J. D. C. Linnell, J. Krishnaswamy, U. Karanth, Big cats in our backyards: Persistence of large carnivores in a human dominated landscape in India. *PLOS ONE* **8**, e57872 (2013). doi: 10.1371/journal.pone.0057872; pmid: 23483933

ACKNOWLEDGMENTS

We appreciate the constructive input of D. Ackerly, B. Brunner, A. Campbell, F. DeClerck, and A. Knight. **Competing interests:** The authors declare no competing interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/eaau6020/suppl/DC1
Tables S1 and S2
References (*107*–*158*)

10.1126/science.aau6020

RESEARCH ARTICLE SUMMARY

ION CHANNELS

Structure of the human voltage-gated sodium channel $Na_v1.4$ in complex with β1

Xiaojing Pan*, Zhangqiang Li*, Qiang Zhou*, Huaizong Shen*, Kun Wu*, Xiaoshuang Huang, Jiaofeng Chen, Juanrong Zhang, Xuechen Zhu, Jianlin Lei, Wei Xiong, Haipeng Gong, Bailong Xiao, Nieng Yan†

INTRODUCTION: The nine subtypes of mammalian voltage-gated sodium (Na_v) channels, $Na_v1.1$ to $Na_v1.9$, are responsible for the initiation and propagation of action potentials in specific excitable systems, among which $Na_v1.4$ functions in skeletal muscle. Responding to membrane potential changes, Na_v channels undergo sophisticated conformational shifts that lead to transitions between resting, activated, and inactivated states. Defects in Na_v channels are associated with a variety of neurological, cardiovascular, muscular, and psychiatric disorders. In addition, Na_v channels are targets for natural toxins and clinical therapeutics.

Understanding the physiological and pathophysiological mechanisms of Na_v channels requires knowing the structure of each conformational state. All eukaryotic Na_v channels comprise a single polypeptide chain, the α subunit, that folds to four homologous repeats I to IV. Channel properties are modulated by one or two subtype-specific β subunits. Cryo–electron microscopy (cryo-EM) structures of two Na_v channels, one from American cockroach and the other from electric eel, were resolved in two distinct conformations. However, the inability to record currents of either channel in heterologous systems prevented functional assignment of these structures. Structural elucidation of a functionally well-characterized Na_v channel is required to establish a model for structure-function relationship studies.

RATIONALE: After extensive screening for expression systems, protein boundaries, chimeras, affinity tags, and combination with subtype-specific β subunits, we focused on human $Na_v1.4$ in the presence of β1 subunit for cryo-EM analysis. The complex, which was transiently coexpressed in human embryonic kidney (HEK) 293F cells with BacMam viruses and purified through tandem affinity columns and size exclusion chromatography, was concentrated to ~0.5 mg/ml for cryo-EM sample preparation and data acquisition.

RESULTS: The cryo-EM structure of human $Na_v1.4$-β1 complex was determined to 3.2-Å resolution. The extracellular and transmembrane domains, including the complete pore domain, all four voltage-sensing domains (VSDs), and the β1 subunit, were clearly resolved, enabling accurate model building (see the figure).

ON OUR WEBSITE

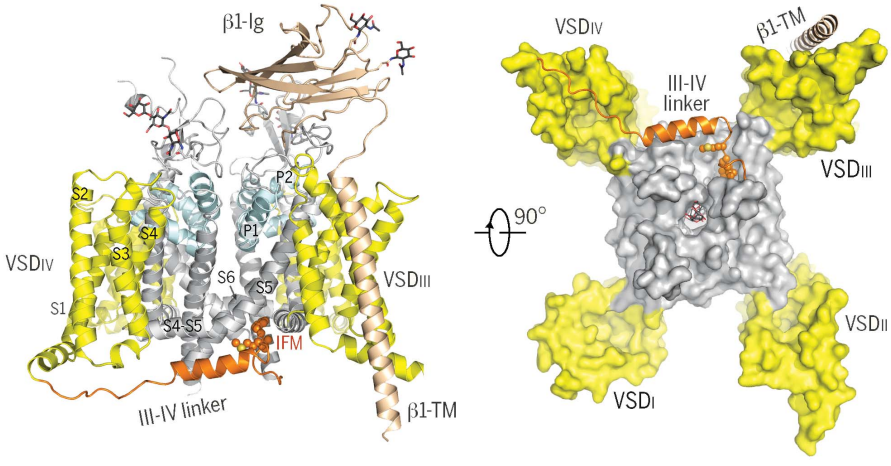
Read the full article at http://dx.doi.org/10.1126/science.aau2486

The well-resolved Asp/Glu/Lys/Ala (DEKA) residues, which are responsible for specific Na^+ permeation through the selectivity filter, exhibit identical conformations to those seen in the other two Na_v structures. A glyco-diosgenin (GDN) molecule, the primary detergent used for protein purification and cryo-EM sample preparation, penetrates the intracellular gate of the pore domain, holding it open to a diameter of ~5.6 Å. The central cavity of the pore domain is filled with lipid-like densities, which traverse the side wall fenestrations.

Voltage sensing involves four to six Arg/Lys residues on helix S4 of the VSD. This helix moves "up" (away from the cytoplasm) in response to changes of the membrane potential, and this opens the channel finally. All four VSDs display up conformations. The movement of the gating charge residues is facilitated by coordination to acidic and polar residues on S1 to S3. The improved resolution allows detailed analysis of the coordination.

The fast inactivation Ile/Phe/Met (IFM) motif on the short linker between repeats III and IV inserts into a hydrophobic cavity enclosed by the S6 and S4-S5 segments in repeats III and IV. Analysis of reported functional residues and disease mutations corroborates our recently proposed allosteric blocking mechanism for fast inactivation.

CONCLUSION: The structure provides important insight into the molecular basis for Na^+ permeation, electromechanical coupling, asynchronous activation, and fast inactivation of the four repeats. It opens a new chapter for studying the structure-function relationships of Na_v channels, affords an accurate template to map mutations associated with diseases such as myotonia and periodic paralysis hyperkalemic, and illuminates a path toward precise understanding and intervention with specific Na_v channelopathies. ■



Structure of the human $Na_v1.4$-β1 complex. Two perpendicular views are shown. Left: Side view in ribbon cartoon. The VSDs are colored yellow, and the selectivity filter and supporting helices P1 and P2 are colored light cyan. The IFM motif is shown as spheres, and the III-IV linker is colored orange. The transmembrane segments in repeat IV are labeled. Right: Surface presentation for the bottom view to highlight the intracellular gate and the cavity that accommodates the IFM motif. The GDN molecule that penetrates the intracellular gate is shown as thin sticks.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: nyan@princeton.edu
Cite this article as X. Pan *et al.*, *Science* **362**, eaau2486 (2018). DOI: 10.1126/science.aau2486

RESEARCH ARTICLE

ION CHANNELS

Structure of the human voltage-gated sodium channel $Na_v1.4$ in complex with β1

Xiaojing Pan[1,2,3,4]*, Zhangqiang Li[1,2,3,5]*, Qiang Zhou[1,2,4]*, Huaizong Shen[1,2,4]*, Kun Wu[3,6]*, Xiaoshuang Huang[1,2,5], Jiaofeng Chen[3,5], Juanrong Zhang[2,3,5], Xuechen Zhu[5], Jianlin Lei[5,7], Wei Xiong[3,5], Haipeng Gong[2], Bailong Xiao[3,6], Nieng Yan[1,2,3,5]†‡

Voltage-gated sodium (Na_v) channels, which are responsible for action potential generation, are implicated in many human diseases. Despite decades of rigorous characterization, the lack of a structure of any human Na_v channel has hampered mechanistic understanding. Here, we report the cryo–electron microscopy structure of the human $Na_v1.4$-β1 complex at 3.2-Å resolution. Accurate model building was made for the pore domain, the voltage-sensing domains, and the β1 subunit, providing insight into the molecular basis for Na^+ permeation and kinetic asymmetry of the four repeats. Structural analysis of reported functional residues and disease mutations corroborates an allosteric blocking mechanism for fast inactivation of Na_v channels. The structure provides a path toward mechanistic investigation of Na_v channels and drug discovery for Na_v channelopathies.

The physiological significance of voltage-gated sodium (Na_v) channels is underscored by their primary function in governing membrane excitability (*1*–*5*). More than 1000 point mutations have been identified in human Na_v channels associated with neurological, cardiovascular, muscular, and psychiatric disorders such as epilepsy, arrhythmia, muscle paralysis, pain syndrome, and autism spectrum disorder (*6*–*9*). Na_v channels represent major targets for a variety of natural toxins and clinical therapeutics (*10*–*12*).

In human, there are nine subtypes of Na_v channels, $Na_v1.1$ to $Na_v1.9$ (fig. S1) (*13*). $Na_v1.1$, $Na_v1.2$, $Na_v1.3$, and $Na_v1.6$ primarily function in the central nervous system. $Na_v1.4$ and $Na_v1.5$ work in skeletal muscle and heart, respectively, and $Na_v1.7$, $Na_v1.8$, and $Na_v1.9$ are mainly found in the peripheral nervous system. The ion-conducting core subunits (α subunits) are sufficient for voltage-dependent ion permeation; however, membrane trafficking and channel properties are modulated by one or two β subunits among four subtypes, β1 to β4 (*14*).

Na_v channels belong to the voltage-gated ion channel (VGIC) superfamily (*15*). Unlike most VGIC members that are homotetramers, eukaryotic Na_v channels comprise a single polypeptide chain of about 2000 residues that folds to four homologous but nonidentical repeats (repeats I to IV) (*16*). Each repeat contains six transmembrane segments (S1 to S6), among which S1 to S4 constitute the voltage-sensing domain (VSD), S5 and S6 from the four repeats together enclose the pore domain, and the sequences between S5 and S6 form the extracellular domains, the selectivity filter (SF), and the SF-supporting helices P1 and P2. Compared to Ca_v and K_v channels, Na_v channels have a highly asymmetric SF. Four distinct residues, Asp/Glu/Lys/Ala (DEKA), at corresponding locus in the SF of each repeat confer Na^+ selectivity (fig. S1) (*17*, *18*).

A simplified diagram of the working cycle of Na_v channels is featured with transitions between resting, activated, and inactivated states (*19*). VSDs undergo conformational shifts in response to membrane potential changes, leading to the opening and closing of the pore domain (*20*, *21*). Repetitively occurring basic residues Arg/Lys on the S4 segment, which represent the molecular basis for "gating charges" (GCs), are responsible for voltage sensing (*16*, *22*, *23*). At resting potential, the GC residues are attracted to the cytoplasmic side, and the pore domain is closed. Upon membrane depolarization, the outward movement of GC results in pore opening through a process known as electromechanical coupling (*24*).

Upon activation, Na_v channels immediately undergo fast inactivation, a critical mechanism that ensures recurring generation of action potentials. Decades of characterization have established the III-IV linker to be a key element for fast inactivation. The hydrophobic cluster Ile/Phe/Met (IFM) at the N terminus of the III-IV linker was defined as the fast inactivation motif (*25*–*29*). The binding site for the inactivation motif was less well understood, except that mutations on the S4-S5 linkers and S6 segments in repeats III and IV led to altered kinetics of fast inactivation (*30*, *31*).

Understanding the physiological and pathophysiological mechanism of Na_v channels requires structural elucidation of a channel in multiple functional states. In contrast to their foundational status in modern biophysics, the structural study of eukaryotic Na_v channels lagged behind other VGIC members (*32*–*35*). The pseudo-symmetry, heavy posttranslational modifications, and medium molecular weight make Na_v channels one of the most challenging targets for both x-ray crystallography and electron cryo–electron microscopy (cryo-EM) analyses.

The eukaryotic Na_v channels were structurally unveiled in 2017. The cryo-EM structures of Na_vPaS, a subtype from American cockroach, and $EeNa_v1.4$, a prototypal Na_v channel from electric eel, were determined to resolutions of 3.8 and 4.0 Å, respectively, in different conformations (*36*, *37*). In the structure of Na_vPaS, the intracellular gate is tightly closed and the side walls are sealed without fenestrations, reminiscent of what is expected for the resting state. However, the VSDs exhibit "up" conformations but are distinct from fully activated conformations. Meanwhile, the III-IV linker is sequestered by the globular C-terminal domain (CTD), unlikely to be able to execute its inactivating function. In contrast, the structure of $EeNa_v1.4$ exhibits the following features. (i) The intracellular gate is held unsealed by a digitonin-like molecule. (ii) Whereas side groups could not be assigned for VSDs in repeats I and II, the well-resolved VSD_{III} and VSD_{IV} are both up. (iii) The IFM motif is positioned in the corner enclosed by the S6 segments and the S4-S5 linkers in repeats III and IV.

Despite the conformational distinctions, functional states cannot be defined for the two structures in the absence of other reference structures. In addition, the inability to record Na_vPaS and $EeNa_v1.4$ channel currents in heterologous systems prevents functional characterization of the structures. Therefore, structural determination of a functionally well-characterized Na_v channel is required to establish a model for structure-function relationship studies.

Results

Protein generation and structural determination of the human $Na_v1.4$-β1 complex

The bottleneck for cryo-EM analysis of any human Na_v channel is the production of well-behaved recombinant proteins and preparation

[1]State Key Laboratory of Membrane Biology, Tsinghua University, Beijing 100084, China. [2]Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing 100084, China. [3]Tsinghua-Peking Joint Center for Life Sciences, Tsinghua University, Beijing 100084, China. [4]School of Medicine, Tsinghua University, Beijing 100084, China. [5]School of Life Sciences, Tsinghua University, Beijing 100084, China. [6]School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China. [7]Technology Center for Protein Sciences, Ministry of Education Key Laboratory of Protein Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.
†Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA.
‡Corresponding author. Email: nyan@princeton.edu

of cryo-samples. We have exploited different eukaryotic expression systems for all nine subtypes of human Na_v channels and screened a variety of boundaries, internal truncations, chimeras, affinity tags, and combination with β subunits (*38*). The expression level of $Na_v1.4$, a functionally well-defined subtype (*39*–*41*), was higher than that of the other isoforms. $Na_v1.4$ has been extensively investigated with multiple biophysical approaches for functional and mechanistic characterization (*42*–*46*). We therefore focused on $Na_v1.4$ for further optimization of protein production and purification.

The channel properties of the overexpressed human $Na_v1.4$, with or without the subtype-specific β1 subunit, are consistent with previous reports (fig. S2) (*47*, *48*). When coexpressed with the human β1 subunit, ~20 to 30 μg of purified protein can be obtained from 12 liters of human embryonic kidney (HEK) 293F cells by transient plasmid expression. The expression level was doubled when BacMam viruses were used to transiently coexpress the complex in HEK293F cells (*49*). The protein behaved relatively well upon size exclusion chromatography (Fig. 1A).

Although the preliminary negative staining EM analysis showed intact particles, the proteins were prone to aggregation under cryo-conditions. After another round of systematic screening, we obtained optimal cryo-samples. Details of the purification are in Materials and methods. Briefly, the full-length wild-type human $Na_v1.4$ was tandemly tagged with twin strep and FLAG on the N terminus. The $Na_v1.4$-β1 complex was extracted using 1% (w/v) *n*-dodecyl-β-D-maltopyranoside supplemented with 0.2% (w/v) cholesteryl hemisuccinate and subsequently purified in the presence of 0.06% (w/v) glycodiosgenin (GDN; Anatrace). After tandem affinity purification and size exclusion chromatography, the presence of the complex was confirmed by mass spectrometric analysis.

Details of cryo-grid preparation, data acquisition, and structural determination can be found in Materials and methods. Out of ~1.7 million particles collected on a Titan Krios electron microscope equipped with Cs corrector, Gatan K2 Summit detector, and GIF Quantum energy filter, a total of 191,936 particles were selected to yield an EM map with the overall resolution of 3.2 Å according to the gold standard Fourier shell correlation (FSC) 0.143 criterion (Fig. 1, B to F; figs. S3 and S4; and table S1). Nearly all the extracellular and transmembrane sequences are well resolved except for residues 287 to 335, an extracellular segment between $S5_I$ and $S6_I$ (where "I" indicates repeat I) that is least conserved between $Na_v1.4$ and the other eight subtypes (fig. S1).

Overall structure of the $Na_v1.4$-β1 complex

In total, 1311 side chains were built for the human $Na_v1.4$-β1 complex (table S1). The overall structure is nearly identical to that of the $EeNa_v1.4$-β1 complex, with a root mean square deviation of 0.955 Å over 1111 Cα atoms. Even the glycosylation sites completely overlap, highlighting the evolutionary conservation (fig. S5). Nevertheless, the markedly improved resolution enabled more accurate model building for detailed analysis of the functional entities such as the SF and all four VSDs (Fig. 2, A and B, and fig. S4). In addition, densities corresponding to four phospholipids are found attached to the pore domain. Three additional linear densities traverse the fenestrations on three sides of the pore domain and fill up the central cavity (Fig. 2A and fig. S6).

The common structural features between $Na_v1.4$-β1 complexes from human and electric eel include the interactions between β1 and $Na_v1.4$ (Fig. 2A and fig. S5), the acidic residue–enriched and multiple disulfide bond–stabilized extracellular loops above the pore domain (Fig. 2B), and a detergent molecule–penetrated intracellular gate (Fig. 2, C and D, and fig. S6A). This density, which can now be perfectly docked with a GDN molecule, is plugged into the intracellular gate of $Na_v1.4$, holding the intracellular gate open to a radius of 2.8 Å, slightly wider than that in $EeNa_v1.4$ (Fig. 2, C and D, and fig. S6B). Because these features have been illustrated in detail for $EeNa_v1.4$ (*37*), we will not elaborate on them in this study. Instead, we will focus on the SF and the VSDs, which are now elucidated to near-atomic resolution.

The asymmetric SF

In Na_v channels, the sieve on the extracellular side for ion permeation, or the high–field strength site (*50*), consists of an outer negative ring constituted by Asp/Glu residues on the first turn of the P2 pore helix in each repeat, the signature DEKA motif, and a wide inner site enclosed by eight carbonyl oxygens (*36*). The DEKA motif (Asp^{406}/Glu^{761}/Lys^{1244}/Ala^{1536}) was reliably resolved in the EM reconstruction for human $Na_v1.4$ with a local resolution of up to 2.8 Å (Figs. 1F and 3A).

The outer acidic residue Asp or Glu is separated from the corresponding locus of the DEKA

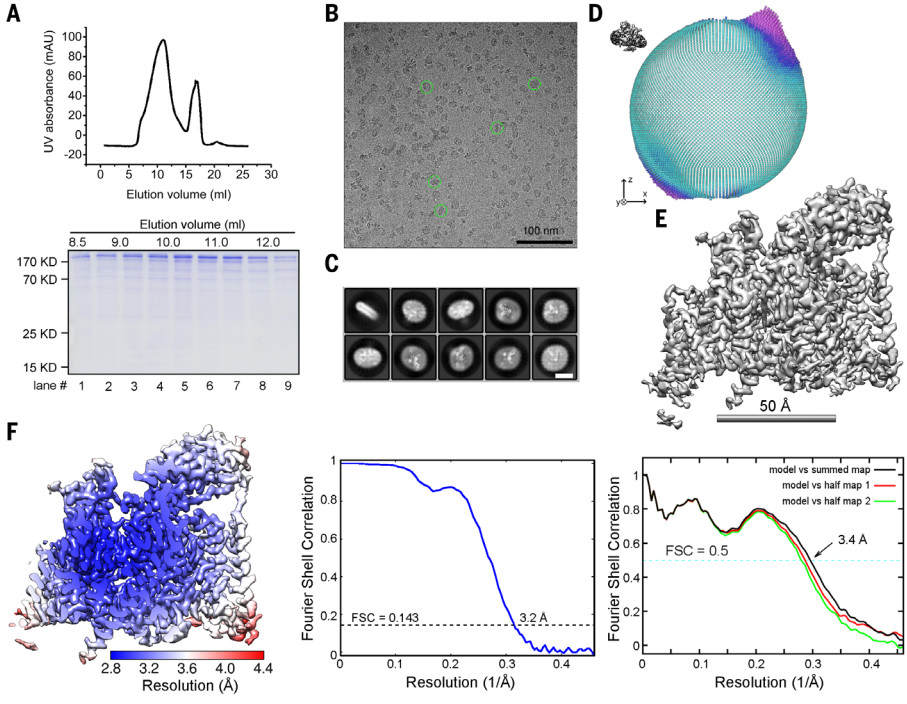


Fig. 1. Cryo-EM analysis of the human $Na_v1.4$-β1 complex. (**A**) Last step purification of recombinantly expressed $Na_v1.4$-β1 complex. A representative chromatogram of gel filtration purification is shown. The indicated fractions were resolved by SDS–polyacrylamide gel electrophoresis followed by Coomassie blue staining. The presence of β1, although invisible on the gel, was confirmed by mass spectrometric analysis. (**B**) Representative electron micrograph of the $Na_v1.4$-β1 complex. Characteristic particles in distinct orientations are highlighted by green circles. (**C**) Representative 2D class averages of $Na_v1.4$-β1 particles. Scale bar, 10 nm. (**D**) Angular distribution of the final reconstruction. Each column represents one view, and the size of the column is proportional to the number of particles in that view. (**E**) EM map for the human $Na_v1.4$-β1 complex. The map is generated in Chimera (*83*). (**F**) Overall resolution of 3.2 Å. Left: Local resolution map calculated with Relion 2.0. Middle: Gold standard FSC curve for the 3D reconstruction. Right: FSC curves of the refined model versus the overall 3.2-Å map that it was refined against (black), of the model refined in the first of the two independent maps used for the gold standard FSC versus that same map (red), and of the model refined in the first of the two independent maps versus the second independent map (green). The small difference between the red and green curves indicates that the refinement did not suffer from overfitting. UV, ultraviolet.

motif by two residues in repeats I, II, and IV, and three in repeat III. Yet, the SF segment in repeat II is one residue shorter than those in the other repeats, with the invariant Trp immediately following Glu instead of being separated by one residue (fig. S1). Consequently, Glu^{764} and Asp^{1248} are aligned above E and K in repeats II and III, respectively, whereas Glu^{409} in repeat I and Asp^{1539} in repeat IV are shifted aside by about 120° (Fig. 3B). This spatial arrangement creates additional asymmetry at the entrance to SF (Fig. 3, B and C). The carboxylate groups of Glu^{764} on $P2_{II}$ and DE (Asp^{406}/Glu^{761} in $hNa_v1.4$) are clustered to constitute a highly electronegative site, which we referred to as the DEE site (Fig. 3C). A Na^+ ion is coordinated by DEE in a 2.6-Å cryo-EM structure of Na_vPaS (*51*), but no density is observed inside or adjacent to the SF vestibule in the EM map for human $Na_v1.4$.

Nearly half a century ago, Hille suggested that the cross-section of the constriction in Na_v channels is about 3 Å by 5 Å (*52*), which can now be explained by the structure. The distance between the amine of K (Lys^{1244}) on repeat III and the opposing carbonyl oxygen of D (Asp^{406}) on repeat I is ~3.5 Å and marks the constriction point along the permeation path (Figs. 2D and 3D), whereas the distance of the corresponding loci on repeats II and IV is longer (Fig. 3D).

In our recent molecular dynamics (MD) simulation analysis of Na_vPaS, the side chains of DEE residues remain relatively rigid during Na^+ penetration (*53*). The well-resolved local densities show that both Glu^{761} and Glu^{764} are stabilized by the guanidinium group of an adjacent invariant Arg^{756} on the $P2_{II}$ helix (Fig. 3D, left). The electronegative surface of the extracellular loops and the outer mouth to the SF vestibule can effectively attract cations and repel anions (Fig. 3E). Once cations are drawn to the outer vestibule of the SF, they are likely driven to the DEE site owing to electrostatic attraction (Fig. 3, C to E). The geometric arrangement of the three carboxylate groups of DEE and the dimension of the constriction site may collectively confer Na^+ selectivity (*53*).

The four VSDs exhibit up conformations

The voltage-dependent movement of GC residues is facilitated by a number of conserved acidic and polar residues exemplified by two residues on S2 (designated An1 and An2) and one invariant Asp on S3 segments (*54*–*57*). An2 and a cyclic hydrophobic residue on S2 and the S3-Asp were functionally defined as the charge transfer center (CTC) (fig. S1) (*58*). Structure-based sequence alignment of the four VSDs in human $Na_v1.4$ shows that the corresponding segments of S2 to S4 in the four domains have different lengths (Fig. 4A). For instance, there are six helical turns in $S4_I$ and $S4_{II}$, and seven in $S4_{III}$ and $S4_{IV}$. The number of GC residues varies from four to six in the four VSDs (Fig. 4A). We defined the one on the last intracellular helical turn of S4 as the sixth GC (*36*, *37*, *59*). There are four GC residues (R2 to R4 and K5) in VSD_I, five (R2 to R4, K5, and K6) in VSD_{II}, five (K1 and R2 to R5) in VSD_{III}, and six (R1 to R5 and R/K6) in VSD_{IV}. In addition, an Arg is positioned four, instead of three, residues away from R5 in VSD_{III}

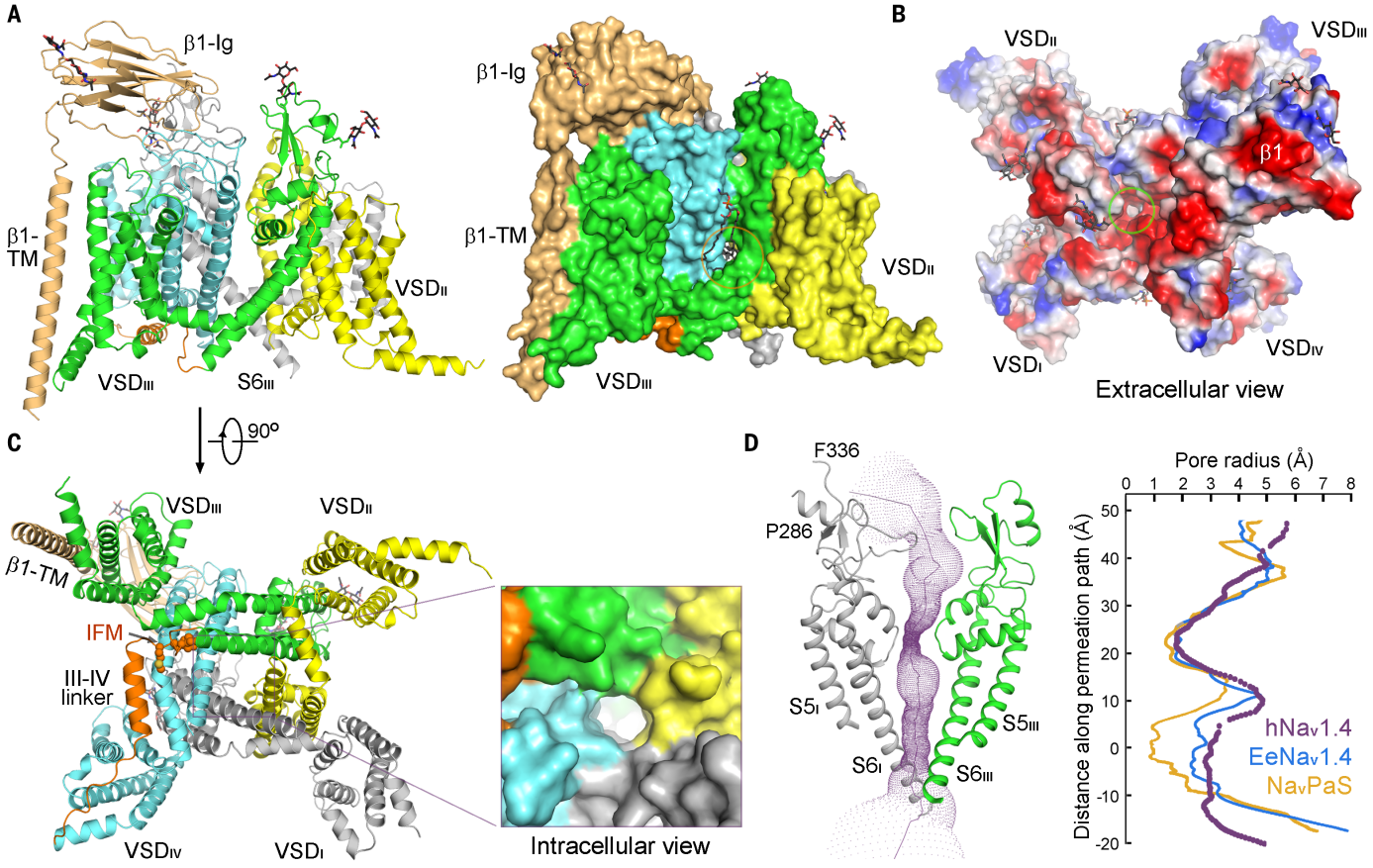


Fig. 2. Structure of the human $Na_v1.4$-β1 complex. (**A**) The extracellular and transmembrane regions of the $Na_v1.4$-β1 complex are clearly resolved. The structure is domain colored. The glycosyl moieties are shown as sticks. The same side view is shown in cartoon (left) and surface (right) representations. (**B**) The extracellular surface of the pore domain is highly electronegative. The surface electrostatic potential is calculated in PyMOL (*87*). (**C**) The intracellular gate of $Na_v1.4$ is kept unclosed. The IFM fast inactivation motif is shown as spheres, and the III-IV linker is colored orange. Inset: Intracellular gate. Detailed analysis of the hydrophobic intracellular gate that is penetrated by a GDN molecule is presented in fig. S6B. (**D**) The SF marks the constriction site along the permeation path in the pore domain. The permeation path, calculated by HOLE (*88*), is illustrated by purple dots in the left panel. The corresponding pore radii of human $Na_v1.4$ (purple), $EeNa_v1.4$ (blue), and Na_vPaS (orange) are compared in the right panel. All structure figures were prepared in PyMOL. Ig, immunoglobulin.

of $Na_v1.1$ to $Na_v1.8$ (Fig. 4A and fig. S1). As the sequences of each corresponding VSD in the nine Na_v subtypes are highly conserved, the observed VSD structures in human $Na_v1.4$ are likely shared by all mammalian Na_v channels (fig. S1).

In the present $Na_v1.4$ structure, the S4 segment harboring R2 to R6 is a 3_{10} helix in all four VSDs, while the R1-carrying segment is relaxed to an α helix in VSDs III and IV (Fig. 4B). When individual VSDs are superimposed relative to the CTC, the height of the Cα atoms of corresponding GCs is III > I > IV > II (Fig. 4B and fig. S7). In all four VSDs, GCs R/K1 to R4 are above the occluding aromatic Tyr/Phe on S2. R5 in I, II, and IV is below the occluding residue and coordinated by the acidic residues in CTC, whereas R5 in VSD_{III} is at a similar height to Phe^{1076}, forming π-cation interaction (Fig. 4C and fig. S7). In total, 15 amine or guanidinium groups in the four VSDs—3 each in VSD_I and VSD_{II}, 5 in VSD_{III}, and 4 in VSD_{IV}—are on the extracellular side of the occluding Phe/Tyr (Fig. 4C and fig. S7). Although the VSDs appear in different up conformations, their functional states cannot be defined in the absence of reference structures in other conformations, such as the resting state.

The improved resolution allows detailed analysis of the coordination of GCs in each VSD. The interactions between GCs and An1 and CTC are presented in detail in Fig. 4C and fig. S7 and will not be enumerated here. It is noteworthy that several conserved polar or acidic residues on the corresponding loci of S1, which have been less characterized before, also participate in the coordination of GCs (Fig. 4, A and C). In VSDs II to IV, an invariant Glu on the extracellular tip of S1 interacts with R2, while a polar residue in the middle of S1 (Ser in VSD_{III} and Asn in the other three VSDs) coordinates R4. On the intracellular tip of S1, an invariant Asp (Asp^{581} in VSD_{II} and Asp^{1356} in VSD_{IV}) interacts with K6/R6, and a Glu (Glu^{1034}) interacts with the C-terminal Arg^{1142} in VSD_{III} (Fig. 4C). The intracellular locus on $S1_I$ is occupied by Ser or Asn (Fig. 4A and fig. S1), which is not involved in the interaction with S4 residues in the present structure, but may participate in GC coordination in other functional states.

The varied numbers and chemical characteristics of GCs (Lys versus Arg) and charge transfer facilitating residues (D/E/N/S) will naturally result in distinct affinities between S4 and surrounding elements in different VSDs, thereby providing the molecular basis for asynchronous movements of the VSDs during voltage-dependent activation (*42*, *60*). In addition, the distinct interface between each VSD and the pore domain further contributes to the functional asymmetry of the four repeats and adds to the complexity in dissecting the electromechanical coupling mechanism of Na_v channels (fig. S8) (*42*, *60*).

Structural basis for fast inactivation

In the cryo-EM structures of Na_vPaS and $Ca_v1.1$, the III-IV linker is both sequestered by the glob-

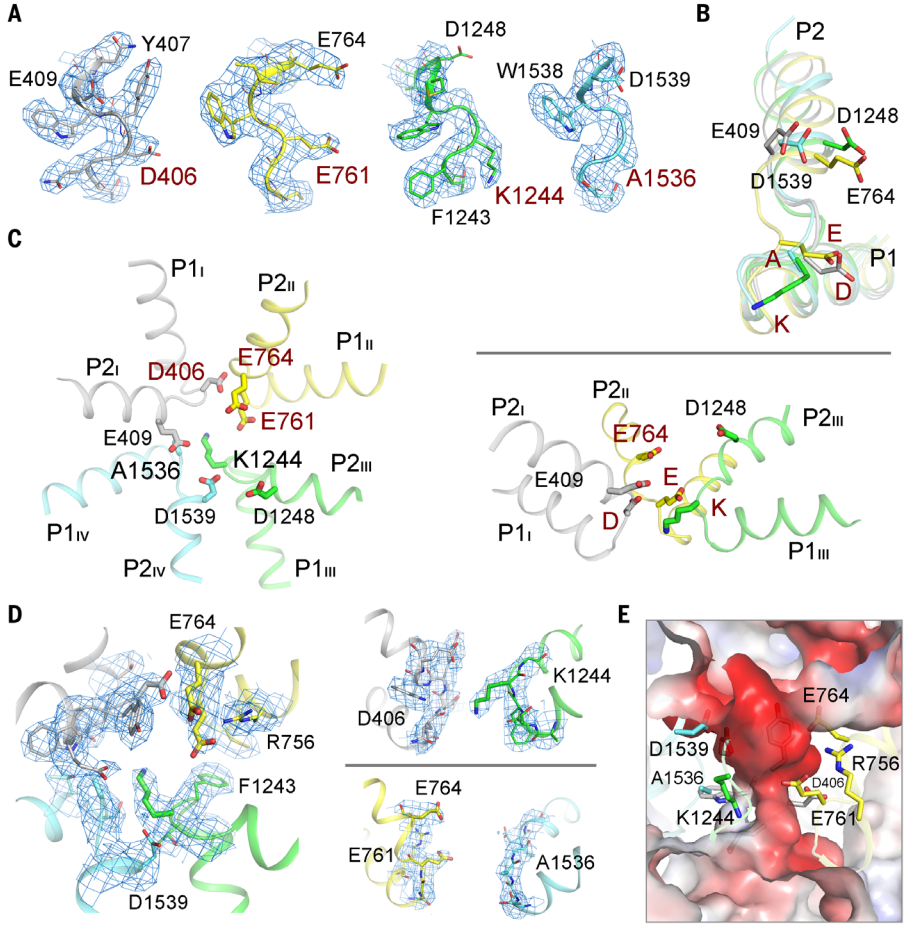


Fig. 3. The asymmetric SF. (**A**) All the residues that constitute the SF are well resolved in $Na_v1.4$. The densities, shown as blue mesh, are contoured at 5 σ. The signature DEKA residues are labeled dark red. (**B**) Superimposition of the P1-SF-P2 segments from the four repeats. DEKA motif and the residues that constitute the outer negative ring are shown as sticks. (**C**) Structure of the SF. A top view is shown on the left, and a side view with repeat IV omitted is shown on the right. Asp^{406} and Glu^{761} in the DEKA motif and Glu^{764} on $P2_{II}$ form a favorite Na^+ binding site, designated the DEE site, consistent with the observation in the structure and MD simulation of Na_vPaS (*51*, *53*). (**D**) Asymmetric SF vestibule. The densities for the residues that constitute the SF vestibule are contoured at 5 σ. An extracellular view (left) and two side views of diagonal repeats (right) are shown. (**E**) Potential Na^+ permeation path. A cut-open side view of the surface electrostatic potential of the SF vestibule is shown.

ular CTD (*36*, *59*), offering a structural basis for the observed functional importance of the CTD in inactivation (*61*–*63*). However, neither Na_vPaS nor $Ca_v1.1$ has the IFM motif. In the structure of $EeNa_v1.4$, the III-IV linker is repositioned relative to Na_vPaS and $Ca_v1.1$, and the IFM motif is plugged into the corner between the S6 and the S4-S5 segments in repeats III and IV. On the basis of the conformational changes between Na_vPaS and $EeNa_v1.4$, we postulated an allosteric blocking mechanism wherein the IFM motif does not directly block the pore but pushes pore closure by squeezing into the space between the S6 helical bundle and the S4-S5 restriction ring (*37*). Nonetheless, the moderate resolution of the $EeNa_v1.4$ structure limited analysis of local interactions.

The IFM motif and surrounding residues were unambiguously resolved in human $Na_v1.4$ (Fig. 5, A and B). Midway between the intracellular gate and VSD_{III}, the S4-S5 and S6 segments in repeats III and IV together with $S5_{IV}$ enclose a hydrophobic cavity (Fig. 5A). The aromatic ring of Phe^{1311} inserts deeply into the cavity, whereas Ile^{1310} and Met^{1312} buttress the interaction by binding to the edges of the cavity (Fig. 5, A and B). This structural feature is consistent with mutagenesis of the IFM motif, which showed that the single point mutation of Phe to Gln had a more pronounced effect on fast inactivation than mutation of either Ile or Met (*29*).

The III-IV helix that follows the IFM motif interacts extensively with $S4\text{-}S5_{IV}$ through van der Waals contacts (Fig. 5, C and D). Mutations of Phe^{1473}, Met^{1476}, and Met^{1477} on $S4\text{-}S5_{IV}$, which are mapped to the interface with the III-IV linker, resulted in fast inactivation defects (*64*, *65*). In

addition, Asn[1484] on $S5_{IV}$ is hydrogen bonded to both the side group of Gln[1316] on the first helical turn of the III-IV helix and the carbonyl oxygen of Phe[1311]. Mutation of the corresponding Asn to Ala in rat $Na_v1.2$ completely abolished fast inactivation (*65*). Ala substitution or mutation of hydrophobic residues to polar residues would disrupt or weaken the association between the III-IV linker and the involved segments. Therefore, these mutagenesis characterizations provide strong support for the critical role of the structurally revealed interface in fast inactivation.

Structural analysis of disease mutations

Near-atomic resolution structures of human Na_v channels can provide accurate templates to analyze disease-related mutations (table S2), thereby illuminating a path toward precise understanding and intervention with specific Na_v channelopathies. More than 70 mutations have been identified in human $Na_v1.4$ associated with different forms of myotonia and periodic paralysis hyperkalemic (Fig. 6 and table S2). Structural mapping identified several hotspots for disease mutations, several of which are discussed below (Fig. 6).

GC residues represent important hotspots in all Na_v channels owing to their fundamental role in voltage sensing and electromechanical coupling (*8*). Only mutations of upper GC residues R1 to R4, but not R5 or R6 in $Na_v1.4$, are found in several types of disease (Fig. 6). Although GC-interacting residues also affect voltage-dependent movements of S4, only one disease mutation D1069N (An1 in VSD_{III}) was discovered (Fig. 6C).

Another hotspot is related to fast inactivation. Some pathogenic mutations were characterized to cause inactivation defects, such as A1152D (*66*, *67*), G1306→A/E/V (*68*, *69*), and T1313M (*70*). Ala[1152] is a residue in the hydrophobic cavity and directly interacts with Ile[1310] and Phe[1311] (Fig. 5B). Gly[1306] is located at the sharp turn between $S6_{III}$ and the IFM motif. Substitution of Gly with side groups may alter the secondary structure. Thr[1313] immediately follows the IFM motif. It contributes to the polar interaction between Gln[1316] on the III-IV helix and Asn[1484] on $S5_{IV}$ (Fig. 5, C and D). The mutation A1481D is responsible for cold-induced myotonia (*71*). Ala is mapped to S4-5_{IV}, in direct contact with both Phe[1311] and Met[1312] (Fig. 5, B and D). The structure thus provides a molecular interpretation for the inactivation defects, and these disease mutations, in turn, serve as evidence for the allosteric blocking mechanism suggested based on the structure (*8*).

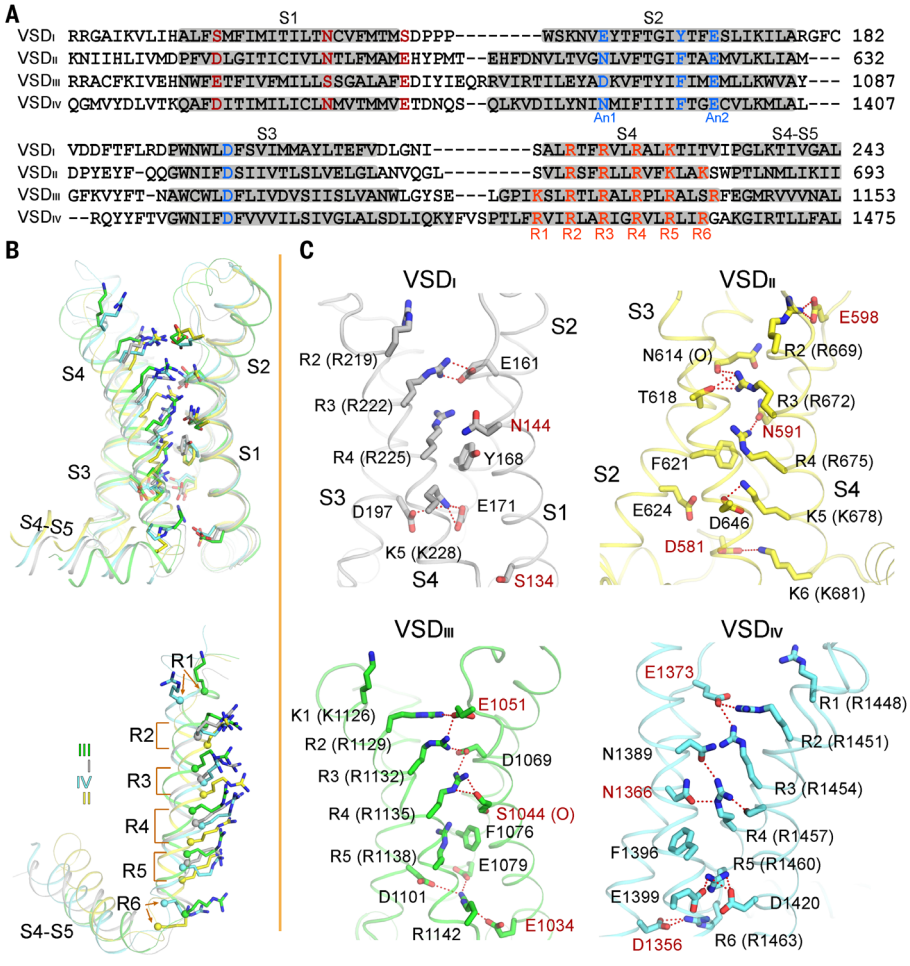


Fig. 4. Molecular basis for the functional asymmetry of the VSDs. (**A**) Structure-based sequence alignment of the four VSDs in human $Na_v1.4$. The VSDs are superimposed in PyMOL, and the resulting sequence alignment is shown. The CTC, GC residues, and the conserved polar or acidic residues on S1 that coordinate GCs are highlighted with blue or red colors. The boundaries for the S1 to S4 helices are shaded gray. (**B**) All four VSDs exhibit up conformations. Top: The four VSD structures are superimposed relative to CTC and An1 on S2. Bottom: The corresponding GC residues in the four VSDs are at different heights relative to the CTC. The Cα atoms of the GC residues are shown as spheres. (**C**) Distinct interior compositions of the four VSDs. The GC residues on S4 and their coordinating residues on S1 to S3 are shown as sticks. The electrostatic interactions are indicated by red dashed lines. The polar residues on S1 that may facilitate charge transfer are labeled red.

Conclusions

Structural determination of an intact human Na_v channel to near-atomic resolution opens a new chapter for studying the structure-function relationships of these channels, which are of remarkable physiological and pathophysiological significance. The structure bridges the large body of experimental and clinical data accumulated in the past seven decades to future structure-based mechanistic investigations, drug discovery, and protein engineering. The methods for protein generation and data processing of human $Na_v1.4$ could also inform the cryo-EM structural analysis of other Na_v subtypes.

The transitions between voltage-dependent working states of Na_v channels entail complicated conformational changes and rearrangement of interactions between adjacent or even distant structural elements through direct or allosteric effects. It is impossible to fully interpret functional and mutational observations based on one structure. It is thus important to elucidate the structures of a specific channel in distinct functional states. In this regard, structural resolution of the functionally well-defined $Na_v1.4$ represents an exciting start point to eventually unveil the choreography of voltage-dependent electromechanical coupling of Na_v channels by means such as mutagenesis and ligand interference.

Materials and methods

Transient coexpression of human $Na_v1.4$ and β1

The optimized coding DNAs for human $Na_v1.4$ (Uniprot: P35499) and β1 (Uniprot: Q07699) were cloned into the pEG BacMam vector (*49*), with tandem twin Strep-tag and FLAG tag at the N terminus of $Na_v1.4$. The BacMam viruses were produced and amplified in Sf9 cells (Invitrogen). The HEK293F cells (Invitrogen) were cultured in SMM 293T-I medium (Sino Biological Inc.) supplemented with 5% CO_2 in a Multitron-Pro shaker (Infors, 130 r.p.m.) at 37°C. When cell density reached 4×10^6 cells per ml, the culture was diluted to 2×10^6 cells per ml with fresh medium and transfected with the P2 (the second

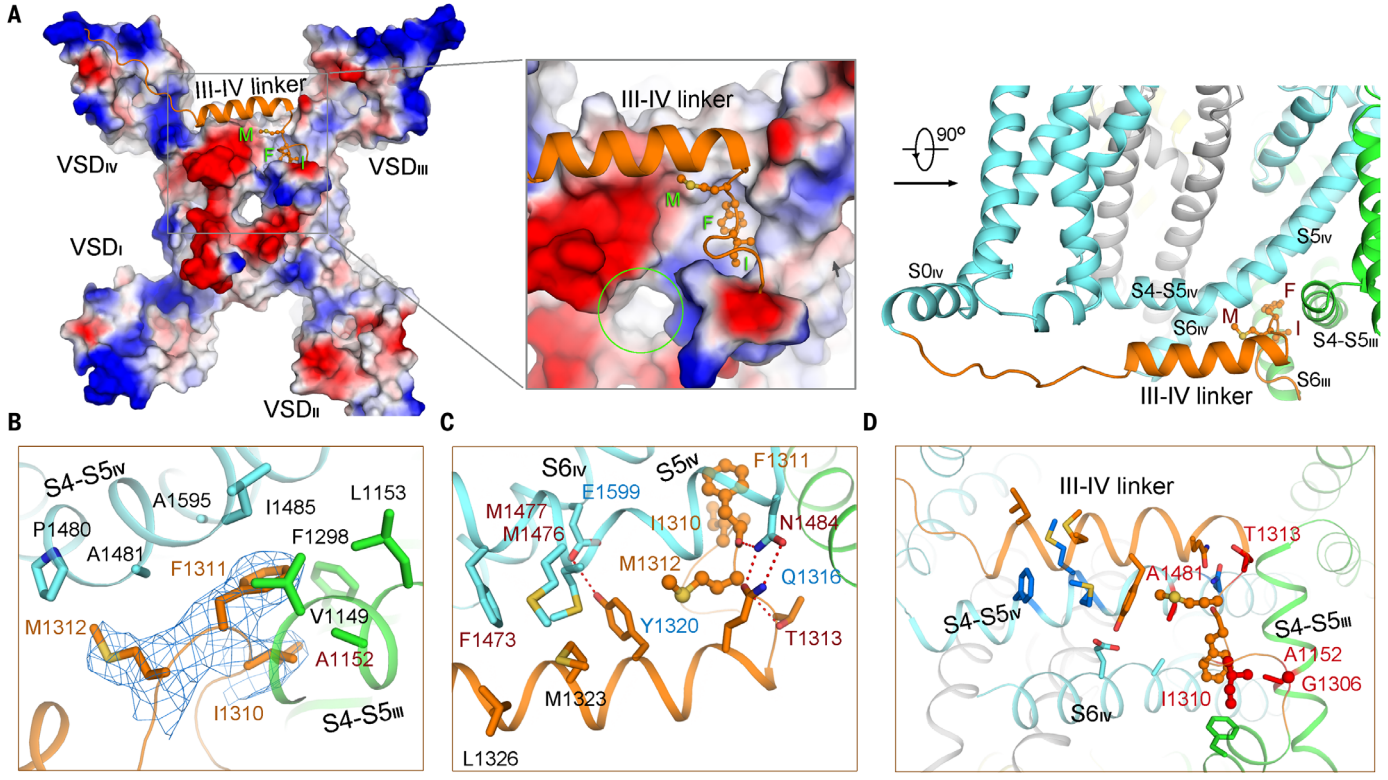


Fig. 5. An allosteric blocking mechanism for fast inactivation. (**A**) The fast inactivation IFM motif inserts into a hydrophobic cavity enclosed by the S4-S5 and S6 segments in repeats III and IV. An intracellular view of the electrostatic surface potential is shown to highlight the hydrophobic cavity. Inset: The intracellular gate is indicated by green circle. (**B**) Accommodation of the IFM motif. The density for the IFM motif, shown as blue mesh, is contoured at 5 σ. (**C**) Extensive interactions between the III-IV motif and adjacent elements. The residues that were previously identified to be important for fast inactivation and those involved in polar interactions are labeled brown and blue, respectively. (**D**) Structural analysis of disease mutations that cause fast inactivation defects. An intracellular view is shown. The residues whose mutations are found in various types of myotonia are colored and labeled red. The residues that were characterized to be involved in fast inactivation are colored blue. IFM residues are shown as ball and sticks.

passage) BacMam viruses of $Na_v1.4$ and β1, each at 25 ml per liter cell culture. Sodium butyrate was added to the cell culture at a final concentration of 10 mM to facilitate protein expression. Transfected cells were cultured for 48 hours before harvesting.

Purification of the $Na_v1.4$-β1 complex

For each batch of protein purification, 12 liters of transfected cells were harvested by centrifugation at 800 g and resuspended in lysis buffer containing 25 mM imidazole, pH 5.9, and 150 mM NaCl. The suspension was supplemented with 1% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM, Anatrace), 0.2% (w/v) cholesteryl hemisuccinate Tris salt (CHS, Anatrace), 0.001% (w/v) dodecyl sulfate sodium salt (SDS, BIO-RAD), and protease inhibitor cocktail including 2 mM phenylmethylsulfonyl fluoride (PMSF), 6.5 μg/ml aprotinin, 3.5 μg/ml pepstatin, and 25 μg/ml leupeptin. After incubation at 4°C for 3 hours, the solution was ultra-centrifuged at 200,000 g for 30 min, and the supernatant was applied to anti-Flag M2 affinity resin (Sigma) by gravity at 4°C. The resin was rinsed four times with the W buffer, which contains 25 mM imidazole pH 5.9, 150 mM NaCl, 0.06% (w/v) glyco-diosgenin (GDN, Anatrace), 0.001% SDS, and the protease inhibitor cocktail. The target proteins were eluted with W buffer plus 200 μg/ml FLAG peptide (Sigma). The eluent was then applied to Strep-Tactin Sepharose (IBA) and the purification protocol was similar to the previous step except that the elution buffer was W buffer plus 2.5 mM D-desthiobiotin (IBA). After incubation at 4°C overnight with addition of GDN to a final concentration of 0.12%, the eluent was then concentrated using a 100-kDa cut-off Centricon (Millipore) and further purified by Superose-6 column (GE Healthcare) also in the W buffer. The presence of the complex was confirmed by mass spectrometry.

Whole-cell electrophysiology

The DNAs of $Na_v1.4$ and β1 were cloned into the pCAG vector (72). The HEK293T cells, which were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 4.5 mg/ml glucose and 10% (v/v) fetal bovine serum (FBS, Gibco), were co-transfected with the expression plasmids for $Na_v1.4$, with or without β1, and an eGFP-encoding plasmid when cell confluency reached 50%. After incubation at 37°C under 5% CO_2 for 24-48 hours, the cells were treated with 0.05% trypsin (Gibco) and transferred to poly-D-lysine (Sigma)-coated 10 mm cover slips for electrophysiological characterizations.

Sodium currents were recorded using Axon 200B amplifier with Clampex10.6 software and glass micropipettes (3-4 MΩ) in HEK293T cells. For recording the voltage-dependent currents as well as the activation and inactivation currents, the electrodes were filled with the internal solution composed of (in mM) 40 CsCl, 10 NaCl, 10 EGTA, 10 HEPES, 105 CsF, pH 7.3 with CsOH. The extracellular solution contained (in mM) 130 NaCl, 4 KCl, 1 $MgCl_2$, 1.5 $CaCl_2$, 5 D-Glucose monohydrate, 5 HEPES, pH = 7.4 with NaOH. The TTX and STX inhibition currents were recorded in the same buffer with indicated concentrations of TTX and STX in the extracellular solution.

The voltage dependence of ion current was analyzed using a protocol consisting of steps from a holding potential of −90 mV to voltages ranging from −100 to 100 mV for 50 ms in 10 mV increment. In the I-V curves, the currents were normalized with the maximum and plotted against the voltage.

To measure the voltage dependence of activation, the cells were held at −90 mV, and

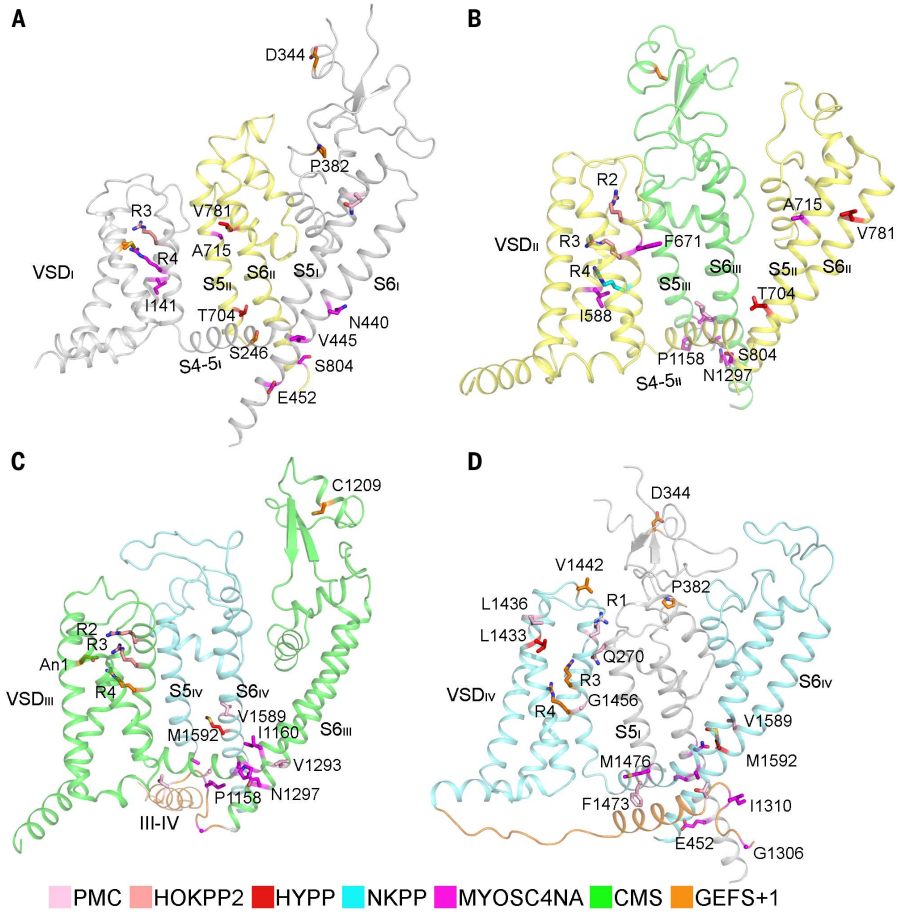


Fig. 6. Structural mapping of disease mutations in human $Na_v1.4$. (**A** to **D**) For better visual effect, one repeat and its interacting pore-domain elements from the adjacent repeat are shown in each panel. The disease-related residues are shown as sticks and color coded for a specific type of disease. PMC, paramyotonia congenita of von Eulenburg; HOKPP2, periodic paralysis hypokalemic 2; HYPP, periodic paralysis hyperkalemic; NKPP, normokalemic periodic paralysis; MYOSCN4A, myotonia SCN4A-related; CMS, myasthenic syndrome, congenital; GEFS+1, generalized epilepsy with febrile seizures plus 1. Details of the disease-associated mutations are presented in table S2.

pre-pulsed at −120 mV for 200 ms. Sodium current was elicited by a stepped depolarization test voltage pulsed from −80 mV to 80 mV for 50 ms with an increment of 5 mV. For the conductance calculation, the equation, G = I/(V − Vr), where Vr (the reversal potential) represents the voltage at which the current is zero, was used. The calculated conductance was normalized and plotted against the voltage from −80 mV to 0 mV.

For the voltage dependence of inactivation, the cells were clamped at a holding potential of −90 mV, and were applied to step pre-pulses from −100 mV to 40 mV for 1000 ms with an increment of 5 mV. The sodium current was then recorded at the test pulse of 0 mV for 50 ms. The peak currents under the test pulses were normalized and plotted against the pre-pulse voltage.

Cryo-EM data acquisition

Aliquots of 3.5 μl of freshly purified $Na_v1.4$-β1 complex, concentrated to approximately 0.5 mg/ml, were placed on glow-discharged holey carbon grids (Quantifoil Au 300 mesh, R1.2/1.3). Grids were blotted for 3.5 s and flash-frozen in liquid ethane cooled by liquid nitrogen using Vitrobot Mark IV (Thermo Fisher). Prepared grids were subsequently transferred to a Titan Krios electron microscopy (Thermo Fisher) operated at 300 kV and equipped with Cs corrector, Gatan K2 Summit detector, and GIF Quantum energy filter. A total of 8,009 movie stacks were automatically collected using AutoEMation (*73*) with a slit width of 20 eV on the energy filter and a preset defocus range from −1.8 μm to −1.5 μm in super-resolution mode at a nominal magnification of 105,000 X. Each stack was exposed for 5.6 s with the exposing time of 0.175 s per frame, resulting in 32 frames per stack. The total dose rate was 48 $e^-/Å^2$ for each stack. The stacks were motion corrected with MotionCor2 (*74*) and binned 2 fold, resulting in a pixel size of 1.091 Å/pixel. Meanwhile, dose weighting was performed (*75*). The defocus values were estimated with Gctf (*76*).

Image processing

A diagram of data processing is presented in fig. S3. A total of 1,714,341 particles were automatically picked using RELION (*77*–*79*) from 7,262 manually selected micrographs. 2D classification identified 580,143 good particles that were subsequently applied to global angular searching 3D classification with K set to 1 (one class). For each of the last several iterations of the global angular searching 3D classification, a local angular searching 3D classification was performed, during which the particles were classified into 4 classes. A total of 391,294 non-redundant particles were selected from the local angular searching 3D classification and subjected to 3 cycles of multi-reference 3D classification to remove bad particles. The selected particles were subjected to 3D auto-refinement, generating a 3D map with an overall resolution of 3.3 Å, which was further improved to 3.2 Å after local defocus correction with Gctf. The final particle number used for the 3D auto-refinement was 191,936. 2D classification, 3D classification, and auto-refinement were performed in RELION 2.1. The resolution was estimated with the gold-standard Fourier shell correlation 0.143 criterion (*80*) with high resolution noise substitution (*81*).

Model building and structure refinement

The sequences were aligned using Clustal W (*82*). Model building was carried out on the basis of the 3.2-Å reconstruction map. The coordinates of $EeNa_v1.4$-β1 (PDB accession number: 5XSY) were fitted into the EM map by CHIMERA (*83*). The sequence of $EeNa_v1.4$-β1 were mutated with corresponding residues in human $Na_v1.4$-β1 in COOT (*84*) and every residue was manually examined. The chemical properties of amino acids were considered during model building.

In total 1,138 residues in $Na_v1.4$ and 173 residues in β1 were assigned with side chains, and 9 sugar moieties, 6 lipid molecules, and 1 GDN were built. For $Na_v1.4$, the N-terminal 118 residues, extracellular loop (residues 287-335), intracellular I-II linker (residues 464-558), II-III linker (residues 807-1014), and C-terminal sequences after Glu1607 were not built due to the lack of corresponding densities.

Structure refinement was performed using phenix.real_space_refine application in PHENIX (*85*) in real space with secondary structure and geometry restraints. Over-fitting of the overall model was monitored by refining the model in one of the two independent maps from the gold-standard refinement approach and testing the refined model against the other map (*86*). Statistics of the map reconstruction and model refinement can be found in table S1.

REFERENCES AND NOTES

1. A. L. Hodgkin, A. F. Huxley, Resting and action potentials in single nerve fibres. *J. Physiol.* **104**, 176–195 (1945). doi: 10.1113/jphysiol.1945.sp004114; pmid: 16991677
2. A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544 (1952). doi: 10.1113/jphysiol.1952.sp004764; pmid: 12991237
3. B. Hille, *Ion Channels of Excitable Membranes* (Sinauer Associates, ed. 3, 2001), pp. 814.

4. W. A. Catterall, From ionic currents to molecular mechanisms: The structure and function of voltage-gated sodium channels. *Neuron* **26**, 13–25 (2000). doi: 10.1016/S0896-6273(00)81133-2; pmid: 10798388
5. C. A. Ahern, J. Payandeh, F. Bosmans, B. Chanda, The hitchhiker's guide to the voltage-gated sodium channel galaxy. *J. Gen. Physiol.* **147**, 1–24 (2016). doi: 10.1085/jgp.201511492; pmid: 26712848
6. C. J. Woolf, R. J. Mannion, Neuropathic pain: Aetiology, symptoms, mechanisms, and management. *Lancet* **353**, 1959–1964 (1999). doi: 10.1016/S0140-6736(99)01307-0; pmid: 10371588
7. W. A. Catterall, Sodium channels, inherited epilepsy, and antiepileptic drugs. *Annu. Rev. Pharmacol. Toxicol.* **54**, 317–338 (2014). doi: 10.1146/annurev-pharmtox-011112-140232; pmid: 24392695
8. W. Huang, M. Liu, S. F. Yan, N. Yan, Structure-based assessment of disease-related mutations in human voltage-gated sodium channels. *Protein Cell* **8**, 401–438 (2017). doi: 10.1007/s13238-017-0372-z; pmid: 28150151
9. M. Rubinstein *et al*., Association of rare missense variants in the second intracellular loop of $Na_V1.7$ sodium channels with familial autism. *Mol. Psychiatry* **23**, 231–239 (2018). doi: 10.1038/mp.2016.222; pmid: 27956748
10. W. A. Catterall *et al*., Voltage-gated ion channels and gating modifier toxins. *Toxicon* **49**, 124–141 (2007). doi: 10.1016/j.toxicon.2006.09.022; pmid: 17239913
11. S. K. Bagal, B. E. Marron, R. M. Owen, R. I. Storer, N. A. Swain, Voltage gated sodium channels as drug discovery targets. *Channels* **9**, 360–366 (2015). doi: 10.1080/19336950.2015.1079674; pmid: 26646477
12. M. R. Israel, B. Tay, J. R. Deuis, I. Vetter, Sodium channels and venom peptide pharmacology. *Adv. Pharmacol.* **79**, 67–116 (2017). doi: 10.1016/bs.apha.2017.01.004; pmid: 28528674
13. W. A. Catterall, A. L. Goldin, S. G. Waxman, International Union of Pharmacology. XLVII. Nomenclature and structure-function relationships of voltage-gated sodium channels. *Pharmacol. Rev.* **57**, 397–409 (2005). doi: 10.1124/pr.57.4.4; pmid: 16382098
14. H. A. O'Malley, L. L. Isom, Sodium channel β subunits: Emerging targets in channelopathies. *Annu. Rev. Physiol.* **77**, 481–504 (2015). doi: 10.1146/annurev-physiol-021014-071846; pmid: 25668026
15. F. H. Yu, V. Yarov-Yarovoy, G. A. Gutman, W. A. Catterall, Overview of molecular relationships in the voltage-gated ion channel superfamily. *Pharmacol. Rev.* **57**, 387–395 (2005). doi: 10.1124/pr.57.4.13; pmid: 16382097
16. M. Noda *et al*., Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature* **312**, 121–127 (1984). doi: 10.1038/312121a0; pmid: 6209577
17. H. Terlau *et al*., Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. *FEBS Lett.* **293**, 93–96 (1991). doi: 10.1016/0014-5793(91)81159-6; pmid: 1660007
18. I. Favre, E. Moczydlowski, L. Schild, On the structural basis for ionic selectivity among Na+, K+, and Ca2+ in the voltage-gated sodium channel. *Biophys. J.* **71**, 3110–3125 (1996). doi: 10.1016/S0006-3495(96)79505-X; pmid: 8968582
19. R. W. Aldrich, D. P. Corey, C. F. Stevens, A reinterpretation of mammalian sodium channel gating based on single channel recording. *Nature* **306**, 436–441 (1983). doi: 10.1038/306436a0; pmid: 6316158
20. C. M. Armstrong, F. Bezanilla, Currents related to movement of the gating particles of the sodium channels. *Nature* **242**, 459–461 (1973). doi: 10.1038/242459a0; pmid: 4700900
21. C. M. Armstrong, F. Bezanilla, Charge movement associated with the opening and closing of the activation gates of the Na channels. *J. Gen. Physiol.* **63**, 533–552 (1974). doi: 10.1085/jgp.63.5.533; pmid: 4824995
22. N. Yang, R. Horn, Evidence for voltage-dependent S4 movement in sodium channels. *Neuron* **15**, 213–218 (1995). doi: 10.1016/0896-6273(95)90078-0; pmid: 7619524
23. N. Yang, A. L. George Jr., R. Horn, Molecular basis of charge movement in voltage-gated sodium channels. *Neuron* **16**, 113–122 (1996). doi: 10.1016/S0896-6273(00)80028-8; pmid: 8562074
24. C. Miller, Biophysics. Lonely voltage sensor seeks protons for permeation. *Science* **312**, 534–535 (2006). doi: 10.1126/science.1127186; pmid: 16645081
25. C. M. Armstrong, F. Bezanilla, E. Rojas, Destruction of sodium conductance inactivation in squid axons perfused with pronase. *J. Gen. Physiol.* **62**, 375–391 (1973). doi: 10.1085/jgp.62.4.375; pmid: 4755846
26. F. Bezanilla, C. M. Armstrong, Inactivation of the sodium channel. I. Sodium current experiments. *J. Gen. Physiol.* **70**, 549–566 (1977). doi: 10.1085/jgp.70.5.549; pmid: 591911
27. C. M. Armstrong, F. Bezanilla, Inactivation of the sodium channel. II. Gating current experiments. *J. Gen. Physiol.* **70**, 567–590 (1977). doi: 10.1085/jgp.70.5.567; pmid: 591912
28. D. E. Patton, J. W. West, W. A. Catterall, A. L. Goldin, Amino acid residues required for fast Na(+)-channel inactivation: Charge neutralizations and deletions in the III-IV linker. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10905–10909 (1992). doi: 10.1073/pnas.89.22.10905; pmid: 1332059
29. J. W. West *et al*., A cluster of hydrophobic amino acid residues required for fast Na(+)-channel inactivation. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10910–10914 (1992). doi: 10.1073/pnas.89.22.10910; pmid: 1332060
30. A. L. Goldin, Mechanisms of sodium channel inactivation. *Curr. Opin. Neurobiol.* **13**, 284–290 (2003). doi: 10.1016/S0959-4388(03)00065-5; pmid: 12850212
31. W. Ulbricht, Sodium channel inactivation: Molecular determinants and modulation. *Physiol. Rev.* **85**, 1271–1301 (2005). doi: 10.1152/physrev.00024.2004; pmid: 16183913
32. Y. Jiang *et al*., X-ray structure of a voltage-dependent K^+ channel. *Nature* **423**, 33–41 (2003). doi: 10.1038/nature01580; pmid: 12721618
33. S. B. Long, E. B. Campbell, R. Mackinnon, Crystal structure of a mammalian voltage-dependent *Shaker* family K^+ channel. *Science* **309**, 897–903 (2005). doi: 10.1126/science.1116269; pmid: 16002581
34. M. Liao, E. Cao, D. Julius, Y. Cheng, Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013). doi: 10.1038/nature12822; pmid: 24305160
35. J. Wu *et al*., Structure of the voltage-gated calcium channel $Ca_V1.1$ complex. *Science* **350**, aad2395 (2015). doi: 10.1126/science.aad2395; pmid: 26680202
36. H. Shen *et al*., Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science* **355**, eaal4326 (2017). doi: 10.1126/science.aal4326; pmid: 28183995
37. Z. Yan *et al*., Structure of the $Na_V1.4$-β1 complex from electric eel. *Cell* **170**, 470–482.e11 (2017). doi: 10.1016/j.cell.2017.06.039; pmid: 28735751
38. Y. He, K. Wang, N. Yan, The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell* **5**, 658–672 (2014). doi: 10.1007/s13238-014-0086-4; pmid: 25119489
39. J. S. Trimmer *et al*., Primary structure and functional expression of a mammalian skeletal muscle sodium channel. *Neuron* **3**, 33–49 (1989). doi: 10.1016/0896-6273(89)90113-X; pmid: 2559760
40. A. L. George Jr., J. Komisarof, R. G. Kallen, R. L. Barchi, Primary structure of the adult human skeletal muscle voltage-dependent sodium channel. *Ann. Neurol.* **31**, 131–137 (1992). doi: 10.1002/ana.410310203; pmid: 1315496
41. M. Chahine, P. B. Bennett, A. L. George Jr., R. Horn, Functional expression and properties of the human skeletal muscle sodium channel. *Pflugers Arch.* **427**, 136–142 (1994). doi: 10.1007/BF00585952; pmid: 8058462
42. B. Chanda, F. Bezanilla, Tracking voltage-dependent conformational changes in skeletal muscle sodium channel during activation. *J. Gen. Physiol.* **120**, 629–645 (2002). doi: 10.1085/jgp.20028679; pmid: 12407076
43. Y. Muroi, M. Arcisio-Miranda, S. Chowdhury, B. Chanda, Molecular determinants of coupling between the domain III voltage sensor and pore of a sodium channel. *Nat. Struct. Mol. Biol.* **17**, 230–237 (2010). doi: 10.1038/nsmb.1749; pmid: 20118934
44. S. Sokolov, T. Scheuer, W. A. Catterall, Ion permeation and block of the gating pore in the voltage sensor of $Na_V1.4$ channels with hypokalemic periodic paralysis mutations. *J. Gen. Physiol.* **136**, 225–236 (2010). doi: 10.1085/jgp.201010414; pmid: 20660662
45. P. Gosselin-Badaroudine, L. Delemotte, A. Moreau, M. L. Klein, M. Chahine, Gating pore currents and the resting state of $Na_V1.4$ voltage sensor domains. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19250–19255 (2012). doi: 10.1073/pnas.1217990109; pmid: 23134726
46. T. Kubota *et al*., Mapping of voltage sensor positions in resting and inactivated mammalian sodium channels by LRET. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1857–E1865 (2017). doi: 10.1073/pnas.1700453114; pmid: 28202723
47. X.-W. Dong, T. Priestley, Electrophysiological analysis of tetrodotoxin-resistant sodium channel pharmacology. *Curr. Protoc. Pharmacol.* **23**, 11.8.1–11.8.33 (2004). doi: 10.1002/0471141755.ph1108s23 pmid: 21956811
48. J. Webb, F.-f. Wu, S. C. Cannon, Slow inactivation of the NaV1.4 sodium channel in mammalian cells is impeded by co-expression of the β1 subunit. *Pflugers Arch.* **457**, 1253–1263 (2009). doi: 10.1007/s00424-008-0600-8; pmid: 18941776
49. A. Goehring *et al*., Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585 (2014). doi: 10.1038/nprot.2014.173; pmid: 25299155
50. B. Hille, The permeability of the sodium channel to metal cations in myelinated nerve. *J. Gen. Physiol.* **59**, 637–658 (1972). doi: 10.1085/jgp.59.6.637; pmid: 5025743
51. H. Shen *et al*., Structural basis for the modulation of voltage-gated sodium channels by animal toxins. *Science* **2018**, eaau2596 (2018). doi: 10.1126/science.aau2596; pmid: 30049784
52. B. Hille, The permeability of the sodium channel to organic cations in myelinated nerve. *J. Gen. Physiol.* **58**, 599–619 (1971). doi: 10.1085/jgp.58.6.599; pmid: 5315827
53. J. Zhang *et al*., Simulating the ion permeation and ion selection for a eukaryotic voltage-gated sodium channel Na_VPaS. *Protein Cell* **9**, 580–585 (2018). doi: 10.1007/s13238-018-0522-y; pmid: 29532417
54. C. M. Armstrong, Sodium channels and gating currents. *Physiol. Rev.* **61**, 644–683 (1981). doi: 10.1152/physrev.1981.61.3.644; pmid: 6265962
55. P. G. DeCaen, V. Yarov-Yarovoy, Y. Zhao, T. Scheuer, W. A. Catterall, Disulfide locking a sodium channel voltage sensor reveals ion pair formation during activation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15142–15147 (2008). doi: 10.1073/pnas.0806486105; pmid: 18809926
56. P. G. DeCaen, V. Yarov-Yarovoy, E. M. Sharp, T. Scheuer, W. A. Catterall, Sequential formation of ion pairs during activation of a sodium channel voltage sensor. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22498–22503 (2009). doi: 10.1073/pnas.0912307106; pmid: 20007787
57. X. Zhang *et al*., Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. *Nature* **486**, 130–134 (2012). doi: 10.1038/nature11054; pmid: 22678295
58. X. Tao, A. Lee, W. Limapichat, D. A. Dougherty, R. MacKinnon, A gating charge transfer center in voltage sensors. *Science* **328**, 67–73 (2010). doi: 10.1126/science.1185954; pmid: 20360102
59. J. Wu *et al*., Structure of the voltage-gated calcium channel $Ca_V1.1$ at 3.6 Å resolution. *Nature* **537**, 191–196 (2016). doi: 10.1038/nature19321; pmid: 27580036
60. M. P. Goldschen-Ohm, D. L. Capes, K. M. Oelstrom, B. Chanda, Multiple pore conformations driven by asynchronous movements of voltage sensors in a eukaryotic sodium channel. *Nat. Commun.* **4**, 1350 (2013). doi: 10.1038/ncomms2356; pmid: 23322038
61. H. K. Motoike *et al*., The Na^+ channel inactivation gate is a molecular complex: A novel role of the COOH-terminal domain. *J. Gen. Physiol.* **123**, 155–165 (2004). doi: 10.1085/jgp.200308929; pmid: 14744988
62. R. S. Kass, Sodium channel inactivation goes with the flow. *J. Gen. Physiol.* **124**, 7–8 (2004). doi: 10.1085/jgp.200409123; pmid: 15226362
63. R. S. Kass, Sodium channel inactivation in heart: A novel role of the carboxy-terminal domain. *J. Cardiovasc. Electrophysiol.* **17** (suppl. 1), S21–S25 (2006). doi: 10.1111/j.1540-8167.2006.00381.x; pmid: 16686678
64. L. Tang, R. G. Kallen, R. Horn, Role of an S4-S5 linker in sodium channel inactivation probed by mutagenesis and a peptide blocker. *J. Gen. Physiol.* **108**, 89–104 (1996). doi: 10.1085/jgp.108.2.89; pmid: 8854339
65. J. C. McPhee, D. S. Ragsdale, T. Scheuer, W. A. Catterall, A critical role for the S4-S5 intracellular loop in domain IV of the sodium channel α-subunit in fast inactivation. *J. Biol. Chem.* **273**, 1121–1129 (1998). doi: 10.1074/jbc.273.2.1121; pmid: 9422778
66. M. R. Smith, A. L. Goldin, Interaction between the sodium channel inactivation linker and domain III S4-S5. *Biophys. J.* **73**, 1885–1895 (1997). doi: 10.1016/S0006-3495(97)78219-5; pmid: 9336184
67. M. Bouhours *et al*., A1152D mutation of the Na^+ channel causes paramyotonia congenita and emphasizes the role of DIII/S4–S5 linker in fast inactivation. *J. Physiol.* **565**, 415–427 (2005). doi: 10.1113/jphysiol.2004.081018; pmid: 15790667
68. A. I. McClatchey *et al*., Temperature-sensitive mutations in the III–IV cytoplasmic loop region of the skeletal muscle sodium channel gene in paramyotonia congenita. *Cell* **68**, 769–774 (1992). doi: 10.1016/0092-8674(92)90151-2; pmid: 1310898

69. H. Lerche *et al*., Human sodium channel myotonia: Slowed channel inactivation due to substitutions for a glycine within the III-IV linker. *J. Physiol.* **470**, 13–22 (1993). doi: 10.1113/jphysiol.1993.sp019843; pmid: 8308722
70. L. J. Hayward, R. H. Brown Jr., S. C. Cannon, Inactivation defects caused by myotonia-associated mutations in the sodium channel III-IV linker. *J. Gen. Physiol.* **107**, 559–576 (1996). doi: 10.1085/jgp.107.5.559; pmid: 8740371
71. B. G. H. Schoser, J. M. Schröder, T. Grimm, D. Sternberg, W. Kress, A large German kindred with cold-aggravated myotonia and a heterozygous A1481D mutation in the *SCN4A* gene. *Muscle Nerve* **35**, 599–606 (2007). doi: 10.1002/mus.20733; pmid: 17212350
72. T. Matsuda, C. L. Cepko, Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16–22 (2004). doi: 10.1073/pnas.2235688100; pmid: 14603031
73. J. Lei, J. Frank, Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005). doi: 10.1016/j.jsb.2005.01.002; pmid: 15797731
74. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
75. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015). doi: 10.7554/eLife.06980; pmid: 26023829
76. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
77. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi: 10.1016/j.jsb.2012.09.006; pmid: 23000701
78. S. H. W. Scheres, Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015). doi: 10.1016/j.jsb.2014.11.010; pmid: 25486611
79. D. Kimanius, B. O. Forsberg, S. H. W. Scheres, E. Lindahl, Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016). doi: 10.7554/eLife.18722; pmid: 27845625
80. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003). doi: 10.1016/j.jmb.2003.07.013; pmid: 14568533
81. S. Chen *et al*., High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013). doi: 10.1016/j.ultramic.2013.06.004; pmid: 23872039
82. M. A. Larkin *et al*., Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007). doi: 10.1093/bioinformatics/btm404; pmid: 17846036
83. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
84. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
85. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
86. A. Amunts *et al*., Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014). doi: 10.1126/science.1249410; pmid: 24675956
87. W. L. DeLano, The PyMOL Molecular Graphics System (2002); www.pymol.org.
88. O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. P. Sansom, HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996). doi: 10.1016/S0263-7855(97)00009-X; pmid: 9195488

ACKNOWLEDGMENTS

We thank X. Li (Tsinghua University) for technical support during EM image acquisition. **Funding:** This work was funded by the National Key Basic Research (973) Program (2015CB910101 to N.Y.) and the National Key R&D Program (2016YFA0500402 to N.Y. and 2016YFA0501100 to J.L.) from Ministry of Science and Technology of China, the National Natural Science Foundation of China (projects 31621092, 31630017, and 81861138009 to N.Y.), and China Postdoctoral Science Foundation (BX201600089 to X.P.). We thank the Tsinghua University Branch of China National Center for Protein Sciences (Beijing) for providing the cryo-EM facility support. We thank the computational facility support on the cluster of Bio-Computing Platform (Tsinghua University Branch of China National Center for Protein Sciences Beijing) and the "Explorer 100" cluster system of Tsinghua National Laboratory for Information Science and Technology. N.Y. is supported by the Shirley M. Tilghman endowed professorship from Princeton University. **Author contributions:** N.Y. conceived the project. X.P., Z.L., Q.Z., H.S., K.W., X.H., J.C., J.Z., X.Z., and J.L. performed the experiments. All authors contributed to data analysis. N.Y. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The atomic coordinates and EM map for the human $Na_v1.4$-β1 complex have been deposited in the Protein Data Bank (www.rcsb.org) and EMDB (www.ebi.ac.uk/pdbe/emdb/) with accession codes 6AGF and EMD-9617, respectively.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/eaau2486/suppl/DC1
Figs. S1 to S8
Tables S1 and S2

22 May 2018; accepted 20 August 2018
10.1126/science.aau2486

RESEARCH ARTICLE SUMMARY

ION CHANNELS

Structural basis for the modulation of voltage-gated sodium channels by animal toxins

Huaizong Shen*, Zhangqiang Li*, Yan Jiang*, Xiaojing Pan*, Jianping Wu†, Ben Cristofori-Armstrong, Jennifer J. Smith, Yanni K.Y. Chin, Jianlin Lei, Qiang Zhou‡, Glenn F. King‡, Nieng Yan†‡

INTRODUCTION: Almost all venoms contain toxins that modulate the activity of voltage-gated sodium (Na_v) channels in order to incapacitate prey or predators. The single-chain eukaryotic Na_v channels comprise four homologous repeats. The central pore domain is constituted by the carboxyl-terminal segments from all four repeats, and each repeat also has a voltage-sensing domain (VSD). Toxins are broadly divided into two categories—pore blockers that physically occlude the channel pore and gating modifiers that alter channel gating by interfering with the VSDs. Whereas small-molecule neurotoxins such as tetrodotoxin (TTX) and saxitoxin (STX) function as pore blockers, most peptidic Na_v channel toxins are gating modifiers that trap the channel in a particular stage of the gating cycle through interactions with one or more VSDs. In neither case is the structural basis of channel modulation fully understood.

RATIONALE: Dc1a is a peptidic gating modifier toxin (GMT) from venom of the desert bush spider *Diguetia canities* that specifically binds to VSD_{II} of insect Na_v channels to promote channel opening. We showed through biochemical analysis that Dc1a interacts with Na_vPaS, a Na_v channel from the American cockroach *Periplaneta americana*, for which a cryo–electron microscopy (cryo-EM) structure was recently determined at 3.8-Å resolution. We therefore sought to solve the structure of the complex between Na_vPaS and Dc1a. As Dc1a occupies a distinctly different channel binding site to pore blockers, we also attempted to supplement the complex with TTX or STX to obtain structures of the ternary complexes.

RESULTS: The cryo-EM structure of Na_vPaS-Dc1a was determined to an overall resolution of 2.8 Å in the presence of 300 mM NaCl, whereas those of Na_vPaS-Dc1a-TTX and Na_vPaS-Dc1a-STX were resolved at 2.6 Å and 3.2 Å, respectively, in the presence of 150 mM NaCl.

VSD_{II} constitutes the primary docking site for Dc1a, which undergoes considerable structural rearrangement upon binding to the channel. The toxin inserts into the cleft between VSD_{II} and the pore region, making intimate contacts with both domains. The network of intermolecular interactions seen in the cryo-EM structure was validated through examination of the effect of toxin and channel mutations using the orthologous Na_vBg channel from the German cockroach *Blattella germanica*.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau2596

Four residues, Asp/Glu/Lys/Ala (DEKA), at a corresponding locus in the selectivity filter (SF) of each repeat confer Na^+ selectivity. A Na^+ ion was observed in the same position in the structures of Na_vPaS-Dc1a and Na_vPaS-Dc1a-TTX, coordinated by the Asp and Glu residues in the DEKA motif of the SF, and an invariant Glu on the P2 helix in repeat II, a helix in the entryway to the SF on the extracellular side. Both TTX and STX form extensive electrostatic interactions with residues in the outer electronegative ring that attracts cations into the SF and Asp and Glu in the DEKA motif, completely blocking access of Na^+ ions to the SF.

CONCLUSION: The structure of the Na_vPaS-Dc1a complex suggests that the network of interactions between Na_v channels and GMTs is more complex than previously anticipated. Therefore, caution has to be applied when using isolated Na_v channel VSDs for drug discovery or for understanding the molecular basis of GMT action. The current structures elucidate the molecular basis for the insect selectivity of Dc1a and the subtype-specific binding of TTX or STX to Na_v channels. Unambiguous structural elucidation of the bound TTX and STX, whose molecular weights are both around 300 Da, showcases the power of cryo-EM and its potential for structure-aided drug discovery. ■

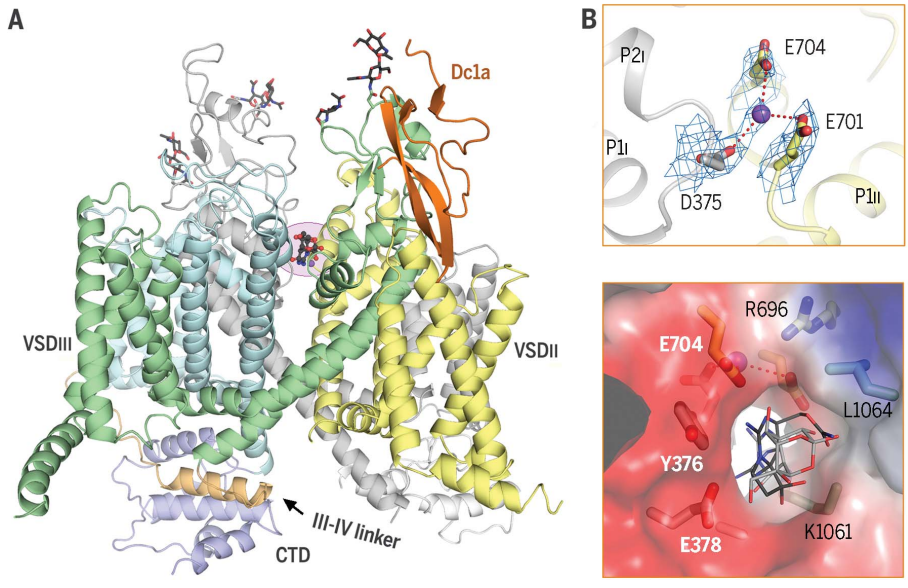


Fig. 1. Structural basis for specific binding of GMT Dc1a and guanidinium pore blockers TTX and STX by Na_vPaS. (**A**) Dc1a inserts into the extracellular cavity between VSD_{II} and the pore elements of repeat III. (**B**) Molecular mechanism for pore blockade by TTX and STX. Top: The carboxylate groups of Asp (D) and Glu (E) residues in the DEKA motif and an invariant Glu on $P2_{II}$ together constitute a potential Na^+ binding site (designated the DEE site). Bottom: TTX and STX block access of Na^+ to the DEE site from the extracellular side. A semitransparent presentation of the electrostatic surface potential of the entrance to the SF viewed from the extracellular side is shown. CTD, C-terminal domain; R, Arg; L, Leu; Y, Tyr; K, Lys.

*These authors contributed equally to this work.
†Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08540, USA.
‡Corresponding author. Email: nyan@princeton.edu (N.Y.); glenn.king@imb.uq.edu.au (G.F.K.); zhouq2015@mail.tsinghua.edu.cn (Q.Z.)
Cite this article as H. Shen *et al.*, *Science* **362**, eaau2596 (2018). DOI: 10.1126/science.aau2596

RESEARCH ARTICLE

ION CHANNELS

Structural basis for the modulation of voltage-gated sodium channels by animal toxins

Huaizong Shen[1,2,3]*, Zhangqiang Li[1,2,3]*, Yan Jiang[4]*, Xiaojing Pan[1,2,3]*, Jianping Wu[1,2,3]†, Ben Cristofori-Armstrong[4], Jennifer J. Smith[4], Yanni K. Y. Chin[4], Jianlin Lei[5], Qiang Zhou[1,2,3]‡, Glenn F. King[4]‡, Nieng Yan[1,2,3]†‡

Animal toxins that modulate the activity of voltage-gated sodium (Na_v) channels are broadly divided into two categories—pore blockers and gating modifiers. The pore blockers tetrodotoxin (TTX) and saxitoxin (STX) are responsible for puffer fish and shellfish poisoning in humans, respectively. Here, we present structures of the insect Na_v channel Na_vPaS bound to a gating modifier toxin Dc1a at 2.8 angstrom-resolution and in the presence of TTX or STX at 2.6-Å and 3.2-Å resolution, respectively. Dc1a inserts into the cleft between VSD_{II} and the pore of Na_vPaS, making key contacts with both domains. The structures with bound TTX or STX reveal the molecular details for the specific blockade of Na^+ access to the selectivity filter from the extracellular side by these guanidinium toxins. The structures shed light on structure-based development of Na_v channel drugs.

Voltage-gated sodium (Na_v) channels play a critical role in generating membrane excitability (*1*) and are targeted by numerous chemical insecticides and human drugs. Na_v channels are also the most common target of venom neurotoxins. Na_v channels comprise one single polypeptide chain that folds to four homologous repeats (repeats I to IV), each containing six transmembrane helices designated S1 to S6. The S1 to S4 segments in each repeat constitute the voltage-sensing domain (VSD), and the S5, S6, and their intervening segments from the four repeats together enclose the ion-conducting pore domain.

Although small-molecule neurotoxins such as tetrodotoxin (TTX) and saxitoxin (STX) function as pore blockers, the vast majority of peptidic Na_v channel toxins are gating modifiers that trap the channel in a particular stage of the gating cycle through interactions with one or more VSDs (*2*). In contrast to pore blockers, gating modifier toxins (GMTs) have more complex allosteric effects on Na_v channel function, and they can inhibit (*3*) or agonize (*4*) the channel. GMTs, which generally have greater selectivity than pore blockers, are valuable leads for the development of subtype-selective Na_v channel drugs (*3*, *5*, *6*).

Despite extensive studies of the molecular basis by which GMTs modulate Na_v channel function, no consensus model of this interaction has emerged. Early studies suggested a dominant role for the extracellular S3-S4 loop in GMT binding (*7*, *8*), but subsequent studies have revealed a key role for the S1-S2 loop in many GMT interactions (*4*, *9*, *10*). More recent studies suggest that GMTs nestle into an extracellular-facing cavity between the S1 and S4 helices, enabling them to act as a wedge that impedes voltage sensor movement (*5*, *11*). It has been suggested that large GMTs such as those found in scorpion venom might be able to simultaneously contact the VSD and the extracellular loop connecting the pore helix P2 and the S6 segment in pore domain (*12*), but no studies to date have predicted a role for any of the pore-domain membrane helices in GMT binding.

Small molecules that occlude the pore of Na_v channels are rare in animal venoms, but TTX and STX are exceptions. As the name indicates, TTX was originally found in tetrodontoid fish exemplified by the puffer fish (fugu). Puffer fish poisoning, resulting from consumption of TTX-containing fish, was documented thousands of years ago in China and Egypt and later in Japan and Mexico (*13*, *14*). TTX was subsequently shown to be present in venom of the deadly blue-ringed octopus, in the poisonous secretions of frogs and newts, and in predatory moon snails; these animals do not synthesize TTX but rather acquire it from endosymbiotic bacteria (*15*). It was discovered in the mid–20th century that the potent toxicity of TTX is due to suppression of action potential generation through specific inhibition of Na^+ influx (*14*, *16*–*18*). STX is a related guanidinium neurotoxin, produced by marine dinoflagellates and cyanobacteria, that competes with TTX for binding to Na_v channels (*15*). The term saxitoxin is also used to refer to a class of >50 structurally related toxins that are responsible for paralytic shellfish poisoning (*19*, *20*).

Because of their stringent specificity for Na_v channels, TTX and STX are widely used for pharmacological characterization of Na_v channels (*21*–*24*). The nine subtypes of mammalian Na_v channels are classified as TTX resistant or TTX sensitive based on their susceptibility to TTX. The latter are inhibited by nanomolar concentrations of TTX, whereas the TTX-resistant subtypes $Na_v1.5$, $Na_v1.8$, and $Na_v1.9$ only respond to micromolar concentrations of the toxin (*24*, *25*). Despite comprehensive studies over the past six decades (*26*–*29*), our molecular understanding of the mechanism of action of these toxins has been impeded by the lack of structural information. Crystal structures of several bacterial Na_v channels have been elucidated (*30*–*32*), but these homotetrameric prokaryotic orthologs are insensitive to TTX and STX because they lack the receptor site found in their single-chain, asymmetric eukaryotic counterparts (*33*).

We recently elucidated the structure of the eukaryotic Na_v channel Na_vPaS from the American cockroach *Periplaneta americana* at 3.8-Å resolution (*34*). Here, we present a 2.8-Å resolution cryo–electron microscopy (cryo-EM) structure of this channel in complex with Dc1a, a peptidic GMT from venom of the desert bush spider *Diguetia canities* that promotes channel opening (*10*). We also report cryo-EM structures of the Na_vPaS-Dc1a complex in the presence of the pore blockers TTX and STX at 2.6 Å and 3.2 Å, respectively. A Na^+ binding site in the selectivity filter (SF) constituted by three carboxylate groups is observed. The structures elucidate the molecular basis for pore blockade by TTX and STX.

Results

Structural determination of Na_vPaS in complex with Dc1a, TTX, and STX

Details of cryo-sample preparation, image acquisition, data processing, model building, and structure refinement can be found in the materials and methods. Briefly, micrographs collected on a Titan Krios electron microscope equipped with Gatan K2 Summit detector, GIF Quantum energy filter, and spherical aberration (Cs) image corrector were used to reconstruct a three-dimensional (3D) EM map for the Na_vPaS-Dc1a complex purified in the presence of 300 mM NaCl to an overall resolution of 2.8 Å. Following a similar protocol, the structures of Na_vPaS-Dc1a bound to TTX and STX were obtained at 2.6 and 3.2 Å, respectively. The central region of Na_vPaS

[1]State Key Laboratory of Membrane Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. [2]Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. [3]Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. [4]Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Queensland 4072, Australia. [5]Technology Center for Protein Sciences, Ministry of Education Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.
†Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08540, USA.
‡Corresponding author. Email: nyan@princeton.edu (N.Y.); glenn.king@imb.uq.edu.au (G.F.K.); zhouq2015@mail.tsinghua.edu.cn (Q.Z.)

exhibits higher resolution in all three structures. Application of a mask for the central region during postprocessing further improved the resolution of this region to 2.7 Å for Na_vPaS-Dc1a and 3.1 Å for Na_vPaS-Dc1a-STX, whereas that for Na_vPaS-Dc1a-TTX remained at 2.6 Å (Fig. 1, A and B; figs. S1 to S4; and table S1). The excellent quality of the EM maps ensured reliable assignment of the ligands and surrounding residues. All residues of the SF, including the invariant residues from the four repeats, Asp/Glu/Lys/Ala (DEKA), and surrounding segments are unambiguously resolved in the high-resolution EM reconstructions (Fig. 1C and fig. S4).

VSD_{II} and the pore domain together accommodate Dc1a

The structure of the Dc1a-Na_vPaS complex (Fig. 2, A to C) confirms that VSD_{II} constitutes the primary docking site for Dc1a, as we reported previously (*10*). Comparison with the ligand-free Na_vPaS structure (*34*) reveals minor conformational changes in the channel upon Dc1a binding, mainly affecting VSD_{II} (fig. S5). In contrast, the structure of Dc1a undergoes considerable rearrangement (Fig. 2B). The nuclear magnetic resonance (NMR) structure of Dc1a alone contains five short β strands that are organized into an N-terminal β sheet and a C-terminal inhibitor cystine knot (knottin) motif (*10*). In the complex, however, the two C-terminal β strands extend into the previously unstructured connecting loop region to form an elongated β hairpin that inserts deeply into the extracellular cavity enclosed by the four segments in VSD_{II} and the adjacent pore-forming S5 segment from repeat III ($S5_{III}$; Fig. 2, A and B).

Dc1a makes extensive polar and hydrophobic interactions with Na_vPaS that are more complex than predicted by any model of GMT binding, involving interactions with the S1-S2 loop (the loop that connects S1 and S2), the extracellular pore loops, and the S5 pore-domain helix of repeat III. The toxin makes no interactions with the S3-S4 loop. The edge of the β sheet of Dc1a interacts with the short extracellular helix in repeat III above the pore domain (designated EαIII; Fig. 2B); specifically, Tyr^{33} and Asp^{56} on Dc1a interact with His^{1032} and Arg^{1027} on Na_vPaS, respectively (Fig. 2C, left). On one side of the VSD_{II} cavity, the toxin interacts extensively with the S1-S2 loop (designated $L1\text{-}2_{II}$); in particular, the guanidinium group of Dc1a-Arg^{41} interacts with the main-chain carbonyl oxygen and the side-chain carboxyl group of Asp^{542} (Fig. 2C, middle).

The β3-β4 hairpin of Dc1a inserts deeply into the VSD_{II} cavity, with Phe^{47} and Phe^{48} at the tip of the hairpin surrounded by hydrophobic residues from $S1_{II}$ and the side wall of the pore domain involving $S5_{III}$. Meanwhile, the aromatic ring of Dc1a-Phe^{48} makes a π-cation interaction with the gating-charge residue Arg^{613} (R3; Fig. 2C, right). Gln^{1002} on $S5_{III}$ makes extensive polar interactions with the side chain of Dc1a-Lys^{44} and the backbone amide of Dc1a-Phe^{48}. These specific interactions with both VSD_{II} and the pore domain collectively stabilize VSD_{II} in the "up" state, consistent with Dc1a inducing opening of the channel (*10*).

We investigated the importance of these intermolecular interactions by examining the effect of toxin and channel mutations using the orthologous Na_vBg channel from the German cockroach *Blattella germanica* (Fig. 2, D and E). Na_vBg is potently activated by Dc1a, but unlike Na_vPaS, it is amenable to electrophysiological analysis (*10*). Modulation of Na_vBg activity by Dc1a was almost abolished when residues Asp^{21}, Tyr^{33}, Arg^{41}, Lys^{44}, and Asp^{56} were mutated to Ala, whereas Ala substitutions of Phe^{47}, Phe^{48}, and Ser^{49} severely diminished but did not completely abrogate Dc1a activity (Fig. 2D and fig. S6). On the basis of ^1H NMR chemical shifts, none of these mutations perturb the structure of Dc1a (figs. S7 and S8); thus, we conclude that they all contribute to Dc1a modulation of insect Na_v channels.

Mutation of Na_vBg residues involved in Dc1a binding caused minor shifts in the conductance-voltage (*G-V*) relationship for the channel (fig. S9). Thus, for each channel mutant, we quantified the previously noted ability of Dc1a to induce a hyperpolarizing shift ($\Delta_{G\text{-}V}$) in the *G-V* relationship (*10*), thereby allowing each mutant channel to serve as its own control (Fig. 2E and fig. S9). Both conservative (D→E) and harsher (D→A) mutations of Asp^{805} and Asp^{808} in $L1\text{-}2_{II}$ (corresponding to Asp^{539} and Asp^{542} in Na_vPaS) greatly reduced $\Delta_{G\text{-}V}$, highlighting the critical importance of these residues to Dc1a binding. Notably, neither residue is conserved in mammalian Na_v channels (*34*), providing a molecular rationale for the insect selectivity of Dc1a (*10*). Mutation of Arg^{1447} to Lys in EαIII (Arg^{1027} in Na_vPaS) reduced Dc1a activity, providing support for this unexpected toxin-channel interaction, whereas mutation of His^{1452} in EαIII (His^{1032} in Na_vPaS) had minimal impact. Consistent with the cryo-EM structure (Fig. 2C), mutation of Arg^{876} (Arg^{610} in Na_vPaS), one of the uppermost gating-charge residues, had minimal impact on Dc1a activity, indicating that this residue is not crucial for the Dc1a-Na_vPaS interaction. Last, a conservative mutation of Gln^{1422} (Gln^{1002} in Na_vPaS) on $S5_{III}$ to Asn greatly reduced toxin activity (Fig. 2E), consistent with the interactions observed between this pore domain residue and residues Lys^{44} and Phe^{48} in Dc1a (Fig. 2C).

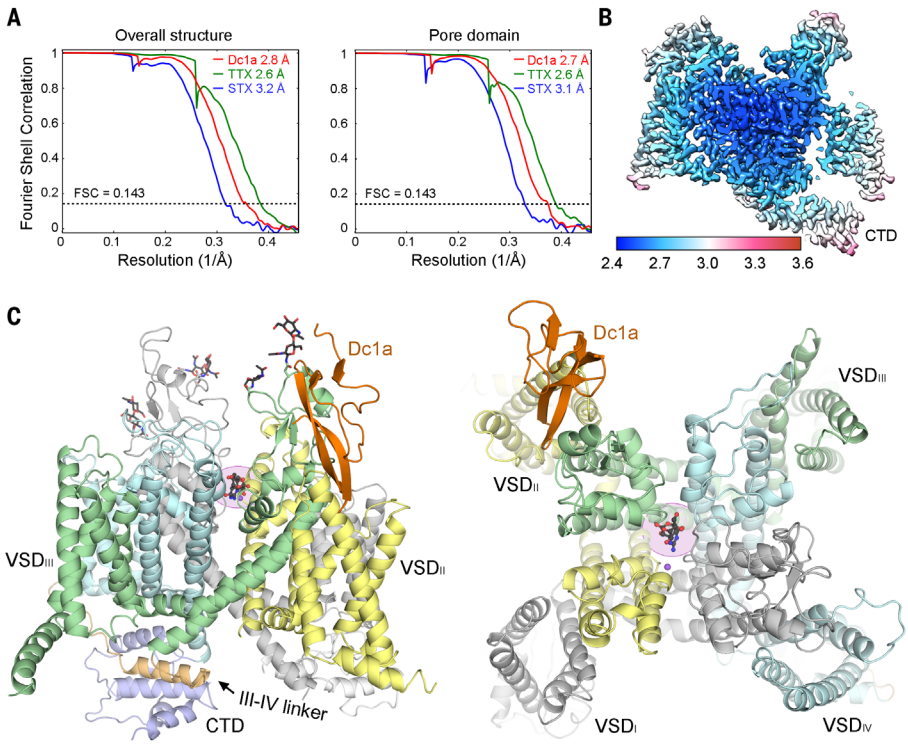


Fig. 1. Structures of the complex between Na_vPaS and the peptide toxin Dc1a with or without TTX or STX. (**A**) Gold standard Fourier shell correlation (FSC) curves for the 3D EM reconstructions of the Na_vPaS-Dc1a complex in the absence or presence of TTX or STX. Left: FSC curves for the overall structures. Right: FSC curves for the pore domains that were masked during postprocessing. (**B**) Local resolution map of the Na_vPaS-Dc1a-TTX complex. The map was estimated with RELION 2.0 and generated in Chimera. CTD, C-terminal domain. (**C**) Overall structure of the Na_vPaS-Dc1a-TTX complex. Side view and top view are shown. Because the three overall structures are nearly identical, only one is shown as a representative. The four repeats in Na_vPaS are shown in different colors, and Dc1a is colored orange. The sugar moieties are shown as black sticks. TTX, shown as black ball and sticks, is highlighted by the pink shade. The putative Na^+ ion is shown as purple sphere. All structure figures were prepared with PyMOL (*67*).

Notably, the corresponding residue is Asn in human $Na_v1.1$–1.8 and Tyr in $Na_v1.9$, again consistent with the insect selectivity of Dc1a (*10*).

In summary, the mutagenesis data provide strong support for the physiological relevance of the complex network of intermolecular interactions observed in the Dc1a-Na_vPaS structure.

Recognition of TTX

TTX and STX, both of which have a molecular weight of ~300 Da, are clearly resolved in the cryo-EM reconstructions (Figs. 3 and 4 and fig. S4). TTX contains one guanidinium, two ether bonds, one oxygen anion, and multiple hydroxyl groups (Fig. 3A and fig. S4A). At lower pH, the oxygen anion is protonated. The map also reveals a density that likely belongs to a coordinated Na^+ (fig. S4C), which we discuss further below.

TTX blocks the entrance to the SF vestibule through an extensive network of electrostatic interactions. The invariant acidic residues on the corresponding locus of the P2 segment in repeats I, II, and IV each form multiple hydrogen bonds or salt bridges with the polar groups of TTX (Fig. 3, B to D). Asp^{375} and Glu^{701} in the DEKA motif also directly contribute to TTX binding (Fig. 3C). Three consecutive backbone amides of the residues that demarcate the P2 helix from the preceding SF loop in repeat III simultaneously bind to C10-OH and the oxygen atom of the ether bond between C7 and C10, whereas Trp^{1063} and the carbonyl oxygen of Phe^{1060} coordinate C9-OH (Fig. 3, B and C).

All of the aforementioned residues are invariant in human Na_v channels (Fig. 3D). Tyr^{376} on repeat I is positioned adjacent to the guanidinium group in TTX, enabling it to make π-cation interactions with the toxin (Fig. 3, B and C). The corresponding locus is occupied by either Phe or Tyr in TTX-sensitive Na_v channel subtypes but replaced by Cys or Ser in the TTX-resistant subtypes $Na_v1.5$, $Na_v1.8$, and $Na_v1.9$ (Fig. 3D). The structure therefore explains why substitution of Cys with Tyr at this position in $Na_v1.5$ confers TTX sensitivity (*24*, *35*–*37*). The position after the invariant Glu on the first helical turn of $P2_I$ is occupied by Arg or Lys in TTX-resistant subtypes, whereas an Asn residue occupies this locus in TTX-sensitive channels (Fig. 3D). Although this residue is too far away to directly participate in TTX coordination, a basic residue at this site may reduce the local electronegativity and further lower channel affinity for TTX.

Recognition of STX

The functional groups of STX include the 1,2,3- and 7,8,9-guanidinium groups, the C12 hydroxyls (C12-OHs), and the 13-carbamoyl group. The distinctive shape of this small molecule allows reliable structural assignment (Fig. 4A and fig. S4, B and D). Polar and charged residues from all four repeats that are positioned at the outer entrance to the SF form extensive interactions with the functional groups of STX (Fig. 4, A and B).

The acidic residues on the first helical turn of the P2 helix in each repeat, which together constitute the outer electronegative ring, provide the primary docking site for STX. The 7,8,9- and 1,2,3-guanidinium groups are respectively bound to the invariant Glu residues on the P2 helix in repeats I and II, whereas the carbamoyl and C12-OH groups are coordinated by polar groups in repeat III and the invariant Asp in repeat IV, respectively. Tyr^{376} in repeat I contributes to coordination of the toxin through π-cation interaction with the 1,2,3-guanidinium group of STX (Fig. 4B). The carbonyl oxygen in the carbamoyl group and one adjacent C12-OH are hydrogen bonded to the backbone amide of the invariant Gly (Gly^{1062} in Na_vPaS) in repeat III. The DEKA-motif residue Glu^{701} in repeat II forms a hydrogen bond with N7 of STX (Fig. 4B, right).

All of the STX-coordinating residues in Na_vPaS, except for Tyr^{376} and Gln^{1065}, are conserved in mammalian Na_v channels (Fig. 4C). The corresponding residue for Na_vPaS-Gln^{1065} is Asp in

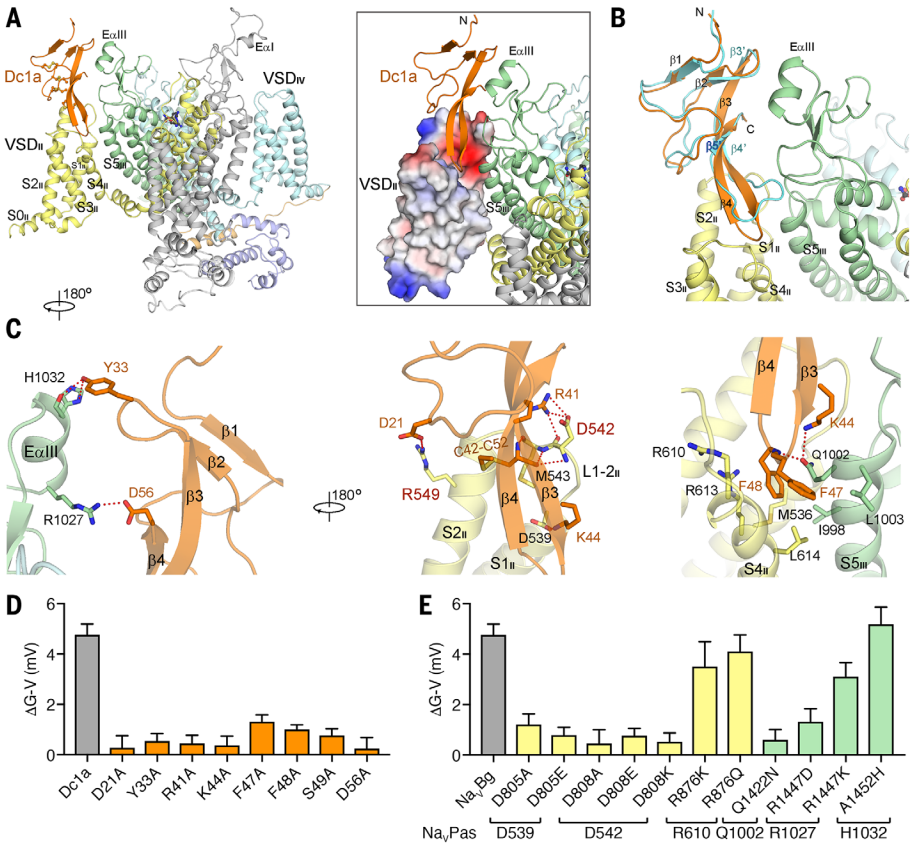


Fig. 2. The interaction between Dc1a and Na_vPaS. (**A**) Dc1a inserts into the extracellular cavity between VSD_{II} and the pore elements of repeat III. The four disulfide bonds in Dc1a are shown as ball and sticks. EαIII: Extracellular α helix in repeat III. Inset: VSD_{II} is shown as surface electrostatic potential calculated in PyMOL. (**B**) Conformational changes of Dc1a upon binding to Na_vPaS. The NMR-determined structure of free Dc1a (cyan) contains five short β strands, with β4 and β5 connected by a flexible linker. When binding to Na_vPaS, the segments containing β3-β5 (labeled as β3′-β5′ to be distinguished from those in the complex structure) become rigidified to form an elongated β hairpin. (**C**) Specific interactions between Dc1a and Na_vPaS. Electrostatic interactions are shown as red dashed lines. The three panels illustrate the contacts from top to bottom. Left: Interactions between Dc1a and the extracellular segments above the pore domain in repeat III of Na_vPaS. Middle: Interactions between Dc1a and the L1-2_{II} loop (the loop that connects the S1 and S2 segments in VSD_{II}). Asp^{542} and Arg^{549}, which are not conserved in mammalian Na_v channels, are highlighted with red labels. Right: Interactions of Dc1a with $S4_{II}$ and $S5_{III}$. (**D** and **E**) Structure-guided mutagenesis characterizations corroborate the structural observations. (D) Shift in *G-V* curve ($\Delta_{G\text{-}V}$) for Na_vBg induced by wild-type (WT) and mutant Dc1a peptides (1 μM). WT Dc1a is shown in gray, whereas mutants are colored orange (n = 5 to 7). (E) Shift in *G-V* curve ($\Delta_{G\text{-}V}$) for WT and mutant Na_vBg channels induced by WT Dc1a (1 μM). Mutants are labeled according to the sequence of WT Na_vBg, with corresponding Na_vPaS numbering below. WT channel is shown in gray, whereas residues located in VSD_{II} and S5-$S6_{III}$ are colored yellow and green, respectively (n = 5 to 6). All data are means ± SEM. Please refer to figs. S6 to S9 for experimental details. Single-letter abbreviations for amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

all human Na_v subtypes except $Na_v1.7$, where this position is occupied by Ile. The Na_vPaS-STX structure provides a molecular basis for the lower affinity of STX for $Na_v1.7$ than for $Na_v1.4$ (*38*), as replacement of Asp with a hydrophobic Ile at this locus would lead to a loss of electrostatic interactions with the carbamoyl amine of STX (Fig. 4, B and C).

Molecular mechanism for pore blockade by TTX and STX

The cryo-EM structures of Na_vPaS in complex with TTX or STX reveal the details of their interaction with the channel. However, elucidation of their mechanism of action also requires a molecular understanding of Na^+ permeation through the SF. Our recent molecular dynamics (MD) simulation of the pore domain of Na_vPaS suggests a preferred path involving the acidic residues from repeats I and II (fig. S10) (*39*). Examination of the EM map for the Na_vPaS-Dc1a complex, which was purified in the presence of 300 mM NaCl, identified a strong density encaged by three acidic residues, Asp^{375} and Glu^{701} from the DEKA motif and Glu^{704} on $P2_{II}$, which is positioned above Glu^{701} (Fig. 5A). Coordination of this density is nearly identical to that observed in the 2.6-Å reconstruction of the Na_vPaS-Dc1a-TTX complex purified in 150 mM NaCl (fig. S4C). The stable conformation of the three carboxylate side chains suggests that they may be stabilized by a cation. The density therefore likely belongs to a Na^+ ion rather than a water molecule. In addition, this site coincides with the energetic minimum observed in the MD simulation of Na^+ penetration through the SF of Na_vPaS. We therefore assigned a Na^+ ion to this density and refer to this Na^+ binding site as the "DEE site" (Fig. 5A).

The preference of Na^+ for the DEE site can be explained by the distinct chemical compositions of the four repeats (Fig. 5B, left). The invariant Arg on $P2_{II}$ (Arg^{696} in Na_vPaS), the Lys in the DEKA motif on repeat III, and a hydrophobic residue on $P2_{III}$ (Leu^{1064} in Na_vPaS and Met in human Na_v channels) may together repel cations to the DEE site (Fig. 5B and fig. S10). Placement of TTX or STX at the entrance to the SF vestibule preserves the configuration of the DEE site but completely blocks access of Na^+ to this site from the extracellular milieu (Fig. 5B).

SUMMARY

In this study, we report the structures of a eukaryotic Na_v channel, Na_vPaS, in complex with three natural toxins. The structures of Na_vPaS in complex with the well-characterized neurotoxins TTX and STX provide a molecular explanation for a wealth of functional studies (*39*–*48*). It is noteworthy that the molecular weights of TTX and STX are both around 300 Da. The clear resolution of these small molecules bound to a Na_v channel with datasets collected in just a few days showcases the power of cryo-EM, which is likely to play an increasingly important role in structure-aided drug discovery.

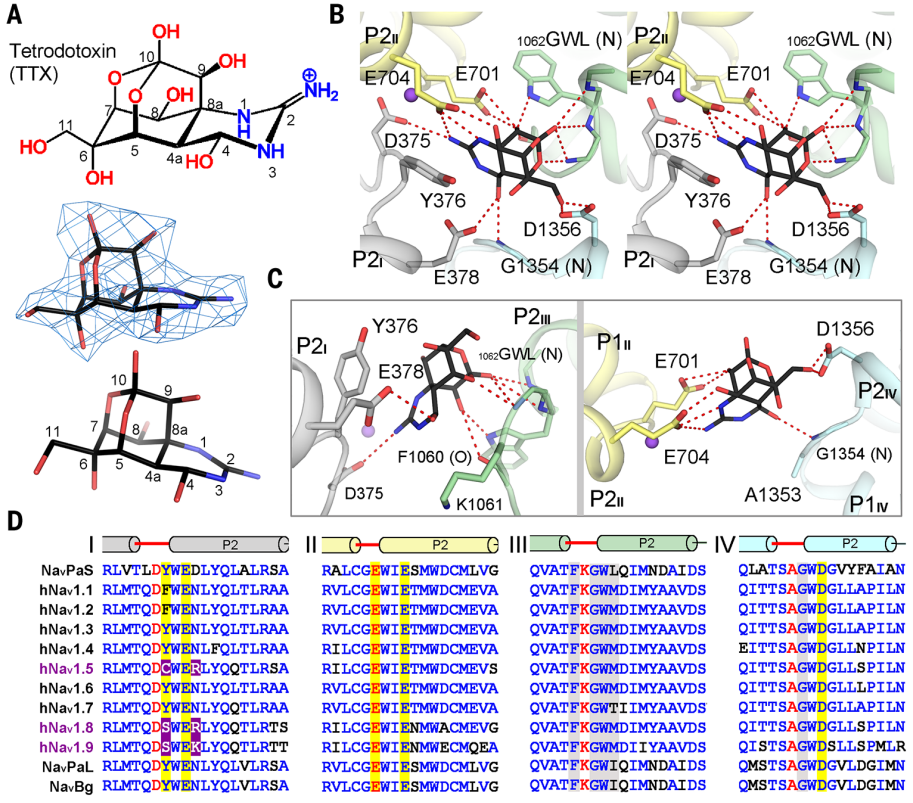


Fig. 3. Specific interactions between Na_vPaS and TTX. (**A**) Structure of TTX. Top: Chemical structure of TTX. Middle: The density for TTX, shown as blue mesh, is contoured at 10 σ. Bottom: Resolved 3D structure of TTX bound to Na_vPaS. (**B**) TTX is specifically coordinated by acidic residues and backbone amides at the outer vestibule of the SF. A stereo view from the extracellular side is shown. The putative Na^+ ion is shown as purple sphere. (**C**) Detailed coordination of TTX by residues from the diagonal repeats shown in side views. (**D**) Sequence alignment of the SF elements and P2 helices in the four channel repeats. The panel is adapted from the reported sequence alignment (*34*). The residues whose side chains are involved in TTX coordination via polar interactions are shaded yellow. The residues whose backbone amides bind to the oxygen anion are shaded gray. In the TTX-resistant Na_v subtypes, the equivalent of Tyr^{376}, which appears to form π-cation interaction with the 1,2,3-guanidinium group of TTX, is Cys ($Na_v1.5$) or Ser ($Na_v1.8$ and $Na_v1.9$).

The structure of the Na_vPaS-Dc1a complex confirmed the important role of VSD_{II} in binding this GMT. However, it also revealed that the network of intermolecular interactions is much more complex than previously anticipated, with key interactions between the toxin and both the $S5_{III}$ pore-domain helix and the extracellular dome above the pore. Thus, one has to apply caution when using isolated Na_v channel VSDs for drug discovery or for understanding the molecular basis of GMT action. Last, the recently determined structure of the $EeNa_v1.4$-β1 complex revealed that the extracellular dome provides a docking site for β subunits (*49*), which, together with the current structure, might explain why the sensitivity of Na_v channels to some GMTs is modulated by the presence of an accessory β subunit (*50*).

Materials and methods

Purification of Na_vPaS in complex with toxins

Recombinant Na_vPaS and Dc1a proteins were expressed and purified as reported (*10*, *34*). To assemble the complex, Dc1a (40 μM) was added to the concentrated Na_vPaS solution and incubated at 4°C for 0.5 hours before size exclusion chromatography (SEC, Superose 6 10/300 GL GE Healthcare). For Na_vPaS-Dc1a complexes bound to TTX or STX, TTX (50 μM) or STX (4 μM) were respectively added to the concentrated Na_vPaS solution 15 min before adding Dc1a. The peak fractions of size exclusion chromatography were pooled and concentrated to approximately 2 mg/ml. For the sample without TTX or STX, 300 mM NaCl was used during purification while for the sample with TTX or STX, 150 mM NaCl was used.

Production of Dc1a analogs

Plasmids encoding Dc1a mutants were generated via PCR-based mutagenesis using a plasmid encoding wild-type Dc1a [pLIC-NSB3; (*10*)] as template. The DNA sequence of all mutants was confirmed by Sanger sequencing. Peptide concentrations were determined by calculating the area under the RP-HPLC peak (at 214 nm) of all analogs, then comparing these to the peak area

obtained from a Dc1a standard whose concentration had been determined by amino acid analysis.

Cryo-EM data acquisition

Cryo-EM samples were prepared as described (*34*). In brief, aliquots (3.5 μl) of freshly purified Na_vPaS complex were placed on glow-discharged holey carbon grids (Quantifoil Cu R1.2/1.3), which were blotted for 3.5 s and flash-frozen in liquid ethane cooled by liquid nitrogen with a Vitrobot Mark IV (Thermo Fisher Scientific Inc.). The grids were subsequently transferred to a Titan Krios (Thermo Fisher Scientific Inc.) electron microscope operating at 300 kV equipped with Cs-corrector (Thermo Fisher Scientific Inc.), Gatan K2 Summit detector and GIF Quantum energy filter. A total of 2764, 3050 or 4539 movie stacks, for Na_vPaS-Dc1a complex, TTX or STX-supplemented samples respectively, were automatically collected using AutoEMation (*51*) with a slit width of 20 eV on the energy filter and a preset defocus range from −1.8 μm to −1.5 μm in superresolution mode at a nominal magnification of 105,000×. Each stack was exposed for 5.6 s with an exposing time of 0.175 s per frame, resulting in a total of 32 frames per stack. The total dose rate was about 48 e/Å^2 for each stack. The stacks were motion corrected with MotionCor2 (*52*) and binned two fold, resulting in a pixel size of 1.091 Å/pixel. Meanwhile, dose weighting was performed (*53*). The defocus values were estimated with Gctf (*54*).

Image processing

The procedure for data processing is summarized in fig. S2. A total of 895,227, 1,506,774, or 1,211,302 particles, for Na_vPaS-Dc1a complex, TTX or STX-supplemented samples respectively, were automatically picked using RELION (*55*–*58*) and Gautomatch (K. Zhang, www.mrc-lmb.cam.ac.uk/kzhang/). After 2D classification, a total of 838,471, 1,042,430 or 447,993 particles were selected for the Na_vPaS-Dc1a complex, TTX or STX-supplemented samples and subjected to global angular searching 3D classification with only one class. For each of the last several iterations of the global angular searching 3D classification, a local angular searching 3D classification was performed, during which the particles were classified into 4 classes. A total of nonduplicated 595,020, 742,093 or 378,538 particles were selected from the local angular searching 3D classification for Na_vPaS-Dc1a complex, TTX or STX-supplemented samples, respectively. For the Na_vPaS-Dc1a complex and STX-supplemented samples, the particles were subjected to multi-reference classification to remove bad particles. Then the good particles were subjected to 3D auto-refinement with 255,265, 742,093, or 166,805 particles for Na_vPaS-Dc1a complex, TTX or STX-supplemented samples, respectively. The 3D auto-refinements were further optimized with a larger box size of 320

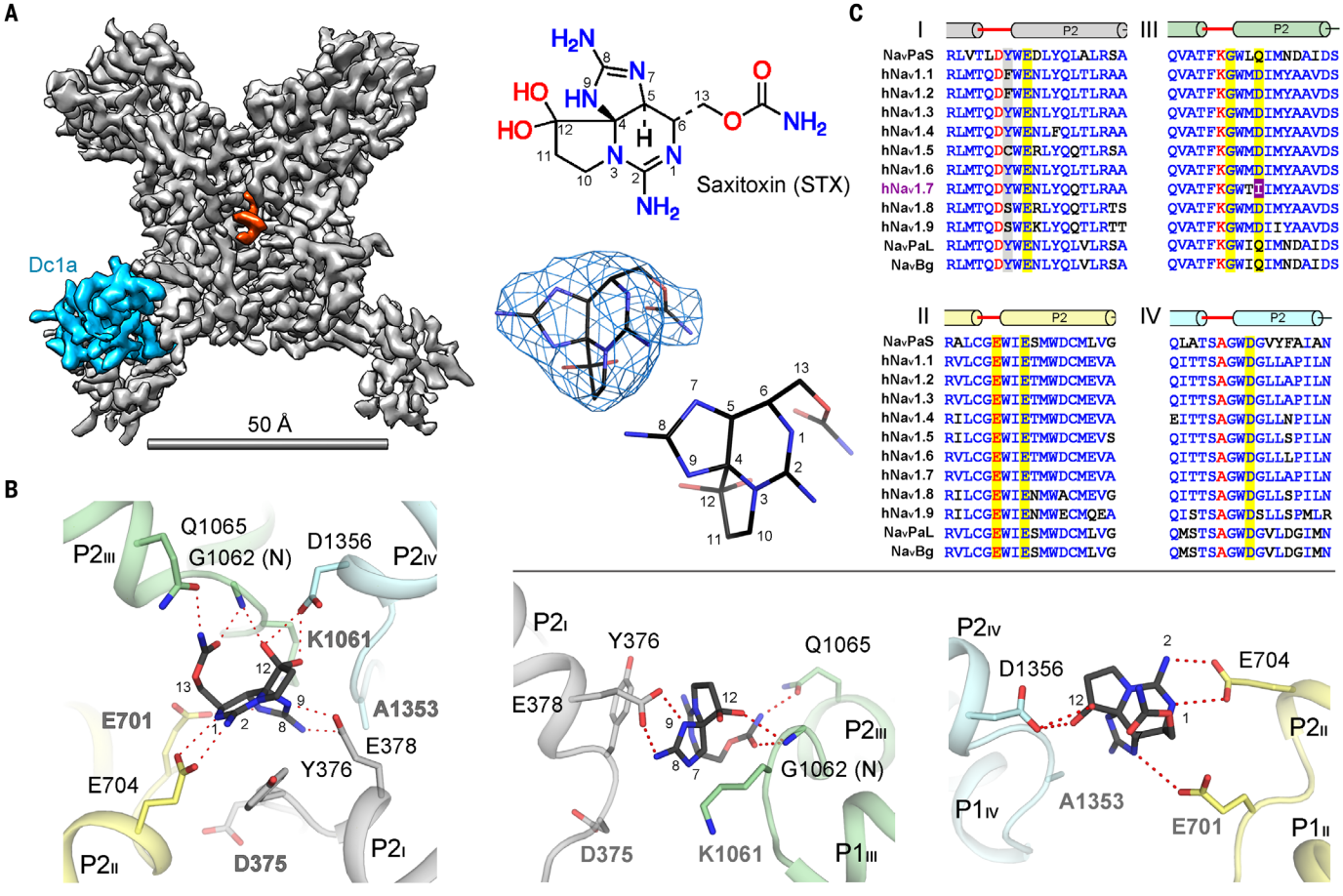


Fig. 4. Specific interactions between Na_vPaS and STX. (**A**) STX is well resolved in the 3.1-Å resolution EM reconstruction of the central domain of Na_vPaS. Left: An extracellular view of the overall EM map. The densities corresponding to STX and Dc1a are colored red and cyan, respectively. Right: Chemical structure (top) and 3D structure (bottom) of the Na_vPaS-bound STX. Middle: Density (blue mesh) for the bound STX at 5 σ. (**B**) STX is specifically coordinated by charged and polar residues from the four repeats at the outer vestibule of the SF. Left: STX tightly blocks the entrance to the SF. An extracellular view is shown. Middle and right: Side views of the coordination of STX by residues from the diagonal repeats. (**C**) Sequence alignment of the SF elements and P2 helices in the four repeats. Similar to Fig. 3D, the residues that participate in STX coordination via polar interactions are shaded yellow. Tyr^{376}, which forms a π-cation interaction with the 1,2,3-guanidinium group of STX, is shaded gray. The locus corresponding to Na_vPaS-Gln^{1065} is Ile in human $Na_v1.7$, which has a much lower affinity for STX than other subtypes of human Na_v channels in which this locus is occupied by Asp.

pixels and local defocus values determined with Gctf. Finally, a local mask was applied during postprocessing to improve the local resolution of the pore domain. 2D classification, 3D classification and auto-refinement were performed with RELION 2.1. The resolution was estimated with the gold-standard Fourier shell correlation 0.143 criterion (*59*) with high resolution noise substitution (*60*).

Model building and structure refinement

Model building was first carried out based on the 2.8-Å reconstruction map of the Na_vPaS-Dc1a complex. The structures of Na_vPaS and Dc1a (PDB accession codes: 5X0M and 2MI5, respectively) were fitted into the EM map by CHIMERA (*61*). Afterwards, the fitted models were manually adjusted in COOT (*62*).

In total, 1380 residues were built with 1272 side chains assigned for the structure. In addition, 7 sugar and 2 lipid moieties were assigned. The intracellular I-II linker, II-III linker, the N-terminal sequence preceding the NTD, and the C-terminal segment following the CTD were not modeled due to the lack of corresponding densities.

The model of Na_vPaS-Dc1a-STX complex was generated using the structure of the Na_vPaS-Dc1a complex as the starting model, which was fitted into the 3.2-Å 3D reconstruction map. 3D conformer of STX (PubChem CID: 37165) was processed with phenix.elbow application in PHENIX (*63*) and the resulted structure can be fitted into the map unambiguously in COOT. The docked models and residues were manually adjusted in COOT.

The model of Na_vPaS-Dc1a-TTX complex was generated using the structure of the Na_vPaS-Dc1a-STX complex as a starting model, which was fitted into the 2.6-Å 3D reconstruction map. The 3D structure of TTX (PubChem CID: 11174599) was used to replace STX. Every residue was manually checked in COOT.

Structure refinement was performed using the phenix.real_space_refine application in PHENIX in real space with secondary structure and geometry restraints to prevent structure over-fitting. Over-fitting of the overall model was monitored by refining the model in one of the two independent maps from the gold-standard refinement approach and testing the refined model against the other map (*64*) (fig. S1D). Statistics of the map reconstruction and model refinement can be found in table S1.

Electrophysiology

Channel mutants were generated using PCR-based mutagenesis with Na_vBg (*65*) as template, then confirmed by DNA sequencing. Fragments from these mutant clones were excised and cloned back into the original Na_vBg containing plasmid to produce final mutant constructs. The DNA sequence of all constructs was confirmed by Sanger sequencing and cRNA synthesized using T7 polymerase (mMessage mMachine kit, Life Technologies, USA) after linearizing the DNA with Not I restriction enzyme. *Xenopus laevis* oocytes were injected with cRNA (0.5–4 ng depending on the channel) encoding wild-type or mutant Na_vBg together with the TipE subunit (*66*) (1:5 molar ratio), then they were incubated at 17°C in ND96 solution (in mM: 96 NaCl, 2 KCl,

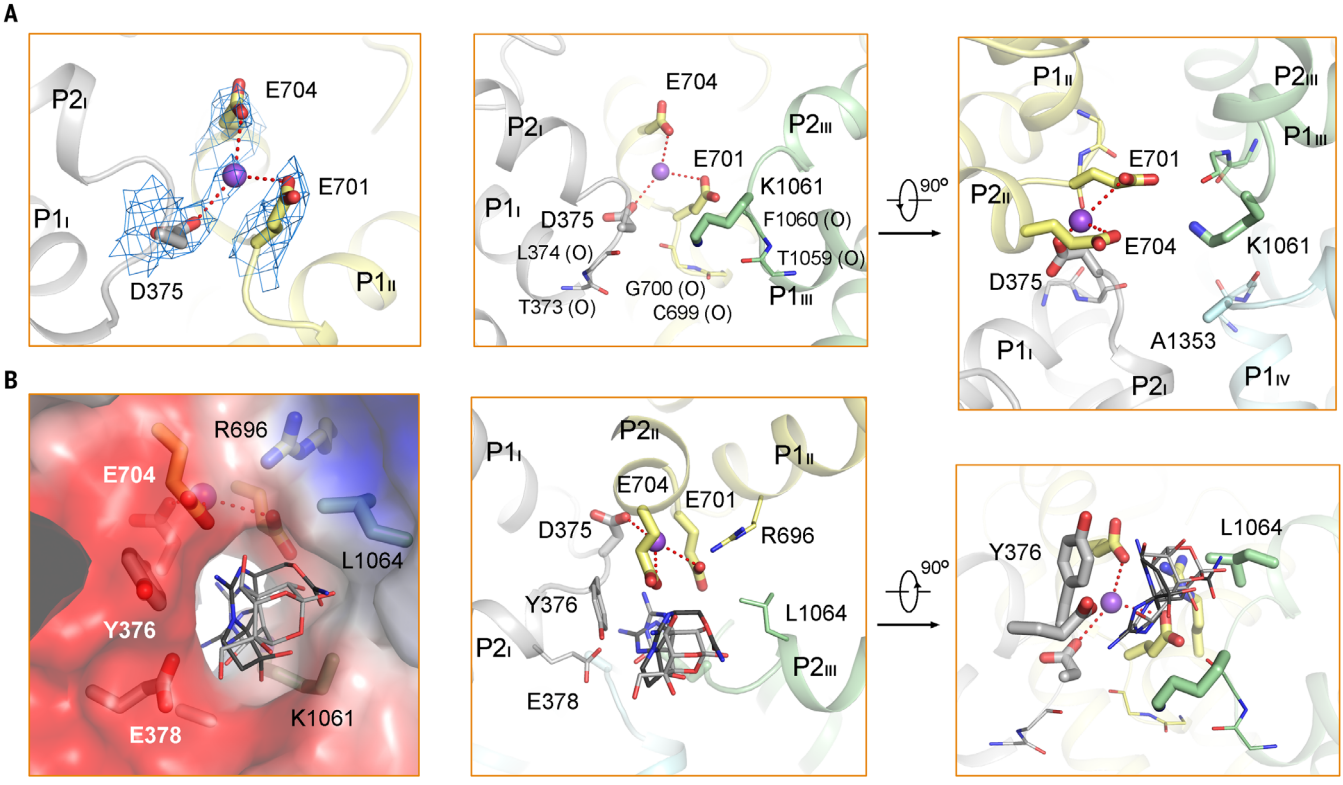


Fig. 5. Molecular mechanism for pore blockade by TTX/STX. (**A**) Potential Na^+ binding site within SF. The carboxylate groups of DE and an invariant Glu on $P2_{II}$ together constitute a potential Na^+ binding site (designated the DEE site). Left: A density surrounded by three carboxylate groups may belong to a bound Na^+ ion in the structure of Na_vPaS-Dc1a, which was purified in the presence of 300 mM NaCl. A similar density is observed in the EM map for Na_vPaS-Dc1a-TTX. Please refer to fig. S4C. Middle and right: Asymmetric coordination of the Na^+ ion within the SF vestibule. Side and top views of the SF vestibule are shown. Repeat IV is omitted in the side view for visual clarity. The carbonyl oxygens that constitute the potential inner site for Na^+ within the SF are shown as thin sticks. The structural observation is consistent with a recent MD simulation analysis of the Na^+ permeation path (*39*). See fig. S10 for details. (**B**) TTX or STX completely blocks access of Na^+ to the DEE site from the extracellular side. Left: The asymmetric chemical composition of the SF outer vestibule determines the permeation path for Na^+. A semitransparent presentation of the electrostatic surface potential of the entrance to the SF viewed from the extracellular side is shown. The residues that give rise to the surface feature are shown in sticks. TTX and STX are shown as silver and black sticks, respectively. Middle and right: Relative position of TTX or STX with respect to the bound Na^+ ion. Placement of TTX or STX at the outer mouth to SF prevents the access of Na^+ to the DEE site from the extracellular side.

1.8 $CaCl_2$, 2 $MgCl_2$ and 5 HEPES; pH 7.6) supplemented with 5 mM pyruvic acid, 50 μg/ml gentamicin sulfate, and 2.5% horse serum. Currents were recorded 1 to 3 days after injections using the two-electrode voltage-clamp technique (Axoclamp 900A, Molecular Devices, USA) with a 30 μl recording chamber. Microelectrodes were filled with 3 M KCl, and resistances were 0.5 to 1 MΩ. All experiments were performed at room temperature (~21°C) in ND96 solution containing 0.1% fatty acid–free bovine serum albumin to prevent adsorption of peptides to plastic. After addition of peptides to the recording chamber, equilibration between peptide and channel was monitored using weak depolarizations elicited at 5 s intervals. For all recordings, voltage-activation relationships were recorded in the absence and presence of peptide. To determine conductance-voltage relationships, oocytes were held at −90 mV and depolarized in 5-mV steps from −90 mV to +30 mV for 50 ms. Data were digitized at 20 kHz; leak and background conductance were identified by blocking channels with TTX and subtracting background currents. Data analyses were performed using Clampfit 10.5 (Molecular Devices, USA) and Prism 7 (GraphPad Software, USA).

REFERENCES AND NOTES

1. B. Hille, *Ion Channels of Excitable Membranes* (Sinauer Associates, ed. 3, 2001), vol. 1, pp. 814.
2. J. Kalia *et al*., From foe to friend: Using animal toxins to investigate ion channel function. *J. Mol. Biol.* **427**, 158–175 (2015). doi: 10.1016/j.jmb.2014.07.027; pmid: 25088688
3. J. R. Deuis *et al*., Pharmacological characterisation of the highly $Na_v1.7$ selective spider venom peptide Pn3a. *Sci. Rep.* **7**, 40883 (2017). doi: 10.1038/srep40883; pmid: 28106092
4. J. D. Osteen *et al*., Selective spider toxins reveal a role for the $Na_v1.1$ channel in mechanical pain. *Nature* **534**, 494–499 (2016). doi: 10.1038/nature17976; pmid: 27281198
5. N. A. Minassian *et al*., Analysis of the structural and molecular basis of voltage-sensitive sodium channel inhibition by the spider toxin Huwentoxin-IV (μ-TRTX-Hh2a). *J. Biol. Chem.* **288**, 22707–22720 (2013). doi: 10.1074/jbc.M113.461392; pmid: 23760503
6. J. K. Murray *et al*., Engineering potent and selective analogues of GpTx-1, a tarantula venom peptide antagonist of the $Na_v1.7$ sodium channel. *J. Med. Chem.* **58**, 2299–2314 (2015). doi: 10.1021/jm501765v; pmid: 25658507
7. S. Cestèle *et al*., Voltage sensor-trapping: Enhanced activation of sodium channels by β-scorpion toxin bound to the S3-S4 loop in domain II. *Neuron* **21**, 919–931 (1998). pmid: 9808476
8. F. Bosmans, M.-F. Martin-Eauclaire, K. J. Swartz, Deconstructing voltage sensor function and pharmacology in sodium channels. *Nature* **456**, 202–208 (2008). doi: 10.1038/nature07473; pmid: 19005548
9. J. Wang *et al*., Mapping the receptor site for α-scorpion toxins on a Na^+ channel voltage sensor. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15426–15431 (2011). doi: 10.1073/pnas.1112320108; pmid: 21876146
10. N. S. Bende *et al*., A distinct sodium channel voltage-sensor locus determines insect selectivity of the spider toxin Dc1a. *Nat. Commun.* **5**, 4350 (2014). doi: 10.1038/ncomms5350; pmid: 25014760
11. C. H. Y. Lau, G. F. King, M. Mobli, Molecular basis of the interaction between gating modifier spider toxins and the voltage sensor of voltage-gated ion channels. *Sci. Rep.* **6**, 34333 (2016). doi: 10.1038/srep34333; pmid: 27677715
12. J. Z. Zhang *et al*., Mapping the interaction site for a β-scorpion toxin in the pore module of domain III of voltage-gated Na^+ channels. *J. Biol. Chem.* **287**, 30719–30728 (2012). doi: 10.1074/jbc.M112.370742; pmid: 22761417
13. H. S. Mosher, F. A. Fuhrman, H. D. Buchwald, H. G. Fischer, Tarichatoxin—Tetrodotoxin: A potent neurotoxin. *Science* **144**, 1100–1110 (1964). doi: 10.1126/science.144.3622.1100; pmid: 14148429
14. C. Y. Kao, Tetrodotoxin, saxitoxin and their significance in the study of excitation phenomena. *Pharmacol. Rev.* **18**, 997–1049 (1966). pmid: 5328391
15. L. M. Durán-Riveroll, A. D. Cembella, Guanidinium toxins and their interactions with voltage-gated sodium ion channels. *Mar. Drugs* **15**, E303 (2017). doi: 10.3390/md15100303; pmid: 29027912
16. J. H. Fleisher, P. J. Killos, C. S. Harrison, Effects of puffer poison on neuromuscular transmission. *J. Pharmacol. Exp. Ther.* **133**, 98–105 (1961). pmid: 13700208
17. T. Narahashi, J. W. Moore, W. R. Scott, Tetrodotoxin blockage of sodium conductance increase in lobster giant axons. *J. Gen. Physiol.* **47**, 965–974 (1964). doi: 10.1085/jgp.47.5.965; pmid: 14155438
18. W. S. Agnew, S. R. Levinson, J. S. Brabson, M. A. Raftery, Purification of the tetrodotoxin-binding component associated with the voltage-sensitive sodium channel from *Electrophorus electricus* electroplax membranes. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2606–2610 (1978). doi: 10.1073/pnas.75.6.2606; pmid: 275831
19. H. Sommer, K. F. Meyer, Mussel poisoning. *Cal. West. Med.* **42**, 423–426 (1935). pmid: 18743279
20. J. F. Lawrence, B. Niedzwiadek, C. Menard, Quantitative determination of paralytic shellfish poisoning toxins in shellfish using prechromatographic oxidation and liquid chromatography with fluorescence detection: Collaborative study. *J. AOAC Int.* **88**, 1714–1732 (2005). pmid: 16526455
21. S. Hagiwara, S. Nakajima, Differences in Na and Ca spikes as examined by application of tetrodotoxin, procaine, and manganese ions. *J. Gen. Physiol.* **49**, 793–806 (1966). doi: 10.1085/jgp.49.4.793; pmid: 5943615
22. J. K. Reed, M. A. Raftery, Properties of the tetrodotoxin binding component in plasma membranes isolated from *Electrophorus electricus*. *Biochemistry* **15**, 944–953 (1976). doi: 10.1021/bi00650a002; pmid: 3213
23. T. Clairfeuille, H. Xu, C. M. Koth, J. Payandeh, Voltage-gated sodium channels viewed through a structural biology lens. *Curr. Opin. Struct. Biol.* **45**, 74–84 (2017). doi: 10.1016/j.sbi.2016.11.022; pmid: 27988421
24. J. Satin *et al*., A mutant of TTX-resistant cardiac sodium channels with TTX-sensitive properties. *Science* **256**, 1202–1205 (1992). doi: 10.1126/science.256.5060.1202; pmid: 1375397
25. N. T. Blair, B. P. Bean, Roles of tetrodotoxin (TTX)-sensitive Na^+ current, TTX-resistant Na^+ current, and Ca^{2+} current in the action potentials of nociceptive sensory neurons. *J. Neurosci.* **22**, 10277–10290 (2002). doi: 10.1523/JNEUROSCI.22-23-10277.2002; pmid: 12451128
26. S. Cestèle, W. A. Catterall, Molecular mechanisms of neurotoxin action on voltage-gated sodium channels. *Biochimie* **82**, 883–892 (2000). doi: 10.1016/S0300-9084(00)01174-3; pmid: 11086218
27. G. M. Lipkind, H. A. Fozzard, A structural model of the tetrodotoxin and saxitoxin binding site of the Na^+ channel. *Biophys. J.* **66**, 1–13 (1994). doi: 10.1016/S0006-3495(94)80746-5; pmid: 8130328
28. B. Hille, The receptor for tetrodotoxin and saxitoxin. A structural hypothesis. *Biophys. J.* **15**, 615–619 (1975). doi: 10.1016/S0006-3495(75)85842-5; pmid: 1148362
29. A. P. Thottumkara, W. H. Parsons, J. Du Bois, Saxitoxin. *Angew. Chem. Int. Ed.* **53**, 5760–5784 (2014). doi: 10.1002/anie.201308235; pmid: 24771635
30. J. Payandeh, T. Scheuer, N. Zheng, W. A. Catterall, The crystal structure of a voltage-gated sodium channel. *Nature* **475**, 353–358 (2011). doi: 10.1038/nature10238; pmid: 21743477
31. X. Zhang *et al*., Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. *Nature* **486**, 130–134 (2012). doi: 10.1038/nature11054; pmid: 22678295
32. E. C. McCusker *et al*., Structure of a bacterial voltage-gated sodium channel pore reveals mechanisms of opening and closing. *Nat. Commun.* **3**, 1102 (2012). doi: 10.1038/ncomms2077; pmid: 23033078
33. D. Ren *et al*., A prokaryotic voltage-gated sodium channel. *Science* **294**, 2372–2375 (2001). doi: 10.1126/science.1065635; pmid: 11743207
34. H. Shen *et al*., Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science* **355**, eaal4326 (2017). doi: 10.1126/science.aal4326; pmid: 28183995
35. D. A. Dougherty, Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163–168 (1996). doi: 10.1126/science.271.5246.163; pmid: 8539615
36. S. H. Heinemann, H. Terlau, K. Imoto, Molecular basis for pharmacological differences between brain and cardiac sodium channels. *Pflugers Arch.* **422**, 90–92 (1992). doi: 10.1007/BF00381519; pmid: 1331981
37. P. H. Backx, D. T. Yue, J. H. Lawrence, E. Marban, G. F. Tomaselli, Molecular localization of an ion-binding site within the pore of mammalian sodium channels. *Science* **257**, 248–251 (1992). doi: 10.1126/science.1321496; pmid: 1321496
38. J. R. Walker *et al*., Marked difference in saxitoxin and tetrodotoxin affinity for the human nociceptive voltage-gated sodium channel ($Na_v1.7$). *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18102–18107 (2012). doi: 10.1073/pnas.1206952109; pmid: 23077250
39. J. Zhang *et al*., Simulating the ion permeation and ion selection for a eukaryotic voltage-gated sodium channel Na_vPaS. *Protein Cell* **9**, 580–585 (2018). doi: 10.1007/s13238-018-0522-y; pmid: 29532417
40. M. Noda, H. Suzuki, S. Numa, W. Stühmer, A single point mutation confers tetrodotoxin and saxitoxin insensitivity on the sodium channel II. *FEBS Lett.* **259**, 213–216 (1989). doi: 10.1016/0014-5793(89)81531-5; pmid: 2557243
41. H. Terlau *et al*., Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. *FEBS Lett.* **293**, 93–96 (1991). doi: 10.1016/0014-5793(91)81159-6; pmid: 1660007
42. K. J. Kontis, A. L. Goldin, Site-directed mutagenesis of the putative pore region of the rat IIA sodium channel. *Mol. Pharmacol.* **43**, 635–644 (1993). pmid: 8386312
43. H. Nakayama *et al*., Photolabeled sites with a tetrodotoxin derivative in the domain III and IV of the electroplax sodium channel. *Biochem. Biophys. Res. Commun.* **184**, 900–907 (1992). doi: 10.1016/0006-291X(92)90676-C; pmid: 1315538
44. J. L. Penzotti, H. A. Fozzard, G. M. Lipkind, S. C. Dudley Jr., Differences in saxitoxin and tetrodotoxin binding revealed by mutagenesis of the Na^+ channel outer vestibule. *Biophys. J.* **75**, 2647–2657 (1998). doi: 10.1016/S0006-3495(98)77710-0; pmid: 9826589
45. J. L. Penzotti, G. Lipkind, H. A. Fozzard, S. C. Dudley Jr., Specific neosaxitoxin interactions with the Na^+ channel outer vestibule determined by mutant cycle analysis. *Biophys. J.* **80**, 698–706 (2001). doi: 10.1016/S0006-3495(01)76049-3; pmid: 11159437
46. G. Choudhary, L. Shang, X. Li, S. C. Dudley Jr., Energetic localization of saxitoxin in its channel binding site. *Biophys. J.* **83**, 912–919 (2002). doi: 10.1016/S0006-3495(02)75217-X; pmid: 12124273
47. G. Choudhary, M. Yotsu-Yamashita, L. Shang, T. Yasumoto, S. C. Dudley Jr., Interactions of the C-11 hydroxyl of tetrodotoxin with the sodium channel outer vestibule. *Biophys. J.* **84**, 287–294 (2003). doi: 10.1016/S0006-3495(03)74849-8; pmid: 12524282
48. E. Carbonneau, K. Vijayaragavan, M. Chahine, A tryptophan residue (W736) in the amino-terminus of the P-segment of domain II is involved in pore formation in $Na_v1.4$ voltage-gated sodium channels. *Pflugers Arch.* **445**, 18–24 (2002). doi: 10.1007/s00424-002-0887-9; pmid: 12397382
49. Z. Yan *et al*., Structure of the $Na_v1.4$-β1 complex from electric eel. *Cell* **170**, 470–482.e11 (2017). doi: 10.1016/j.cell.2017.06.039; pmid: 28735751
50. J. Gilchrist, S. Das, F. Van Petegem, F. Bosmans, Crystallographic insights into sodium-channel modulation by the β4 subunit. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E5016–E5024 (2013). doi: 10.1073/pnas.1314557110; pmid: 24297919
51. J. Lei, J. Frank, Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005). doi: 10.1016/j.jsb.2005.01.002; pmid: 15797731
52. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
53. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015). doi: 10.7554/eLife.06980; pmid: 26023829
54. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
55. S. H. W. Scheres, A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012). doi: 10.1016/j.jmb.2011.11.010; pmid: 22100448
56. S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi: 10.1016/j.jsb.2012.09.006; pmid: 23000701
57. S. H. W. Scheres, Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015). doi: 10.1016/j.jsb.2014.11.010; pmid: 25486611

58. D. Kimanius, B. O. Forsberg, S. H. W. Scheres, E. Lindahl, Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016). doi: 10.7554/eLife.18722; pmid: 27845625
59. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003). doi: 10.1016/j.jmb.2003.07.013; pmid: 14568533
60. S. Chen *et al*., High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013). doi: 10.1016/j.ultramic.2013.06.004; pmid: 23872039
61. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
62. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
63. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
64. A. Amunts *et al*., Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014). doi: 10.1126/science.1249410; pmid: 24675956
65. K. Dong, A single amino acid change in the para sodium channel protein is associated with knockdown-resistance (*kdr*) to pyrethroid insecticides in German cockroach. *Insect Biochem. Mol. Biol.* **27**, 93–100 (1997). doi: 10.1016/S0965-1748(96)00082-3; pmid: 9066120
66. G. Feng, P. Deák, M. Chopra, L. M. Hall, Cloning and functional analysis of TipE, a novel membrane protein that enhances Drosophila *para* sodium channel function. *Cell* **82**, 1001–1011 (1995). doi: 10.1016/0092-8674(95)90279-1; pmid: 7553842
67. W. L. DeLano, The PyMOL Molecular Graphics System (2002); www.pymol.org.

ACKNOWLEDGMENTS

We thank X. Li (Tsinghua University) for technical support during EM image acquisition. We thank K. Dong (Michigan State University) for sharing $BgNa_v1$ (Na_vBg) and TipE constructs. **Funding:** This work was funded by the National Key Basic Research (973) Program (2015CB910101 to N.Y.) and the National Key R&D Program (2016YFA0500402 to N.Y. and 2016YFA0501100 to J.L.) from Ministry of Science and Technology of China, the National Natural Science Foundation of China (projects 31621092, 31630017, and 31611130036 to N.Y.), the Australian Research Council (DP160104411), and the Australian National Health and Medical Research Council (Principal Research Fellowship and Program Grant APP1072113 to G.F.K.). We thank the Tsinghua University Branch of China National Center for Protein Sciences (Beijing) for providing the cryo-EM facility support. We thank the computational facility support on the cluster of Bio-Computing Platform (Tsinghua University Branch of China National Center for Protein Sciences Beijing) and the "Explorer 100" cluster system of Tsinghua National Laboratory for Information Science and Technology. N.Y. is supported by the Shirley M. Tilghman endowed professorship from Princeton University. **Ethics statement:** This study was carried out in accordance with the recommendations in the Australian code of practice for the care and use of animals for scientific purposes (8th ed., 2013). The protocol for *Xenopus laevis* studies was approved by the Animal Ethics Committee of The University of Queensland (approval number QBI/059/13/ARC/NHMRC). Surgeries for harvesting *X. laevis* oocytes with recovery endpoint were performed under anesthesia [animals exposed to tank filled with MS-222 (1.3 mg/ml)], with a minimum of 3 months between surgeries. **Author contributions:** N.Y., G.F.K., and Q.Z. conceived the project. H.S., Q.Z., Z.L., Y.J., X.P., J.W., B.C.-A., J.J.S., Y.K.Y.C., and J.L. performed the experiments. All authors contributed to data analysis. N.Y., G.F.K., Q.Z., and H.S. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The atomic coordinates for Na_vPaS-Dc1a complex, Na_vPaS-Dc1a-TTX complex, and Na_vPaS-Dc1a-STX complex have been deposited in the Protein Data Bank (www.rcsb.org) with accession codes 6A90, 6A95, and 6A91, respectively. The EM maps have been deposited in EMDB (www.ebi.ac.uk/pdbe/emdb/) with accession codes EMD-6995, EMD-6997, and EMD-6996.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/eaau2596/suppl/DC1
Figs. S1 to S10.
Table S1
References

23 May 2018; accepted 17 July 2018
10.1126/science.aau2596

RESEARCH ARTICLE SUMMARY

SOCIAL SCIENCE

Relationship of gender differences in preferences to economic development and gender equality

Armin Falk* **and Johannes Hermle***

INTRODUCTION: Understanding determinants of gender differences in economic and social domains has been of interest, both in academic and public debates. Previous research has shown that gender differences in fundamental economic preferences are important in explaining gender differences in economic outcomes, such as for occupational choice, financial investment, or educational decisions, among many others. However, gaps remain in understanding the sources of gender differences in preferences and their variation.

RATIONALE: We contrasted and tested two hypotheses that make opposite predictions concerning the cross-country association of gender differences in preferences with economic development and gender equality. On one hand, the attenuation of gender-specific social roles that arises in more developed and gender-egalitarian countries may alleviate differences in preferences between women and men. As a consequence, one would expect gender differences in preferences to be negatively associated with higher levels of economic development and gender equality (social role hypothesis). On the other hand, greater availability of material and social resources removes the gender-neutral goal of subsistence, which creates the scope for gender-specific ambitions and desires. In addition, more gender-equal access to those resources may allow women and men to express preferences independently from each other. As a consequence, one would expect gender differences in preferences to be positively associated with higher levels of economic development and gender equality (resource hypothesis).

We tested these competing predictions using data on experimentally validated measures of willingness to take risks, patience, altruism, positive and negative reciprocity, and trust for 80,000 individuals in 76 representative country samples. So that the data would be geographically representative, the dataset was chosen so as to include all continents and a broad range of cultures and economic development levels. In total, the data represent about 90% of both the world population and global income.

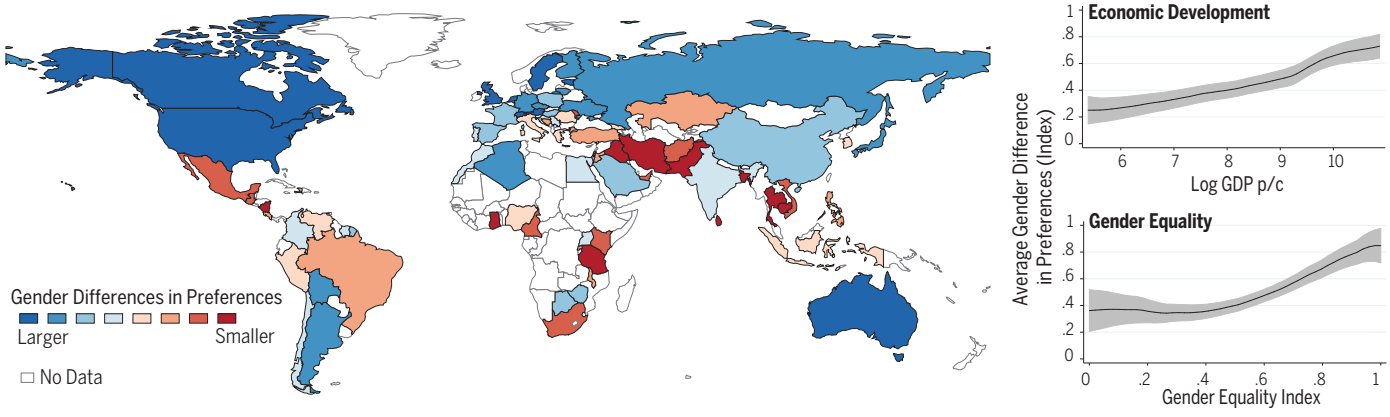
RESULTS: The data revealed substantial cross-country variation in gender differences in preferences. Gender differences were found to be strongly positively associated with economic development as well as gender equality. These relationships held for each preference separately as well as for a summary index of differences in all preferences jointly. Quantitatively, this summary index exhibited correlations of 0.67 ($P < 0.0001$) with log GDP per capita and 0.56 ($P < 0.0001$) with a Gender Equality Index (a joint measure of four indices of gender equality), respectively. To isolate the separate impacts of economic development and gender equality, we conducted a conditional analysis, finding a quantitatively large and statistically significant association between gender differences and log GDP per capita conditional on the Gender Equality Index, and vice versa. These findings remained robust in several validation tests, such as accounting for potential culture-specific survey response behavior, aggregation bias, and nonlinear relationships.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aas9899

CONCLUSION: The reported evidence indicates that higher levels of economic development and gender equality favor the manifestation of gender differences in preferences across countries. Our results highlight the critical role of availability of material and social resources, as well as gender-equal access to these resources, in facilitating the independent formation and expression of gender-specific preferences. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: armin.falk@briq-institute.org (A.F.); j.hermle@berkeley.edu (J.H.)
Cite this article as A. Falk, J. Hermle, *Science* **362**, eaas9899 (2018). DOI: 10.1126/science.aas9899



Gender differences in preferences across countries and their association with economic development and gender equality. (Left) World map visualizing a summary index of gender differences in all six preferences (risk-taking, patience, altruism, trust, and positive and negative reciprocity). (Right) The relationship between the summary index of gender differences in preferences and (top) log GDP per capita and (bottom) a Gender Equality Index comprising measures of material, social, and political gender equality. The relationships are predicted from local polynomial regressions. Shaded areas indicate 95% confidence intervals.

RESEARCH ARTICLE

SOCIAL SCIENCE

Relationship of gender differences in preferences to economic development and gender equality

Armin Falk[1]* and Johannes Hermle[2]*

Preferences concerning time, risk, and social interactions systematically shape human behavior and contribute to differential economic and social outcomes between women and men. We present a global investigation of gender differences in six fundamental preferences. Our data consist of measures of willingness to take risks, patience, altruism, positive and negative reciprocity, and trust for 80,000 individuals in 76 representative country samples. Gender differences in preferences were positively related to economic development and gender equality. This finding suggests that greater availability of and gender-equal access to material and social resources favor the manifestation of gender-differentiated preferences across countries.

Fundamental preferences such as altruism, risk-taking, reciprocity, patience, or trust constitute the foundation of choice theories and govern human behavior. A growing literature in economics (*1*, *2*) and psychology (*3*) documents important gender differences in preferences. These differences provide a key explanation for differential choices and outcomes between women and men in contexts such as occupational choice, financial investment, or educational decisions (*4*, *5*), among many others. The origins of gender differences in preferences, and their variability across countries and cultures, are addressed by an extensive literature that discusses biological and evolutionary determinants (*6*, *7*) and the role of the social environment (*8*–*10*).

Hypotheses

We contrast two competing hypotheses that make opposite predictions concerning the cross-country correlational patterns of gender differences in preferences with respect to economic development and gender equality. Following social role theory, one may hypothesize that gender differences in preferences attenuate in more developed, gender-egalitarian countries (social role hypothesis). This hypothesis rests on two premises. First, economic development is a key determinant of societal progression toward gender equality (*11*, *12*), which is critical for the dissolution of traditional gender roles (*13*, *14*). Second, as discussed by a large body of literature (*8*–*10*), gender-specific roles instill distinct preferences in women and men and hence constitute a crucial component in explaining the gender preference gap. As a consequence, according to the social role hypothesis, higher economic development and gender equality (and the associated dissolution of traditional gender roles) should lead to a narrowing of gender differences in preferences.

In contrast, there is reason to expect that gender differences in preferences expand with economic development and gender equality (resource hypothesis). As suggested by post-materialist theory (*15*, *16*), a critical societal precondition for self-expression is the fulfillment of basic material needs. In line with this, existing research shows that the unrestricted expression of preferences hinges on the availability of sufficient material and social resources (*17*–*20*). Therefore, gender differences in preferences should manifest themselves only if both women and men obtain sufficient access to these resources to independently develop and express their intrinsic preferences (*21*). Specifically, greater availability of material resources eliminates the gender-neutral goal of subsistence. This creates scope for attending to gender-specific ambitions and desires. As a consequence, economic development may facilitate the unfolding of differences between women and men. More developed countries also feature higher levels of gender equality in political, social, and economic domains (*11*), which is a critical requirement for the acceptance of gender-specific desires and preferences. In particular, as women become less exposed and vulnerable to male influence, gender differentiation may be reinforced through women's greater opportunities for self-expression. In sum, greater availability of material and social resources to both women and men may facilitate the independent development and expression of gender-specific preferences, and hence may lead to an expansion of gender differences in more developed and gender-egalitarian countries.

Data and measures

An empirical test of the two competing hypotheses requires data that meet three critical conditions: (i) reliability of preference measures, (ii) extensive cultural variation as well as comprehensive global coverage, and (iii) representativeness of country samples. Our investigation used the Global Preference Survey (GPS) (*22*, *23*). The GPS was collected as part of the Gallup World Poll 2012 and contains measures of six fundamental preferences with regard to social and nonsocial domains: willingness to take risks; patience, which captures preferences over the intertemporal timing of rewards; altruism; trust (*24*); and positive and negative reciprocity, which capture the costly willingness to reward kind actions or to punish unkind actions, respectively.

Before the launch of the international survey, multiple survey items were selected for these preferences through an ex ante experimental validation (*25*). For each preference, subjects responded to a large set of survey items and participated in incentivized choice experiments. The subset of survey items that maximized adjusted R^2 in predicting incentivized behavior in the corresponding experiment was selected for the international survey. The selected items, described below, comprise a combination of qualitative self-assessments and quantitative items that involve economic trade-off decisions. The qualitative items elicit participants' subjective assessment of their willingness to act in a certain way, such as whether participants are generally willing to take risks. Complementarily, the quantitative items provide revealed preference measures by using participants' choices in monetary trade-off decisions. As an example, the quantitative item for risk-taking provides the participants with a sequence of five interdependent choices between a fixed and a risky payment (lottery). This allows one to progressively approach the point of indifference between the fixed payment and the lottery, which serves as a revealed preference measure for risk-taking behavior. The presence of both qualitative and quantitative items allows for robustness tests with respect to potential culture-specific response behavior. So that survey items were comparable across cultures, all items were translated back and forth by professionals, and monetary values mentioned in the survey questions were adjusted along median household income across countries. The survey items were pretested in 22 countries of varied cultural heritage as part of the Gallup World Poll 2012 pretest, conducted in late 2011, to guarantee cross-cultural validity.

After the ex ante experimental validation and pretests, the international survey was implemented in a total of 76 countries, representing about 90% of the global population and global GDP. To provide geographic representativeness as well as developmental and cultural variation, we selected the countries to include all continents and a very broad range of economic development levels. For each country, the data contain samples representative of the resident population aged 15 and older, with a median sample size

[1]Department of Economics, University of Bonn, and Behavior and Inequality Research Institute (briq), 53113 Bonn, Germany. [2]Department of Economics, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: armin.falk@briq-institute.org (A.F.); j.hermle@berkeley.edu (J.H.)

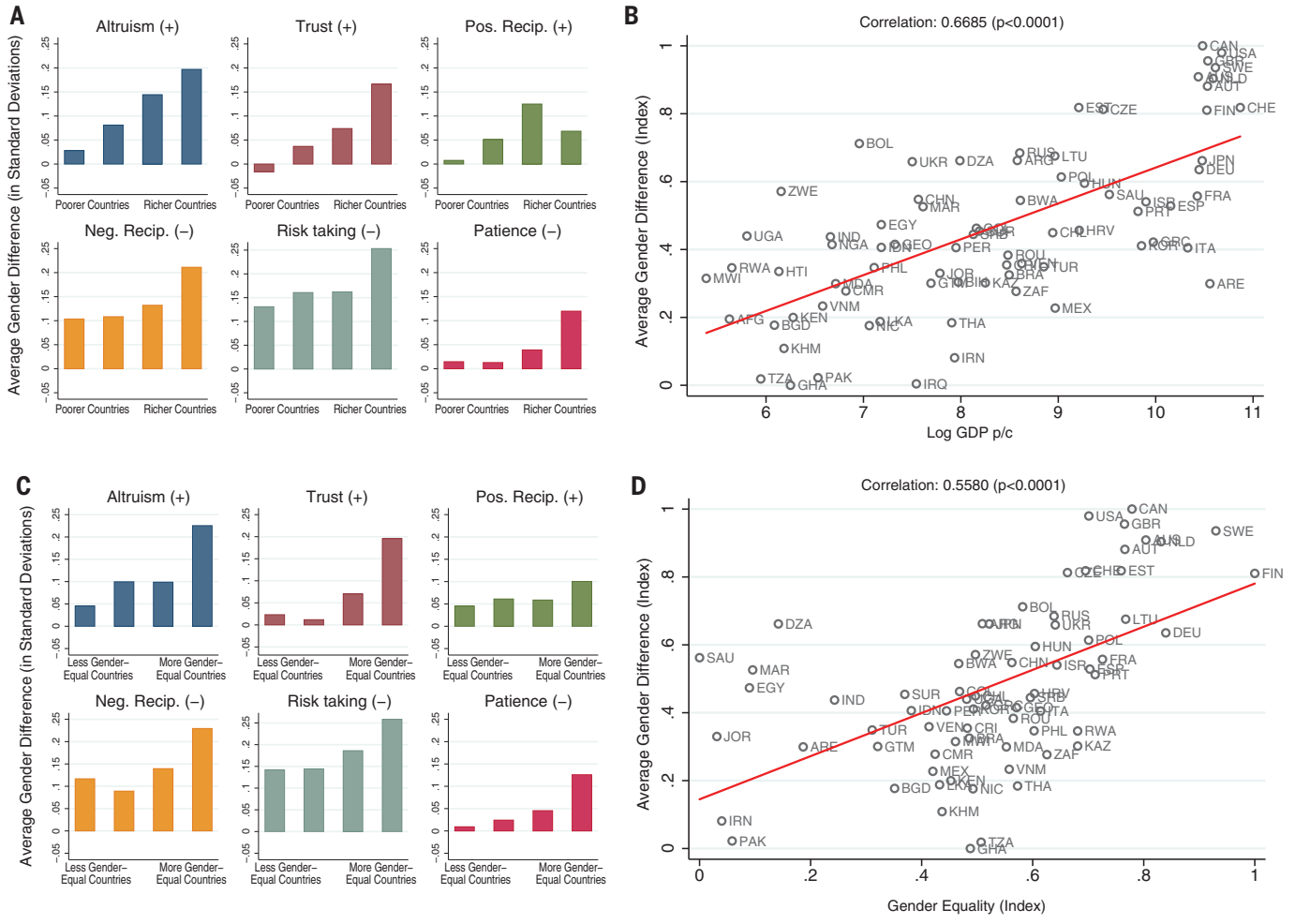


Fig. 1. Analysis of gender differences in preferences in relation to economic development and gender equality. (**A**) Mean country-level gender difference in altruism, trust, positive reciprocity, negative reciprocity, risk-taking, and patience by development level. Countries were sorted into four bins according to their GDP per capita quartile. The symbols + and – in the panel titles indicate the sign of the difference for each preference; + indicates that positive differences correspond to women exhibiting higher levels of the respective preference, whereas – indicates that positive differences correspond to women exhibiting lower levels of the respective preference. (**B**) Relationship between the aggregate index of gender differences in all six preferences and log GDP per capita. (**C** and **D**) Same relationships as in (A) and (B) for the Gender Equality Index. See supplementary materials for country abbreviation key.

of 1000 participants per country; this made generalizable inferences possible. In total, the data include preference measures for about 80,000 participants.

After implementation of the worldwide survey, the measures for the six preferences were generated according to the following procedure. First, each of the survey items was standardized using the mean and variance of the entire worldwide sample. Then, to obtain the preference measures, the relevant z-scores were averaged using weights developed in the experimental validation. Further details on the data collection and construction of our measures are given below and in the supplementary materials.

The data allow assessment of the existence and quantitative relevance of gender differences in preferences at the global level (*22*). For this purpose, global gender differences were calculated as follows: Each preference measure was standardized at the global level to exhibit a mean of 0 and a standard deviation of 1. Then, for each preference, an ordinary least-squares (OLS) regression was performed on the worldwide sample, using as the independent variable a gender indicator in which male is the reference category, controlling for age, age squared, cognitive skills (as proxied by subjective math skills), education level, household income quintile, and country fixed effects. Standard errors were clustered at the country level. The estimated coefficient on the gender indicator served as the gender difference in the respective preference. On the global level, all six preferences featured significant gender differences (fig. S1): Women tended to be more prosocial and less negatively reciprocal than men, with differences in standard deviations of 0.106 for altruism ($P < 0.0001$), 0.064 for trust ($P < 0.0001$), 0.055 for positive reciprocity ($P < 0.0001$), and 0.129 for negative reciprocity ($P < 0.0001$). Turning to nonsocial preferences, women were less risk-taking by 0.168 standard deviations ($P < 0.0001$) and less patient by 0.050 standard deviations ($P < 0.0001$) (*26*). The observed differences in preferences set the stage for our analysis.

Analysis of gender differences in preferences in relation to economic development and gender equality

To test the competing hypotheses, we computed country-level gender differences for each preference. For this purpose, we standardized each preference measure at the country level to exhibit a mean of 0 and a standard deviation of 1. We then performed for each preference and country a separate OLS regression using as independent variable a gender indicator in which male is the reference category. We also included several controls to isolate the gender effect from potentially confounding factors that might differ between women and men. These controls are age, age squared, subjective math skills, education level, and household income quintile. The obtained

coefficient on the gender indicator served as a measure of the gender difference in the respective preference and country.

Using the country-level estimates of gender differences in preferences, we examined variation along levels of economic development and gender equality. As the measure of economic development, we used GDP per capita. To assess the role of gender equality, we created a Gender Equality Index as a joint measure of four indices of gender equality: (i) the Global Gender Gap Index of the World Economic Forum (WEF), (ii) the Gender Equality Index of the United Nations (UN), (iii) the ratio of female to male labor force participation rates, and (iv) the number of years since women's suffrage. The Gender Equality Index was constructed as the predicted main component from a principal components analysis of the four indices.

To study the effect of economic development, we first sorted the 76 countries into four bins according to their level of development, measured by GDP per capita. We then computed for each bin the average country-level gender difference in each preference. Gender differences in all six preferences increased with a country's level of development (Fig. 1A). The positive correlations between log GDP per capita and country-level gender differences were large and statistically significant for all six preferences: 0.58 for altruism ($P < 0.0001$), 0.59 for trust ($P < 0.0001$), 0.31 for positive reciprocity ($P = 0.0067$), 0.35 for negative reciprocity ($P = 0.0017$), 0.37 for risk-taking ($P = 0.0011$), and 0.38 for patience ($P = 0.0006$) (fig. S2). We also analyzed a summary index of gender differences for all preferences jointly. For this purpose, we first performed a principal components analysis of the country-level gender differences in the six preferences. We then created an index of gender differences in preferences as the predicted main component. This index exhibited a correlation of 0.67 ($P < 0.0001$) with log GDP per capita (Fig. 1B) (*27*).

To study the effect of gender equality, we ran the same analysis as for economic development, using the Gender Equality Index as the explanatory variable. Gender differences in preferences were found to increase with gender equality for each preference separately (Fig. 1C) as well as for the index of gender differences in preferences (Fig. 1D). For the individual preferences, the correlation coefficients were 0.51 for altruism ($P < 0.0001$), 0.41 for trust ($P = 0.0005$), 0.13 for positive reciprocity ($P = 0.2875$), 0.40 for negative reciprocity ($P = 0.0005$), 0.34 for risk-taking ($P = 0.0036$), and 0.43 for patience ($P = 0.0002$) (fig. S3). The summary index of gender differences in preferences exhibited a correlation of 0.56 ($P < 0.0001$) with the Gender Equality Index. Reassuringly, the positive relationship between the index of gender differences in preferences and gender equality was also found for the four individual indicators of gender equality (fig. S4).

Economic development and gender equality are strongly intertwined (*11*). To isolate the separate impacts of economic development and gender equality on gender differences in preferences, we conducted a conditional analysis. We constructed partial regression plots illustrating the relationship between the index of gender differences in preferences and log GDP per capita conditional on the Gender Equality Index (Fig. 2A) and vice versa (Fig. 2B). The dependent and independent variables were standardized to exhibit a mean of 0 and a standard deviation of 1. Hence, the slope coefficients can be interpreted as the standard deviation change in the dependent variable in response to a change of one standard deviation in the independent variable.

There was a quantitatively large and statistically significant association of gender differences with log GDP per capita conditional on the Gender Equality Index. The estimated slope coefficient was 0.53 ($P < 0.0001$). Likewise, gender differences were strongly associated with the Gender Equality Index conditional on log GDP per capita, with a somewhat smaller slope coefficient of 0.32 ($P = 0.0033$) (see also table S4, column 7). When conducting an F-test for equality of both coefficients, we failed to reject at $P = 0.2537$, indicating that the strength of the relationships between the index of gender differences in preferences and log GDP per capita and the Gender Equality Index were not statistically different. These findings imply that both economic development and gender equality exhibited an independent and significant association with gender differences in preferences (*28*). Conditional on log GDP per capita, differences in preferences were also significantly and positively associated with the four individual measures of gender equality (Fig. 2, C to F). Slope coefficients were 0.23 ($P = 0.0084$) for the WEF Global Gender Gap Index, 0.29 ($P = 0.0515$) for the UN Gender Equality Index, 0.25 ($P = 0.0123$) for ratio of female to male labor force participation, and 0.30 ($P = 0.0023$) for years since women's suffrage.

In sum, these findings provide evidence in favor of the resource hypothesis that higher levels of economic development and gender quality are associated with stronger gender differentiation in preferences.

A potential concern regarding the reported results involves bias due to culture-specific survey response behavior (*29*–*32*). Note that our data contain two types of items: qualitative self-assessments and quantitative choice measures. Qualitative self-assessments might be affected by response biases such as scaling effects, which might vary across cultures, thereby introducing systematic measurement error (*33*). In contrast, the quantitative items present trade-offs that are well-defined in terms of stakes and probabilities, yielding revealed preference measures that facilitate a culturally fair comparison. To test for robustness with regard to the elicitation method, we constructed two separate indices of gender differences using either qualitative or quantitative items only (analogous to how we constructed the main index). The correlations of the indices with log GDP per capita were found to be very similar, with values of 0.551 ($P < 0.0001$) for qualitative items and 0.516 ($P < 0.0001$) for quantitative items (fig. S7, A and B). A test of the null hypothesis of equality of the correlation coefficients failed to reject at conventional significance levels ($P = 0.744$). Likewise, correlations with the Gender Equality Index were 0.480 ($P < 0.0001$) for qualitative items and 0.479 ($P < 0.0001$) for quantitative items (fig. S7, C and D). Testing equality of the coefficients failed to reject ($P = 0.991$), thus providing no support that culture-specific response behavior contaminated the results.

To further test for the robustness of our results, we conducted several additional analyses. First, because trust reflects a composite trait that captures beliefs about others' behavior, prosocial preferences, and preferences for risk-taking, we repeated our analysis excluding the trust dimension. To do so, we constructed an alternative index of gender differences in preferences in a procedure parallel to the main index but using only the five remaining preferences. Similar to our main results, this alternative index exhibited a quantitatively large association with economic development and measures of gender equality (tables S5 and S6). Second, we tested whether the level of standardization affected our results. We repeated our analysis using preference measures standardized at the global rather than the country level. The results using preferences standardized at the global level were similar to our main results (fig. S8 and tables S7 and S8). Third, we repeated our analysis without using individual-level controls when calculating gender differences, yielding similar results (fig. S9 and tables S9 and S10). Fourth, a common concern in cross-country analysis involves measurement error. Because the experimental validation was conducted in Germany, more linguistically similar countries might exhibit smaller measurement error. To test for robustness against this potential confound, we additionally controlled for linguistic distance to German, which left the results qualitatively unchanged (tables S11 and S12). Fifth, to address concerns of aggregation bias, we tested for the relationship between household income and gender differences in preferences in individual-level regressions, finding a significant relationship for each preference (table S13). Finally, we tested for a nonlinear relationship with economic development. A closer inspection of Fig. 1B suggested a nonlinear, convex relationship, which is confirmed by regression analysis (table S14, column 2). This pattern originated from the fact that richer countries are overproportionally more gender-equal. Therefore, when we investigated the relationship between the index of gender differences in preferences and log GDP per capita after residualizing both variables with respect to the Gender Equality Index, the relationship was found to be linear (table S14). See supplementary text for details of the robustness tests.

Concluding remarks

The reported evidence indicates that higher levels of economic development and gender equality are associated with stronger gender differentiation

in preferences. These findings may also relate to other personality traits, such as the Big Five (*34*, *35*) or value priorities (*36*). Our findings do not rule out an influence of gender-specific roles that drive gender differences in preferences. They also do not preclude a role for biological or evolutionary determinants of gender differences (*37*). Our results highlight, however, that theories not attributing a significant role to the social environment are incomplete (*38*).

In this regard, our findings point toward the critical role of availability of and equal access to material and social resources for both women and men in facilitating the independent formation and expression of gender-specific preferences

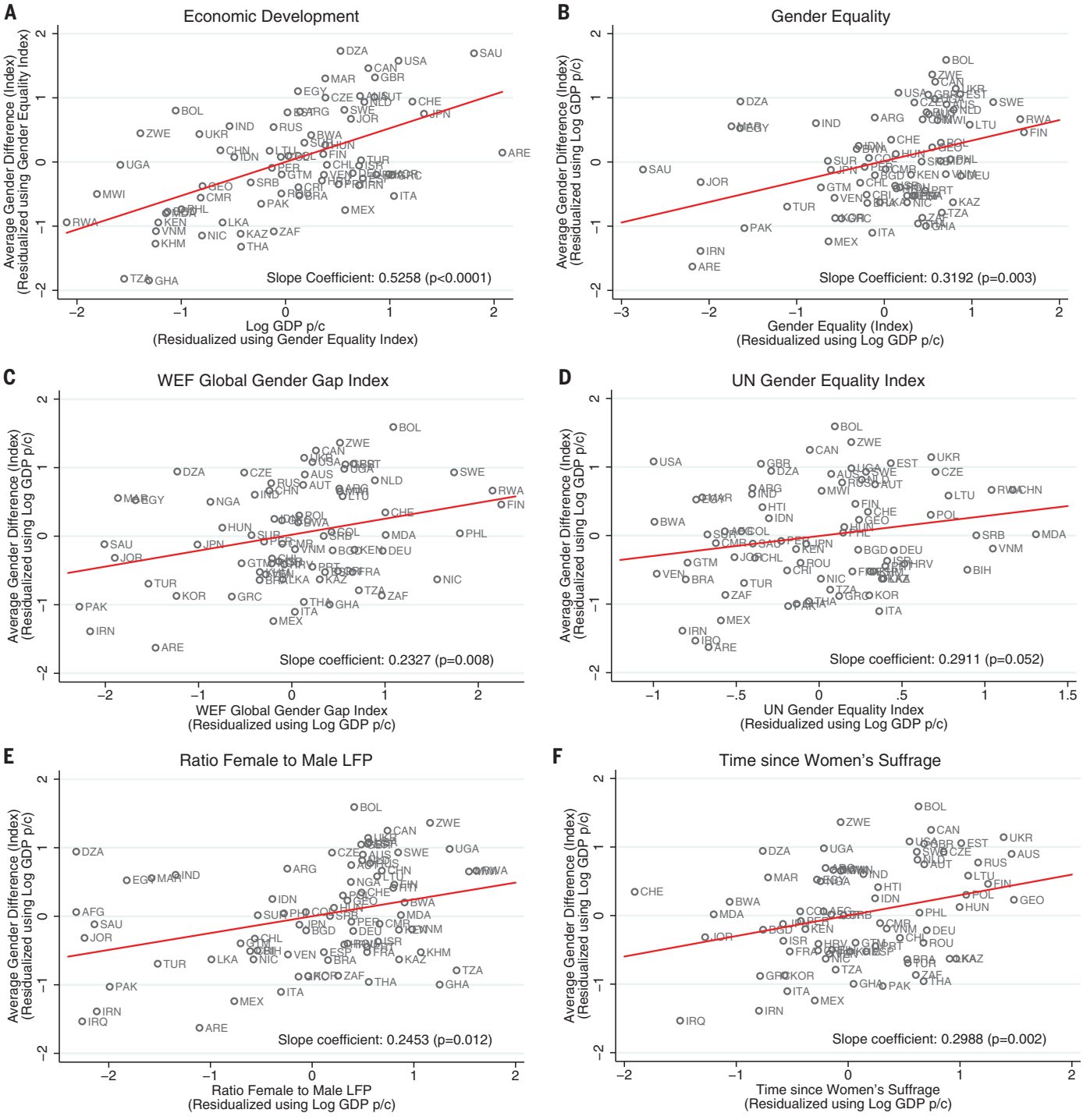


Fig. 2. Analysis of gender differences in preferences in relation to economic development conditional on gender equality, and vice versa. Each panel depicts a partial regression plot. (**A**) Relationship between the aggregate index of gender differences in preferences and log GDP per capita after residualizing both variables with respect to the Gender Equality Index. (**B** to **F**) Relationship between the aggregate index of gender differences in preferences and five indices of gender equality after residualizing all variables with respect to log GDP per capita. Indices of gender equality are (B) the Gender Equality Index, (C) WEF Global Gender Gap Index, (D) UN Gender Equality Index, (E) ratio of female to male labor force participation, and (F) years since women's suffrage. For corresponding regression evidence, see table S4.

across countries. As suggested by the resource hypothesis, greater availability of material resources removes the human need of subsistence, and hence provides the scope for attending to gender-specific preferences. A more egalitarian distribution of material and social resources enables women and men to independently express gender-specific preferences.

Materials and methods

See the supplementary materials for further details, including a list of the 76 countries included in the survey.

Experimental selection of survey items and construction of preference measures

Survey items included in the GPS data were selected in an ex ante experimental validation procedure at the Laboratory for Experimental Economics of the University of Bonn in winter 2010–2011. In this procedure, 402 subjects participated in incentivized-choice experiments and responded to a large set of survey items, which were either newly developed or taken from existing surveys (*25*).

Incentivized-choice experiments were conducted to obtain an incentivized behavioral measure for each preference. Risk-taking was measured as the average response to two multiple price lists in which subjects chose between a lottery and varying safe options. Patience was measured as the average response to two multiple price lists in which subjects chose between receiving a payment on the day of the experiment or a larger payment 12 months later. Trust was measured as the average amount sent as a first mover in two investment games. Altruism was measured as first-mover behavior in a dictator game with a charitable organization as recipient. Positive reciprocity was measured as the average amount sent back as a second mover in two investment games. Negative reciprocity was measured as the average amount invested into punishment after unilateral defection of the opponent in a prisoner's dilemma and the minimum acceptable offer in an ultimatum game.

For each preference, we selected those survey items for constructing the GPS that exhibited the highest predictive power for the corresponding incentivized behavioral measure (*25*). Formally, for each preference the behavioral measure was regressed on different combinations of the survey items. The combination that maximized the adjusted R^2 was then selected for the respective preference.

Twelve survey questions were selected for the GPS; these comprised a mixture of qualitative items, measured on an 11-point Likert scale (0 to 10), and quantitative items involving economic trade-off decisions. Risk-taking was elicited by (i) an item determining the indifference point between a lottery with 50% chance of winning and receiving a fixed certain payment, and (ii) the response to the item "Please tell me, in general, how willing or unwilling you are to take risks." Patience was elicited by (i) an item determining the indifference point between receiving a fixed monetary amount on the day of the survey and a larger amount 12 months later, and (ii) the response to the question "How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?" Positive reciprocity was elicited by (i) an item asking for the value of a thank-you gift the respondent is willing to give in return for help by a stranger and (ii) the response to the item "When someone does me a favor I am willing to return it." Negative reciprocity was elicited by responses to the items (i) "If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so," (ii) "How willing are you to punish someone who treats you unfairly, even if there may be costs for you?", and (iii) "How willing are you to punish someone who treats others unfairly, even if there may be costs for you?" Altruism was elicited by (i) the quantitative value in response to the question "Imagine the following situation: Today you unexpectedly received 1000 euros. How much of this amount would you donate to a good cause?" and (ii) the response to the question "How willing are you to give to good causes without expecting anything in return?" Trust was elicited by the response to the item "I assume that people have only the best intentions." For each preference, the final survey measure was given as the weighted average of the z-scores of the corresponding survey items. The weights were calculated as the coefficients in OLS regressions of the incentivized behavioral measures on the respective survey items.

Selection of countries, translation of survey items, and pretest

For the GPS, 76 countries were selected with the goal of providing representative coverage of the global population. As a key criterion, the selected countries covered all development levels and geographic regions, including 24 in Europe, 22 in Asia, 1 in Oceania, 14 in Africa, and 15 in the Americas. Further, the selection process aimed at maximizing variation along country characteristics (such as language and historical, political, and ecological conditions) and favored culturally distinct and non-neighboring countries.

For each country, the selected survey items were translated into the country's major languages by at least three translators for each language. A first translator suggested, depending on the region of the target language, an English, French, or Spanish version of the item. A second translator conducted the translation into the target language. A third translator conducted a translation back to the original language. If a discrepancy occurred, the process was iterated until all translators agreed. Furthermore, monetary amounts used in the survey questions were adjusted to correspond to the same share in the median income of the target countries.

The survey items were pretested as part of the Gallup World Poll 2012 pretest, conducted at the end of 2011 in 22 countries with a sample size of 10 to 15 respondents per country. No respondent indicated problems in understanding the wording or the quantitative content of the survey items. Some respondents suggested rewording, which was incorporated through minor adjustments of some survey items.

Sampling and selection of respondents

We included the GPS as part of the Gallup World Poll 2012 through the infrastructure of Gallup (*23*). Respondents were sampled to achieve national representativeness of the resident population aged 15 and older. Telephone interviews were conducted in regions where at least 80% of the country's population is covered by telephone or where it is the customary survey methodology. Otherwise, face-to-face interviews were conducted.

The selection of households in countries with telephone interviews used either a random-digit-dialing method or nationally representative lists of phone numbers. In countries with face-to-face interviews, primary sampling units were stratified by population size and/or geography. To select sampled households, we used a random-route procedure. Respondents were selected randomly by either the latest-birthday or Kish grid method.

REFERENCES AND NOTES

1. M. Bertrand, in *Handbook of Labor Economics*, O. Ashenfelter, D. Card, Eds. (Elsevier, 2011), vol. 4b, pp. 1543–1590.
2. R. Croson, U. Gneezy, Gender Differences in Preferences. *J. Econ. Lit.* **47**, 448–474 (2009). doi: 10.1257/jel.47.2.448
3. E. E. Maccoby, C. N. Jacklin, *The Psychology of Sex Differences* (Stanford Univ. Press, 1974).
4. T. DeLeire, H. Levy, Worker Sorting and the Risk of Death on the Job. *J. Labor Econ.* **22**, 925–953 (2004). doi: 10.1086/423159
5. T. Buser, M. Niederle, H. Oosterbeek, Gender, Competitiveness, and Career Choices. *Q. J. Econ.* **129**, 1409–1447 (2014). doi: 10.1093/qje/qju009
6. D. M. Buss, Psychological sex differences: Origins through sexual selection. *Am. Psychol.* **50**, 164–168 (1995). doi: 10.1037/0003-066X.50.3.164
7. D. C. Geary, *Male, Female: The Evolution of Human Sex Differences* (American Psychological Association, 1998).
8. A. H. Eagly, *Sex Differences in Social Behavior: A Social-Role Interpretation* (Erlbaum, 1987).
9. A. H. Eagly, W. Wood, The origins of sex differences in human behavior: Evolved dispositions versus social roles. *Am. Psychol.* **54**, 408–423 (1999). doi: 10.1037/0003-066X.54.6.408
10. U. Gneezy, K. L. Leonard, J. A. List, Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society. *Econometrica* **77**, 1637–1664 (2009). doi: 10.3982/ECTA6690
11. E. Duflo, Women Empowerment and Economic Development. *J. Econ. Lit.* **50**, 1051–1079 (2012). doi: 10.1257/jel.50.4.1051
12. M. Doepke, M. Tertilt, Women's Liberation: What's in It for Men? *Q. J. Econ.* **124**, 1541–1591 (2009). doi: 10.1162/qjec.2009.124.4.1541
13. R. Fernandez, A. Fogli, C. Olivetti, Mothers and Sons: Preference Formation and Female Labor Force Dynamics. *Q. J. Econ.* **119**, 1249–1299 (2004). doi: 10.1162/0033553042476224
14. R. Fernández, A. Fogli, Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *Am. Econ. J. Macroecon.* **1**, 146–177 (2009). doi: 10.1257/mac.1.1.146
15. R. Inglehart, *The Silent Revolution: Changing Values and Political Styles Among Western Publics* (Princeton Univ. Press, 1977).
16. R. Inglehart, P. Norris, *Rising Tide: Gender Equality and Cultural Change Around the World* (Cambridge Univ. Press, 2003).
17. I. Almås, A. W. Cappelen, K. G. Salvanes, E. Ø. Sørensen, B. Tungodden, Willingness to Compete: Family Matters. *Manage. Sci.* **62**, 2149–2162 (2016). doi: 10.1287/mnsc.2015.2244
18. J. Haushofer, E. Fehr, On the psychology of poverty. *Science* **344**, 862–867 (2014). doi: 10.1126/science.1232491

19. T. Tanaka, C. F. Camerer, Q. Nguyen, Risk and time preferences: Experimental and household survey data from Vietnam. *Am. Econ. Rev.* **100**, 557–571 (2010). doi: 10.1257/aer.100.1.557
20. V. McLoyd, Socioeconomic disadvantage and child development. *Am. Psychol.* **53**, 185–204 (1998). doi: 10.1037/0003-066X.53.2.185
21. One may be agnostic about the ultimate determinants of gender-specific preferences. They may be acquired through cross-culturally universal gender roles, or they may be due to biological and evolutionary differences between women and men.
22. A. Falk, A. Becker, T. Dohmen, B. Enke, D. Huffman, U. Sunde, Global Evidence on Economic Preferences. *Q. J. Econ.* **133**, 1645–1692 (2018). doi: 10.1093/qje/qjy013
23. For further details on the Gallup World Poll, see www.gallup.com/analytics/213704/world-poll.aspx.
24. We note that trust is not a preference but a composite trait, including beliefs about others' behavior, prosocial preferences, and preferences for risk-taking. Given its importance, however, we decided to include it in our analysis. The results remained unchanged when excluding trust from the analysis; see supplementary materials.
25. A. Falk, A. Becker, T. Dohmen, D. Huffman, U. Sunde, "The preference survey module: A validated instrument for measuring risk, time, and social preferences" (IZA Discussion Paper 9674, 2016).
26. The raw gender differences without controls are qualitatively very similar (table S1).
27. This relationship remained robust when controlling for country-specific differences such as geographic, demographic, and historical factors (table S2). A complementary analysis further showed that gender differentiation in preferences was driven by those components of economic development that disproportionately benefit females relative to males (table S3).
28. We also investigated, separately for each preference, the relationship between gender differences and economic development conditional on gender equality, and vice versa. For each preference, gender differences were found to be strongly associated with log GDP per capita conditional on the Gender Equality Index (fig. S5). Likewise, gender differences were found to be highly associated with the Gender Equality Index conditional on log GDP per capita (fig. S6).
29. M. Yuki, Intergroup Comparison versus Intragroup Relationships: A Cross-Cultural Examination of Social Identity Theory in North American and East Asian Cultural Contexts. *Soc. Psychol. Q.* **66**, 166–183 (2003). doi: 10.2307/1519846
30. S. Guimond, A. Chatard, D. Martinot, R. J. Crisp, S. Redersdorff, Social comparison, self-stereotyping, and gender differences in self-construals. *J. Pers. Soc. Psychol.* **90**, 221–242 (2006). doi: 10.1037/0022-3514.90.2.221
31. S. Guimond *et al*., Culture, Gender, and the Self: Variations and Impact of Social Comparison Processes. *J. Pers. Soc. Psychol.* **92**, 1118–1134 (2007). doi: 10.1037/0022-3514.92.6.1118
32. S. Guimond, Psychological similarities and differences between women and men across cultures. *Soc. Personal. Psychol. Compass* **2**, 494–510 (2008). doi: 10.1111/j.1751-9004.2007.00036.x
33. A specific concern involves cross-cultural differences in gender-specific social comparison (*29*–*32*): Survey respondents in less developed, less gender-egalitarian countries may be inclined to define themselves relative to people of the same gender (intragroup social comparison). In contrast, those in more developed and gender-egalitarian countries may be likely to compare themselves to both women and men (intergroup social comparison). As a consequence, gender differences estimated through self-assessments may be understated in less developed countries.
34. P. T. Costa, A. Terracciano, R. R. McCrae, Gender differences in personality traits across cultures: Robust and surprising findings. *J. Pers. Soc. Psychol.* **81**, 322–331 (2001). doi: 10.1037/0022-3514.81.2.322
35. D. P. Schmitt, A. Realo, M. Voracek, J. Allik, Why Can't a Man Be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *J. Pers. Soc. Psychol.* **94**, 168–182 (2008). doi: 10.1037/0022-3514.94.1.168
36. S. H. Schwartz, T. Rubel, Sex differences in value priorities: Cross-cultural and multimethod studies. *J. Pers. Soc. Psychol.* **89**, 1010–1028 (2005). doi: 10.1037/0022-3514.89.6.1010
37. An example for the biological role in shaping human preferences comes from twin studies (*39*–*42*).
38. For instance, our results can be accommodated by approaches suggesting interactions of evolutionary or biological determinants with the social environment (*43*).
39. D. Cesarini, C. T. Dawes, M. Johannesson, P. Lichtenstein, B. Wallace, Genetic variation in preferences for giving and risk taking. *Q. J. Econ.* **124**, 809–842 (2009). doi: 10.1162/qjec.2009.124.2.809
40. N. Zethraeus *et al*., A randomized trial of the effect of estrogen and testosteron on economic behavior. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6535–6538 (2009). doi: 10.1073/pnas.0812757106
41. D. Cesarini *et al*., Heritability of cooperative behavior in the trust game. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3721–3726 (2008). doi: 10.1073/pnas.0710069105
42. B. Wallace, D. Cesarini, P. Lichtenstein, M. Johannesson, Heriability of ultimatum game responder behavior. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15631–15634 (2007). doi: 10.1073/pnas.0706642104
43. D. C. Geary, Evolution and developmental sex differences. *Curr. Dir. Psychol. Sci.* **8**, 115–120 (1999). doi: 10.1111/1467-8721.00027

ACKNOWLEDGMENTS

We thank P. Andre, E. Avis, S. Fiedler, K. Jensen, M. Kosfeld, L. Shiferaw, and J. Tungodden for helpful comments; L. Henkel, R. Suchy, and S. Walter for outstanding research assistance; M. Antony for his support in collecting and managing the GPS data; and M. Entner and S. Sauter for administrative support. **Funding:** Supported by the European Research Council through ERC #209214 (A.F.). **Competing interests:** Authors declare no competing interests. **Data availability:** The data on preferences are deposited at www.briq-institute.org/global-preferences/home.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/eaas9899/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S14
References (*44*–*52*)

16 January 2018; accepted 20 August 2018
10.1126/science.aas9899

REPORT

QUANTUM COMPUTING

Quantum advantage with shallow circuits

Sergey Bravyi[1], David Gosset[1]*, Robert König[2]†

Quantum effects can enhance information-processing capabilities and speed up the solution of certain computational problems. Whether a quantum advantage can be rigorously proven in some setting or demonstrated experimentally using near-term devices is the subject of active debate. We show that parallel quantum algorithms running in a constant time period are strictly more powerful than their classical counterparts; they are provably better at solving certain linear algebra problems associated with binary quadratic forms. Our work gives an unconditional proof of a computational quantum advantage and simultaneously pinpoints its origin: It is a consequence of quantum nonlocality. The proposed quantum algorithm is a suitable candidate for near-future experimental realizations, as it requires only constant-depth quantum circuits with nearest-neighbor gates on a two-dimensional grid of qubits (quantum bits).

Quantum nonlocality—the fact that measurements of multiple spatially separated quantum systems can yield certain classically nonlocal correlations—is a remarkable feature of quantum mechanics. Its understanding was revolutionized by Bell, who recognized that this phenomenon can be experimentally verified (*1*). His work also prompted a shift in our approach to quantum mechanical systems: Modern quantum information science seeks to exploit quantum phenomena such as nonlocality, entanglement, and the superposition principle to tackle information-processing tasks that are believed to be difficult or impossible to achieve by purely classical means.

The potential benefits resulting from the use of quantum resources in computation are challenging to quantify. A major difficulty is the fact that the computational hardness of a given task for a classical computing device is not easy to assess. The strongest evidence that quantum computers may be more powerful than classical computers is the existence of quantum algorithms, such as Shor's efficient factoring algorithm (*2*), that achieve better performance than the best currently known classical algorithms for certain tasks. The appeal of Shor's algorithm relies on the conjecture that factoring cannot be achieved in polynomial time on a classical computer. Other existing proposals for quantum speed-ups [see, e.g., (*3*)] similarly rely on complexity-theoretic conjectures and, as a result, the potential for quantum advantage disappears if these conjectures turn out to be false (*4*). To sidestep this issue, one can instead work within the framework of so-called query complexity, in which, in addition to a classical or quantum computer, we are also given black-box access to an "oracle" operation whose internal structure cannot be examined. The computational task is to determine some property of the oracle, using the smallest possible number of queries to the oracle. Notable examples of quantum speed-ups obtained in terms of query complexity include quantum algorithms for unstructured search (*5*), element distinctness (*6*), and formula evaluation (*7*). Unfortunately, a speed-up proved in the oracular setting might not translate into a real-world advantage because the internal structure of an oracle may help to solve the problem classically.

A more practical concern is the fact that most proposed quantum algorithms are beyond current experimental capabilities: Their implementation requires a fully functional quantum computer incorporating error correction. Although the overhead for encoding and manipulating quantum data fault-tolerantly is asymptotically small (*8*, *9*), it remains prohibitive for current technology. Accordingly, it is expected that near-term quantum computers will lack error correction capabilities (*10*–*12*). A quantum computation without error correction can perform only a constant number of operations before the qubits decohere and the entropy builds up. This is evidenced by the impossibility of realizing passive quantum memories (amounting to the identity gate) when qubits experience independent noise with constant decoherence rate (*13*). This leads to an interesting trade-off between the number of qubits n and the depth d of quantum circuits that can be implemented on near-term hardware. Because the total number of operations is proportional to nd, quantum circuits with a large number of qubits that can potentially be hard to simulate classically are limited to a small depth.

This motivates the study of quantum parallel algorithms running in a constant time. These take as input a classical bit string, apply a constant-depth quantum circuit, and output a random bit string obtained by measuring each qubit in the 0,1 basis. By definition, a quantum circuit of depth d consists of d layers of one- and two-qubit gates, where gates within each layer act on disjoint sets of qubits. We assume that such disjoint gates can be applied in parallel. Constant-depth quantum circuits can therefore be implemented in a constant time, which is independent of the problem size or the number of qubits n. For brevity, we refer to such parallel constant-time quantum computations as shallow quantum circuits (SQCs).

Although SQCs constitute a highly restricted form of quantum computation, they are of central practical and theoretical importance. First, as discussed above, SQCs are among those forms of quantum computations most likely to be realized in the near future. Second, there are reasons to believe that SQCs may outperform classical circuits in certain tasks. The first evidence in this direction was provided in pioneering work (*14*) where it was argued that SQCs may be computationally hard to simulate classically. Quite recently, building on earlier studies of so-called instantaneous quantum polynomial-time circuits (*15*–*17*) and relying on complexity-theoretic assumptions, it was shown (*18*) that the output distribution of SQCs may be computationally hard to sample classically even if a constant statistical error is allowed [see also (*19*)]. Various models of computation closely related to SQCs have been studied (*20*–*26*).

Here, we compare the computational power of SQCs and their classical counterparts—that is, constant-depth classical (probabilistic) circuits. We present a simple linear algebra problem associated with binary quadratic forms that can be solved with certainty by a SQC composed of nearest-neighbor gates acting on a two-dimensional (2D) grid; this setup reflects near-term experimental abilities. At the same time,

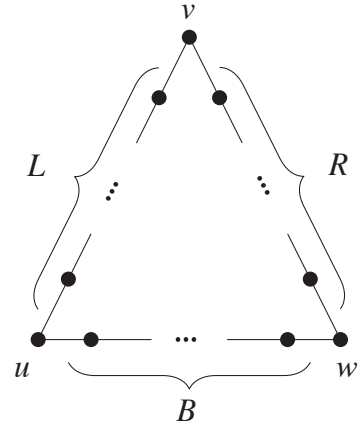


Fig. 1. Cycle graph Γ. Shown is a cycle Γ of even length M and three vertices u, v, w such that all pairwise distances are even. The sides L, R, B may have unequal lengths.

[1]IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA. [2]Technical University of Munich, 85748 Garching, Germany.
*Present address: Department of Combinatorics & Optimization and Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
†Corresponding author. Email: robert.koenig@tum.de

we prove that no constant-depth classical probabilistic circuit can solve the considered problem with a sufficiently small error probability for all instances. The classical circuit does not have to be geometrically local in any sense and may access random bits drawn from an arbitrary probability distribution that depends only on the input size. The only requirement is that all gates in the classical circuit must have bounded fan-in (i.e., each gate has a constant number of input wires). This result provides an unconditional separation between the power of constant-depth quantum and classical circuits.

The computational problem we use to establish the above separation can be viewed as a nonoracular variant of the well-known Bernstein-Vazirani problem (*27*). The latter involves a quantum oracle computing a linear function l: $\{0,1\}^n \to \{0,1\}$ such that $l(x) = \sum_{i=1}^{n} z_i x_i (\mathrm{mod}2)$ for some "hidden" bit string $z \in \{0,1\}^n$. The oracle is a unitary operator that can evaluate $l(x)$ on any superposition of inputs x. The task is to identify the hidden string z. As was shown in (*27*), the problem can be solved by a simple quantum algorithm making a single query to the oracle, whereas any classical algorithm with access to a classical oracle computing ℓ requires n queries to find z. To obtain a non-oracular variant of this problem, we ask: Where else—other than inside an oracle—can we hide a linear function?

We now explain how to hide a linear Boolean function inside a quadratic form. We consider quadratic forms q: $\{0,1\}^n \to \mathbb{Z}_4 = \{0,1,2,3\}$ defined as

$$q(x) = \sum_{i,j=1}^{n} A_{i,j} x_i x_j (\mathrm{mod}4) \qquad (1)$$

where $x = (x_1, \ldots, x_n)$ is a vector of n binary variables and A is a symmetric binary matrix; that is, $A_{i,j} = A_{j,i} \in \{0,1\}$. Evaluation of $q(x)$ modulo 4 ensures that $q(x)$ depends only on the value of input variables modulo 2. As a consequence, many properties of real-valued quadratic forms carry over to the binary case considered here. We are interested in the case when x lies in the binary null-space of A,

$$\mathrm{Ker}(A) = \{x \in \{0,1\}^n : Ax = 0 (\mathrm{mod}2)\} \qquad (2)$$

Here and below, we consider x as a column vector. We write x^T for the transposed row vector. If x is a real vector and $q(x) = x^T A x$ is a real-valued quadratic form, then the restriction of $q(x)$ onto the null-space of A is clearly zero because $Ax = 0$ implies $q(x) = 0$. In the binary case, this is no longer true. However, we will see that the restriction of $q(x)$ onto the binary null-space of A is always a linear function. In other words, for any quadratic form $q(x)$ as above, there exists a (generally not unique) vector $z \in \{0,1\}^n$ such that

$$q(x) = 2 \sum_{i=1}^{n} z_i x_i (\mathrm{mod}4) \text{ for all } x \in \mathrm{Ker}(A) \qquad (3)$$

[See (*28*) for the proof of Eq. 3.] Thus, one can say that the quadratic form $q(x)$ hides a linear function $l(x) = 2z^T x (\mathrm{mod}4)$, which is analogous to the hidden linear function in the Bernstein-Vazirani problem. However, in our case the form $q(x)$ is explicitly specified by its matrix of coefficients A, so there is no need for the oracle.

We consider the following hidden linear function (HLF) problem: Given an $n \times n$ symmetric binary matrix A specifying a quadratic form $q(x) = x^T A x$, which is evaluated modulo 4, find a binary vector $z \in \{0,1\}^n$ such that $q(x) = 2z^T x$ for all x in the binary null-space of A. [We note that any instance of the HLF problem (*29*) can be solved classically with $O(n^3)$ binary arithmetic operations by computing a basis of $\mathrm{Ker}(A)$ and evaluating $q(x)$ on each basis vector $x \in \mathrm{Ker}(A)$.]

An instance of the HLF problem can be alternatively described by a graph $G(A)$ with n vertices $i = 1, \ldots, n$ such that the off-diagonal part of A is the adjacency matrix of $G(A)$ and the nonzero entries on the diagonal of A specify a subset of marked vertices. Thus, by definition, a pair of vertices $i \neq j$ is connected by an edge if $A_{i,j} = 1$. We consider special instances of the HLF, where $n = N^2$ for some integer N and $G(A)$ is a subgraph of the square grid with $N \times N$ vertices. In other words, $A_{i,j} = 0$ unless i and j are nearest-neighbor vertices of the grid or $i = j$. Such instances constitute what we call the 2D HLF problem. As far as we are aware, the best classical algorithm that solves the 2D HLF problem requires $O(n^2)$ arithmetic operations (*30*).

A quantum algorithm solving the 2D HLF problem is similar to the one considered by Bernstein and Vazirani (*27*). Suppose first that one has access to a quantum circuit U_q acting on the basis states $x \in \{0,1\}^n$ as $U_q|x\rangle = i^{q(x)}|x\rangle$. Consider a system of n qubits and a quantum state

$$|\Psi_q\rangle = H^{\otimes n} U_q H^{\otimes n} |0\rangle^{\otimes n} = 2^{-n} \sum_{x,z \in \{0,1\}^n} i^{q(x)} (-1)^{z^T x} |z\rangle \qquad (4)$$

where

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad (5)$$

is the Hadamard gate. It can be easily checked that destructive interference between different terms x in Eq. 4 results in a vanishing amplitude $\langle z|\Psi_q\rangle = 0$ on basis states z that are not solutions of the HLF problem. Conversely, all solutions z appear in the state $|\Psi_q\rangle$ with nonzero amplitudes (*28*). Therefore, measuring each qubit of $|\Psi_q\rangle$ in the standard basis produces a random bit string z that is a solution of the HLF problem. The circuit U_q can be decomposed into a product of controlled-Z gates CZ and phase-shift gates S defined as

$$S_j|x\rangle = i^{x_j}|x\rangle$$
$$CZ_{i,j}|x\rangle = (-1)^{x_i x_j}|x\rangle \qquad (6)$$

Here, the subscripts i, j indicate qubits acted upon by the gate. Indeed, applying a $CZ_{i,j}$ gate for each pair of qubits such that $A_{i,j} = 1$ and applying an S_i gate for each qubit such that $A_{i,i} = 1$ gives U_q.

Let us consider the quantum resources required to implement the circuit U_q. We place the ith qubit at the ith vertex of the graph $G(A)$ embedded into a 2D grid. Then U_q contains only nearest-neighbor CZ gates and single-qubit S gates.

One can decompose the CZ part of U_q into four layers of CZ gates such that each layer is a depth-one circuit composed of pairwise disjoint CZ gates. This shows that U_q is a constant-depth quantum circuit. The above algorithm can be easily converted into a SQC by adding ancillary qubits that store the input matrix A and replacing each gate of U_q by its controlled version. We envision each input bit $A_{i,j}$ as being provided at the edge $\{i, j\}$ and each bit $A_{i,i}$ at the vertex i of the 2D grid. Then each gate in the controlled version of U_q acts on at most three qubits located on the same edge or the same vertex of the grid. Hence, the 2D HLF can be solved with certainty by a SQC with geometrically local gates on a 2D grid.

The above algorithm can be alternatively described as a sequence of single-qubit Pauli measurements performed on the graph state (*31*, *32*) associated with the graph $G(A)$. The latter is a quantum state of n qubits defined as

$$|\Psi_{G(A)}\rangle = \prod_{1 \le i < j \le n} CZ_{i,j}^{A_{i,j}} \cdot H^{\otimes n} |0\rangle^{\otimes n} \qquad (7)$$

Here, a CZ gate is applied for every edge of the graph $G(A)$. We claim that a solution z of the HLF problem can be obtained by preparing the graph state $|\Psi_{G(A)}\rangle$ and measuring each qubit i in either the X basis (if $A_{i,i} = 0$) or the Y basis (if $A_{i,i} = 1$). Indeed, a comparison of Eqs. 4 and 7 reveals that $|\Psi_q\rangle$ can be obtained from $|\Psi_{G(A)}\rangle$ by applying a gate S_i to each qubit i with $A_{i,i} = 1$ and applying a layer of n Hadamard gates. Furthermore, because the states Ψ_q and $\Psi_q{}^*$ (the complex conjugate of Ψ_q) give rise to the same measurement statistics in the Z basis, one can replace all gates S_i by their complex conjugates $S_i{}^* = S_i{}^3$. The above claim then follows from the identities $ZH = HX$ and $Z(HS^3) = (HS^3)Y$.

To understand the difficulty of solving the 2D HLF problem using classical circuits, recall that such a circuit is specified by a directed acyclic graph in which each vertex is either an input (if it has in-degree 0), an output (if it has out-degree 0), or a gate. For each gate we must also specify a Boolean function $\{0,1\}^k \to \{0,1\}$ that is computed by the gate, where k is the in-degree or "fan-in" of the gate. The output bit computed by a gate is copied to all outgoing edges of this gate. The out-degree or "fan-out" of a gate is the number of times the output of the gate is used in the remainder of the circuit. The depth d of a classical circuit is the maximum number of gates along a path from an input (variable) to an output. We are interested in circuits $\mathcal{C}$ with constant depth $d = O(1)$ such that all gates have fan-in at most $K = O(1)$.

The structure of such a circuit $\mathcal{C}$ imposes severe restrictions on the form of correlations present between the input (*33*) $x \in \{0,1\}^m$ and outputs $z = F(x) \in \{0,1\}^n$, where F is the function computed by $\mathcal{C}$. Let us say that the input bit (variable) x_j is correlated with the output bit (variable) z_k if and only if there is some $y \in \{0,1\}^m$ such that flipping the jth bit of y flips the kth bit of $F(y)$. Then the function F computed by $\mathcal{C}$ is local in the sense that each output bit can only be correlated with a constant number of input bits. Indeed, defining the ("backward") lightcone of the output bit z_k as the set $L_{\mathcal{C}}(z_k)$ of input variables x_j correlated with z_k, we have the obvious inequality

$$|L_{\mathcal{C}}(z_k)| \leq K^d \quad \text{for all output bits } z_k \qquad (8)$$

Similarly, we can argue [see claim 5 in (*28*)] that a constant fraction of input bits x_j have a small ("forward") lightcone $L_{\mathcal{C}}(x_j)$. The latter is the set of output variables z_k correlated with x_j. We will see that locality restrictions of this kind ultimately prevent solution of the 2D HLF by constant-depth classical circuits.

To understand why, recall that correlations present in the measurement statistics of entangled quantum states exhibit quantum nonlocality. A famous illustration (*34*, *35*) of this phenomenon is based on the three-qubit GHZ state $|\mathrm{GHZ}\rangle = (1/\sqrt{2})(|000\rangle + |111\rangle)$. Let $b \in \{0,1\}^3$ and consider the measurement outcomes $z \in \{0,1\}^3$ obtained by measuring each qubit j of $|\mathrm{GHZ}\rangle$ in either the X basis (if $b_j = 0$) or the Y basis (if $b_j = 1$). Then the measurement statistics satisfy

$$i^{b_1+b_2+b_3}(-1)^{z_1+z_2+z_3} = 1 \quad \text{whenever} \quad b_1 \oplus b_2 \oplus b_3 = 0 \cdot \qquad (9)$$

In contrast, it is well known (*35*) that Eq. 9 cannot be satisfied by a local hidden-variable model in which every output bit z_j is a function only of the local measurement setting b_j and some shared random string r, that is, $z_j = z_j(b_j, r)$. A local hidden-variable model can equivalently be viewed as a strictly local classical circuit in which every output bit depends only on one input bit (as well as the shared random string). Thus, rephrasing the GHZ example in terms of circuits, we conclude that a strictly local, possibly randomness-assisted classical circuit cannot solve the relational problem of producing a string $z \in \{0,1\}^3$ satisfying Eq. 9 for any input $b \in \{0,1\}^3$.

Our main technical contribution is a proof that this type of quantum nonlocality provides a relational problem that cannot be solved by general constant-depth classical circuits in the case when the GHZ state is replaced by the graph state defined in Eq. 7. To arrive at this result, we first consider 1D instances of the HLF problem and show that a constant-depth classical circuit with geometrically local gates cannot solve all such instances. We then use this result to show that a classical circuit that is "constant-depth local" (in the sense of having constant-size backward lightcones) cannot solve all instances of the 2D HLF problem.

Our starting point is an extension of the GHZ example described in (*36*). It shows that the statistics of single-qubit X and Y measurements performed on the graph state $|\Psi_\Gamma\rangle$ corresponding to a cycle graph Γ possess geometrically nonlocal correlations. In other words, certain measurement outcomes are correlated with measurement settings (i.e., choice of single-qubit measurement bases) that are far away with respect to the shortest path on the cycle. As a consequence, low-depth classical circuits that are geometrically local in one dimension cannot reproduce these statistics. Because the considered Pauli measurements are identical to those required to solve instances of the HLF problem on the cycle graph, we conclude that these instances cannot be solved by geometrically local constant-depth classical circuits.

In more detail, let Γ be a cycle graph of even length M and let Δ be the adjacency matrix of Γ. Suppose u, v, w are vertices of Γ such that all pairwise distances between them are even (Fig. 1). We define the distance $\mathrm{dist}_\Gamma(j, k)$ between any two vertices j, k to be the number of edges in the shortest path between them in Γ. Given a string $b = b_u b_v b_w \in \{0,1\}^3$, we can write $0^{M-3}b \in \{0,1\}^M$ for the string that associates the bits b_u, b_v, b_w to the vertices u, v, w and the value 0 to all other vertices. Let Δ_b be a matrix obtained from Δ by placing the string $0^{M-3}b$ on the main diagonal. This defines eight instances of the HLF problem on the cycle graph Γ with $A = \Delta_b$ and $b \in \{0,1\}^3$. Each instance has M output bits $z_1, \ldots, z_M$. In (*28*) we show that any classical randomness-assisted circuit $\mathcal{C}$ that solves all eight instances Δ_b with probability of at least ⅞ is geometrically nonlocal: The lightcone $L_{\mathcal{C}}(b_i)$ of at least one of the input bits $b_i \in \{b_u, b_v, b_w\}$ contains an output bit z_q such that $\mathrm{dist}_\Gamma(i,q) \geq \frac{1}{2}D_\Gamma(\{u, v, w\})$, where $D_\Gamma(\{u, v, w\})$ is the minimum pairwise distance between two vertices in the set $\{u, v, w\}$. The proof of this intermediate result proceeds by establishing a relation analogous to Eq. 9 for the graph state $|\Psi_\Gamma\rangle$ [similar to (*36*); see (*28*)].

Now consider the more general family of constant-depth classical circuits, which may not be geometrically local. Here, the HLF associated with the cycle graph Γ is not suitable for establishing our desired separation. Classical constant-depth circuits can communicate values b_u, b_v, b_w of input bits to distant locations in the cycle Γ and solve the HLF in this case essentially by simulating geometric nonlocality. We thus turn to the 2D HLF problem: We establish that the correlations between the input A and the solutions z in the 2D HLF problem have a strong form of nonlocality that cannot be reproduced by constant-depth randomness-assisted classical circuits of bounded fan-in. The key difference relative to the cycle graph setting is that the underlying graph $G(A)$ in the problem is also varied. Our main technical result is the following lower bound, which is valid for all sufficiently large N: If a classical circuit $\mathcal{C}_N$ with fan-in at most K solves all $N \times N$ instances of the 2D HLF problem with probability greater than ⅞, then the depth of $\mathcal{C}_N$ is at least $[1/(8 \log K)] \log N$.

Thus, no constant-depth circuit with bounded fan-in can solve all instances of the 2D HLF problem with high probability. The full proof of this result is given in (*28*). It proceeds by contradiction. Suppose a small-depth circuit $\mathcal{C}_N$ with bounded fan-in solves all size-N instances of the 2D HLF. To conform with the previous notations, we write $b_i \equiv A_{i,i}$ for the diagonal elements of A. We restrict our attention to instances A where $G(A)$ is a subgraph Γ of the 2D grid chosen in a particular way (depending on $\mathcal{C}_N$) such that

1) Γ is an even length cycle of length proportional to N,

2) There are vertices u, v, w on Γ, with even pairwise distances at least proportional to N, and

3) Any output bit z_q lying on the cycle Γ and correlated with one of the input bits b_u, b_v, b_w must be located in a small neighborhood of this input bit (of size proportional to $N^{1/2}$).

We establish the existence of such a cycle Γ by a probabilistic argument. Informally, restricting $\mathcal{C}_N$ to the instances of the HLF defined above gives rise to a certain geometrically local structure, as described by condition 3. However, such a geometrically local structure can be ruled out using the example of the HLF on the cycle graph. Indeed, in the latter case we have already shown that the lightcone of at least one input bit $b_i \in \{b_u, b_v, b_w\}$ must contain an output bit z_q located far from b_i, such that $\mathrm{dist}_\Gamma(i,q) \geq \frac{1}{2}D_\Gamma(\{u, v, w\})$. This contradicts condition 3 because $D_\Gamma(\{u, v, w\})$ is proportional to N. Thus, we obtain the desired lower bound on the depth of $\mathcal{C}_N$. This lower bound can be viewed as a worst-case hardness result, as we assume that the classical circuit solves all instances of the problem. In (*28*), we provide an analogous average-case hardness result that assumes only that the classical circuit solves a constant fraction of instances from a suitable random ensemble.

Our result constitutes a provable separation between analogously defined classical and quantum computational models that uses quantum nonlocality in answering a complexity-theoretic question. Our quantum circuit for the 2D HLF problem, which uses only classically controlled Clifford gates in a 2D geometry, is a promising target for future experimental demonstration of a quantum algorithm with provably better scaling (in terms of circuit depth) than any classical algorithm. Future theoretical work may address the question of minimizing resource requirements for practical demonstrations of such a quantum advantage, and may incorporate methods of fault tolerance tailored to hardware architectures of interest.

REFERENCES AND NOTES

1. J. S. Bell, *Physics* **1**, 195–200 (1964).
2. P. W. Shor, *SIAM Rev.* **41**, 303–332 (1999).
3. A. Montanaro, *NPJ Quantum Inf.* **2**, 15023 (2016).
4. Separations between quantum and classical information processing have been established in the area of communication complexity (*37*, *38*), but computational restrictions of communicating nodes are typically not considered in that field.
5. L. K. Grover, in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing* (ACM, 1996), pp. 212–219.
6. A. Ambainis, *SIAM J. Comput.* **37**, 210–239 (2007).

7. B. W. Reichardt, R. Spalek, in *Proceedings of the 40th Annual ACM Symposium on Theory of Computing* (ACM, 2008), pp. 103–112.
8. D. Gottesman, *Quantum Inf. Comput.* **14**, 1338–1371 (2014).
9. H. Bombín, *New J. Phys.* **17**, 083002 (2015).
10. K. Temme, S. Bravyi, J. M. Gambetta, *Phys. Rev. Lett.* **119**, 180509 (2017).
11. Y. Li, S. C. Benjamin, *Phys. Rev. X* **7**, 021050 (2017).
12. S. Boixo *et al.*, *Nat. Phys.* **14**, 595–600 (2018).
13. F. Pastawski, A. Kay, N. Schuch, I. Cirac, *Phys. Rev. Lett.* **103**, 080501 (2009).
14. B. M. Terhal, D. P. DiVincenzo, *Quantum Inf. Comput.* **4**, 134–145 (2004).
15. M. J. Bremner, A. Montanaro, D. J. Shepherd, *Phys. Rev. Lett.* **117**, 080501 (2016).
16. M. J. Bremner, A. Montanaro, D. J. Shepherd, *Quantum* **1**, 8 (2017).
17. E. Farhi, A. W. Harrow, arXiv:1602.07674 [quant-ph] (24 February 2016).
18. J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, J. Eisert, *Phys. Rev. X* **8**, 021010 (2018).
19. X. Gao, S.-T. Wang, L.-M. Duan, *Phys. Rev. Lett.* **118**, 040502 (2017).
20. C. Moore, M. Nilsson, *SIAM J. Comput.* **31**, 799–815 (2001).
21. R. Cleve, J. Watrous, in *Proceedings, 41st Annual Symposium on Foundations of Computer Science* (IEEE, 2000), pp. 526–536.
22. P. Høyer, R. Špalek, *Theory Comput.* **1**, 83–101 (2005).
23. D. Browne, E. Kashefi, S. Perdrix, in *Conference on Quantum Computation, Communication, and Cryptography* (Springer, 2010), pp. 35–46.
24. L. Eldar, A. W. Harrow, in *58th Annual Symposium on Foundations of Computer Science* (IEEE, 2017), pp. 427–438.
25. S. Aaronson, *Proc. R. Soc. London Ser. A* **461**, 3473–3482 (2005).
26. S. Aaronson, L. Chen, in *Proceedings of the 32nd Computational Complexity Conference* (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2017), pp. 22:1–22:67.
27. E. Bernstein, U. Vazirani, *SIAM J. Comput.* **26**, 1411–1473 (1997).
28. See supplementary materials.
29. Our analysis [see Lemma 2 in (*28*)] also shows that the HLF problem can be regarded as a non-oracular variant of the "Fourier Fishing problem" introduced in (*39*), where the task is to find a nonzero Fourier coefficient of a function. In our case, the function of interest is $x \rightarrow i^{q(x)}$.
30. Because our quantum algorithm for the 2D HLF problem only involves classically controlled Clifford gates applied to an initial stabilizer state, it can be simulated classically using the Gottesman-Knill theorem (*40*, *41*). Imagine an "active window" W of width $O(1)$ that sweeps through the $N \times N$ lattice from left to right. At each step of the simulation, we measure qubits in the leftmost column of W in the appropriate basis (X or Y), shift W one column to the right, and then apply CZ gates near the right boundary of W to incorporate a newly arrived column of qubits into the graph state. Note that all qubits outside of W are unentangled: Qubits on the left have been already measured, and qubits on the right have not yet been incorporated into the graph state. Thus, the Gottesman-Knill simulator only needs to keep track of $O(N)$ qubits located in W, leading to a classical algorithm using $O(N^4) = O(n^2)$ arithmetic operations.
31. R. Raussendorf, D. E. Browne, H. J. Briegel, *Phys. Rev. A* **68**, 022312 (2003).
32. M. Hein, J. Eisert, H. J. Briegel, *Phys. Rev. A* **69**, 062311 (2004).
33. Our results also hold for probabilistic circuits in which a subset of input bits may be chosen at random; that is, $x = x'r$, where $r \in \{0,1\}^{\ell}$ is a random string drawn from some (arbitrary) distribution $\rho(r)$ and $x' \in \{0,1\}^{m-\ell}$.
34. D. M. Greenberger, M. A. Horne, A. Shimony, A. Zeilinger, *Am. J. Phys.* **58**, 1131–1143 (1990).
35. N. D. Mermin, *Phys. Rev. Lett.* **65**, 1838–1840 (1990).
36. J. Barrett, C. M. Caves, B. Eastin, M. B. Elliott, S. Pironio, *Phys. Rev. A* **75**, 012103 (2007).
37. G. Brassard, *Found. Phys.* **33**, 1593–1616 (2003).
38. R. de Wolf, *Theor. Comput. Sci.* **287**, 337–353 (2002).
39. S. Aaronson, in *Proceedings of the 42nd ACM Symposium on Theory of Computing* (ACM, 2010), pp. 141–150.
40. D. Gottesman, in *Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics* (International Press, 1999), pp. 32–43.
41. S. Aaronson, D. Gottesman, *Phys. Rev. A* **70**, 052328 (2004).

ACKNOWLEDGMENTS

We thank the anonymous referees for their comments. **Funding:** Supported by the IBM Research Frontiers Institute (S.B. and D.G.) and by the Technical University of Munich–Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement 291763 (R.K.). **Author contributions:** S.B., D.G., and R.K. contributed equally to this work. **Competing interests:** None.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/308/suppl/DC1
Supplementary Text
Figs. S1 and S2

24 October 2017; accepted 30 August 2018
10.1126/science.aar3106

ORGANIC CHEMISTRY

Chiral Lewis acids integrated with single-walled carbon nanotubes for asymmetric catalysis in water

Taku Kitanosono, Pengyu Xu, Shū Kobayashi*

The development of highly reactive and stereoselective catalytic systems is required not only to improve existing synthetic methods but also to invent distinct chemical reactions. Herein, a homogenized combination of nickel-based Lewis acid–surfactant-combined catalysts and single-walled carbon nanotubes is shown to exhibit substantial activity in water. In addition to the enhanced reactivity, stereoselective performance and long-term stability were demonstrated in asymmetric conjugate addition reactions of aldoximes to furnish chiral nitrones in high yields with excellent selectivities. The practical and straightforward application of the designed catalysts in water provides an expedient, environmentally benign, and highly efficient pathway to access optically active compounds.

Synthetic organic chemistry remains reliant on organic solvents as reaction media, despite the inherent safety and sustainability advantages of water (*1*). Most organic substances are insoluble in or immiscible with water, and reagents as well as catalysts are also often hydrolytically unstable. In the late 1990s, Lewis acid catalysts were combined with a micelle-forming functional group such as dodecylsulfate to generate Lewis acid–surfactant-combined catalysts (hereafter LASCs) that could be used to overcome these problems (*2*). Dispersed lipophilic environments were formed around the Lewis acidic cations to entrap and concentrate hydrophobic small molecules in water. Moreover, appropriate reaction environments in water would impose stricter stereochemical regulation compared with those in organic solvents because of hydrogen-bond networks, specific solvation, and hydrophobic interactions (*3*, *4*). Stereoselective installation still remains a more challenging task in water because many metal-ligand complexes are poorly soluble or insoluble in water, and water molecules can interfere with the catalyst-substrate interactions that are vital for stereoselection.

Aldoximes have been identified as ideal nucleophiles to provide oxime ethers or nitrones (Fig. 1A). The higher electron density of the sp^3-hybridized oxygen atom relative to the sp^2-hybridized nitrogen atom imposes a preference for O-conjugate addition reactions upon treatment with a (salen) aluminum complex (*5*) or an iminium catalyst (*6*, *7*). On the other hand, softer Lewis acids such as Zn^{2+} and Cd^{2+} were reported to promote nitrone formation through N-conjugate addition (*8*–*10*). The synthesis of optically active nitrones by this route would provide useful pharmaceutical intermediates (*11*). To attain this goal, we envisioned that a Lewis acid could be integrated with π-conjugated materials that have a hydrophobic large surface area to entrap and concentrate hydrophobic small molecules (reactants and a ligand) and that also have a huge conjugated π network to modify the electronic nature of the metal (Fig. 1B). Among such π-conjugated materials, single-walled carbon nanotubes (SWNTs) have attracted wide attention since their advent in 1991 because of their distinctive electronic and physical properties (*12*, *13*). SWNTs have been used as supports for zero-valent metal clusters (*14*) and nanoparticles (*15*), as well as metal oxides (*16*–*18*), through physical adsorption, adhesion, and chemical modifications (*19*). However, given that the insolubility of SWNTs in most solvents presents a major impediment to their practical application, substantial efforts have been devoted to generating suspensions of SWNTs in a controlled manner. Approaches have included modulation of their surface properties with dispersants, including anionic surfactants such as sodium dodecylsulfate (SDS) (*20*, *21*), polycyclic aromatic compounds, supramolecular biomolecules, and synthetic polymers. For instance, the (*n*, *m*) separation (where *n* and *m* are any integers) of SWNTs relies on selective adsorption, which is dependent on differences in the surfactant structure (*22*). We took advantage of the strong van der Waals force between SWNTs and carbon chains of dodecylsulfate to anchor Lewis acid cationic centers onto the surfaces of SWNTs without any perturbation of the cylindrical structure, unlike covalent SWNT modification. Given that the electronic performance of SWNTs is enhanced by increased dispersion (*23*), we attempted to achieve a uniform dispersion of SWNTs with LASCs by ultrasonication.

We commenced our study with a selection of homogeneous dispersions of LASC–SWNT nanocomposites containing first-row transition metals, and we evaluated the catalytic activities of the systems in the asymmetric conjugate addition reaction of benzaldoxime (**1a**) with non-3-en-2-one (**2a**) (table S2, entries 1 to 6). When combined with chiral 2,2′-bipyridine ligand **L** (*24*, *25*), a homogeneous dispersion of LASC–SWNT nanocomposites based on a vanadium ion, as a representative hard Lewis acid, was ineffective for this reaction. Whereas typical water-compatible Lewis acids such as iron(II), copper(II), and zinc(II) ions showed low activity, a homogeneous dispersion of $Ni(OSO_3C_{12}H_{25})_2$–SWNT nanocomposite (hereafter $OSO_3C_{12}H_{25}$ = DS) was found to promote the conjugate addition reaction of **1a** with **2a** to afford the optically active nitrone in 79% yield with 97% enantiomeric excess (ee) (Table 1, entry 1). In contrast, a micellar environment constructed by $Ni(DS)_2$ without SWNTs suffered from low yield and low selectivity (Table 1, entry 2). The intrinsically poor stability and small rate constant for water exchange (WERC) of the nickel(II) aqua ion can account for these differences (*26*). The aqueous solution of $Ni(DS)_2$ is acidic (pH 4.37), and a precipitate formed when it was allowed to stand for 24 hours, which suggested that hydrolysis occurred (figs. S14 and S15). Moreover, the WERC of the nickel ion is smaller than that of the scandium ion by three orders of magnitude, which is associated with its intrinsically poor Lewis acidity, especially in water (table S1). Inverted microscopic analysis indicated why the experimental outcomes observed with $Ni(DS)_2$ and the $Ni(DS)_2$–SWNT composite in water differed (Fig. 1C). When both reactants were added to the solutions, the average size of the $Ni(DS)_2$ micelle in which both substrates were entrapped varied from ~10 nm to ~1 μm, whereas the overall micelle size of the $Ni(DS)_2$–SWNT composite remained in the nanometer range. After agitation for 24 hours, $Ni(DS)_2$ micelles formed aggregates. When allowed to stand for 30 days, $Ni(DS)_2$ furnished vesicles with diameters larger than 20 μm. In contrast, the small size of the micelles was retained after 24 hours, and micellar aggregates were evidently impeded even after 30 days, in the presence of SWNTs. Reaction profiles further indicated the longevity of the $Ni(DS)_2$–SWNT composite (fig. S16). When SWNTs were not used, $Ni(DS)_2$ micelles lost catalytic activity within 24 hours. The beneficial effect of the integration with SWNTs was not exerted without the prior isolation of $Ni(DS)_2$, even though NaCl or AgCl was removed from the catalyst solution (Table 1, entries 3 to 6). This result suggested that $Ni(DS)_2$ was responsible for the enhanced reactivity and selectivity on the basis of electronic coupling only when strongly bound onto the surface of SWNTs. In principle, once the SWNT surface was saturated with $Ni(DS)_2$ molecules, the remaining excess $Ni(DS)_2$ molecules could participate in the formation of micelles (*27*). The tightly bound state of the composites helps to rationalize the persistence of the dispersion, which was stable for 30 days. SWNT agglomerations with varying electronic properties arising from different chiralities

Department of Chemistry, School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
*Corresponding author. Email: shu_kobayashi@chem.s.u-tokyo.ac.jp

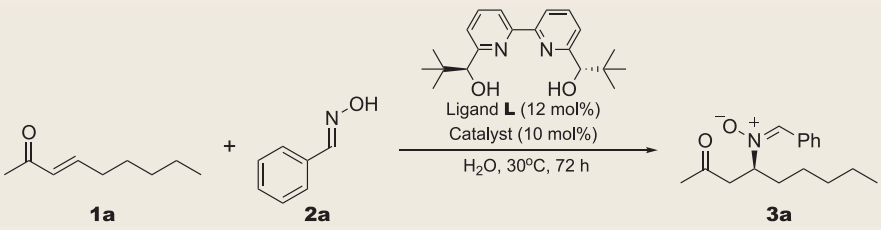
proved ineffective when wrapped with $Ni(DS)_2$, presumably because of weak electronic interactions between the nickel ion and SWNTs (Table 1, entries 7 and 8).

The effect of the integration with other π materials was surveyed. The reaction resulted in much smaller augmentation of the yield and/or selectivity when double-walled or multiwalled carbon nanotubes (DWNTs or MWNTs) were integrated with $Ni(DS)_2$ (Table 1, entries 9 and 10). This contrast is ascribed to the difference in electric properties among the carbon nanotubes (*28*). The ineffectiveness of fullerene and graphene nanoplatelets might be attributed to a poor interaction with surfactants because their spherical and two-dimensional shapes readily lead to coagulation (Table 1, entries 11 and 12). Aldoxime underwent nucleophilic addition, even in MeOH, CH_2Cl_2, and $CHCl_3$, to provide the nitrone, albeit in a racemic manner (Table 1, entries 13, 15, and 16). The reaction did not take place in other organic solvents (Table 1, entries 14 and 17 to 19). Consequently, substantially improved catalytic activity cannot be achieved without the uniform dispersion of SWNTs wrapped with $Ni(DS)_2$ in water. Scanning transmission electron microscopy (STEM) images of the $Ni(DS)_2$–SWNT composite revealed that SWNTs were well disentangled and dispersed (Fig. 2A). Energy-dispersive spectroscopy (EDS) mapping revealed that the $Ni(DS)_2$ molecules were attached to the single line of independent SWNTs. $Ni(DS)_2$ molecules were nevertheless revealed to be stably located on the surface to some extent, even when embedded on the SWNT ropes (fig. S13). In contrast, MWNTs failed to afford well-disentangled composites, and nickel was not attached on the surface of fullerene. These results suggested that $Ni(DS)_2$ molecules showed substantial catalytic performance only when strongly bound onto the surface of SWNTs (Fig. 2B).

A number of simulation studies have been used to understand the morphology of surfactant aggregates self-assembled on SWNTs. As mentioned, after saturation of the SWNT surface with $Ni(DS)_2$ molecules, the remaining $Ni(DS)_2$ molecules could participate in the formation of exterior micelles, providing effective nanotube-nanotube repulsion (*27*). The strongly adsorbed states are expected to exchange slowly or not at all with other states and to be undetectable in the nuclear magnetic resonance (NMR) measurements (Fig. 2C). In addition to the loss of mobility due to the presence of strongly bound states, the large surface area of SWNTs would help hydrophobic substrates and a ligand localize near the strongly bound $Ni(DS)_2$ molecules. The results of dynamic light scattering (DLS) analysis of the $Ni(DS)_2$–SWNT composite were consistent with the translational and rotational diffusion coefficients expressed for long, rigid, rodlike particles, corresponding to discrete SWNTs wrapped with $Ni(DS)_2$ (*29*) (Fig. 2D and fig. S4). We also ascertained that high-power ultrasonication was required for reliable deagglomeration and dispersion (fig. S4 and tables S3 and S4). Furthermore, NMR studies showed dispersity-dependent transitions and broadened resonances, corresponding to expected short T_2 relaxation times of strongly bound $Ni(DS)_2$ molecules (Fig. 2D). Although diffusion NMR measurements have usually probed SWNT–surfactant interactions (*30*), the paramagnetic nature of nickel broadened the signals considerably in our system. The interaction between metal d orbitals and π materials has been discussed mainly on the basis of calculations (*31*, *32*), ion detection (*33*), and chirality separation (*34*).

In this study, Raman, ultraviolet/visible/near-infrared (UV/vis/NIR), and photoluminescence excitation (PLE) analyses offered evidence for the electronic interaction between Ni^{2+} and semiconducting SWNTs. The redshifts of the SO_n vibration peaks in the Raman spectra in the presence of SWNTs are attributed to the increased electron density of the nickel ion (*35*) (Fig. 2E). Moreover, the redshifts of the C–C and C–H vibration peaks would result from lower vibrational frequencies due to orbital overlap with π electrons. When $Ni(OSO_2C_{12}H_{25})_2$ was used instead of $Ni(DS)_2$, C–S stretching vibrations were also found to be redshifted (fig. S1). These results can be explained by anchoring of the long alkyl chains on the surface of the SWNTs and consequent electronic interaction between Ni^{2+} and SWNTs (figs. S1 and S2). Owing to the sparing solubility of chiral 2,2′-bipyridine **L** in water, UV/vis spectra of a solution that was diluted enough below the critical micelle concentration were acquired to detect the $Ni(DS)_2$-**L** complex formed in water (Fig. 2F). The absorption peak at 314 nm (3.95 eV) is assigned to the π-π* transition of the 2,2′-bipyridine ring of the chiral nickel complex (*36*), which corresponds to a peak at 288 nm (4.31 eV) of pristine **L**. The SWNT-integrated catalyst showed the corresponding absorption at 319 nm (3.89 eV) together with two peaks of the π-π* interband transition of the SWNTs. The redshift of the π-π* transition of the 2,2′-bipyridine ring upon SWNT integration implies reinforced coordination of nickel(II) with **L** in this environment, supporting an orbital overlap between the nickel ion and the SWNTs. The Ni^{2+}-**L** complex was found to adopt high-spin octahedral geometry in water, whereas it formed trigonal bipyramidal structures in organic

Table 1. Evaluation of the catalytic activity of the designed SWNT-integrated catalysts in the enantioselective conjugate addition of benzaldoxime in water. Ee, enantiomeric excess; Me, methyl; Et, ethyl; Ph, phenyl; THF, tetrahydrofuran; –, not determined; NR, no reaction.



Entry	Catalyst	Solvent	Yield (%)	Ee (%)
1	$Ni(DS)_2$-SWNT	H_2O	79	95
2	$Ni(DS)_2$	H_2O	33	16
3	$NiCl_2$ + SDS-SWNT	H_2O	18	0
4*	$NiCl_2$ + SDS-SWNT	H_2O	32	0
5	$NiCl_2$ + AgDS-SWNT	H_2O	50	0
6†	$NiCl_2$ + AgDS-SWNT	H_2O	22	0
7	$Ni(DS)_2$-SWNT(aggregates)‡	H_2O	28	0
8	$Ni(DS)_2$ + SWNT§	H_2O	33	16
9	$Ni(DS)_2$-DWNT¶	H_2O	58	49
10	$Ni(DS)_2$-MWNT#	H_2O	39	21
11	$Ni(DS)_2$-C_{60}	H_2O	16	60
12	$Ni(DS)_2$-GNP**	H_2O	25	1
13	$Ni(DS)_2$-SWNT	MeOH	36	3
14	$Ni(DS)_2$-SWNT	EtOH	Trace	–
15	$Ni(DS)_2$-SWNT	CH_2Cl_2	39	0
16	$Ni(DS)_2$-SWNT	$CHCl_3$	20	1
17	$Ni(DS)_2$-SWNT	Et_2O	Trace	–
18	$Ni(DS)_2$-SWNT	THF	Trace	–
19	$Ni(DS)_2$-SWNT	Toluene	NR††	–

*After desalting through cellulose semipermeable membrane. †After centrifugation. ‡Prepared as a precipitate from $NiCl_2$, SDS, and SWNT without debundling in water. §SWNT was added as an additive without any treatment. ¶1.5- to 2.0-nm diameter DWNT. #20- to 40-nm diameter MWNT. **Graphene nanoplatelets of 6 to 8 nm thickness and 25 mm width. ††The reaction did not proceed at all, even when $Ni(OTf)_2$ (Tf, triflate) was used as a catalyst instead of $Ni(DS)_2$-SWNT composites.

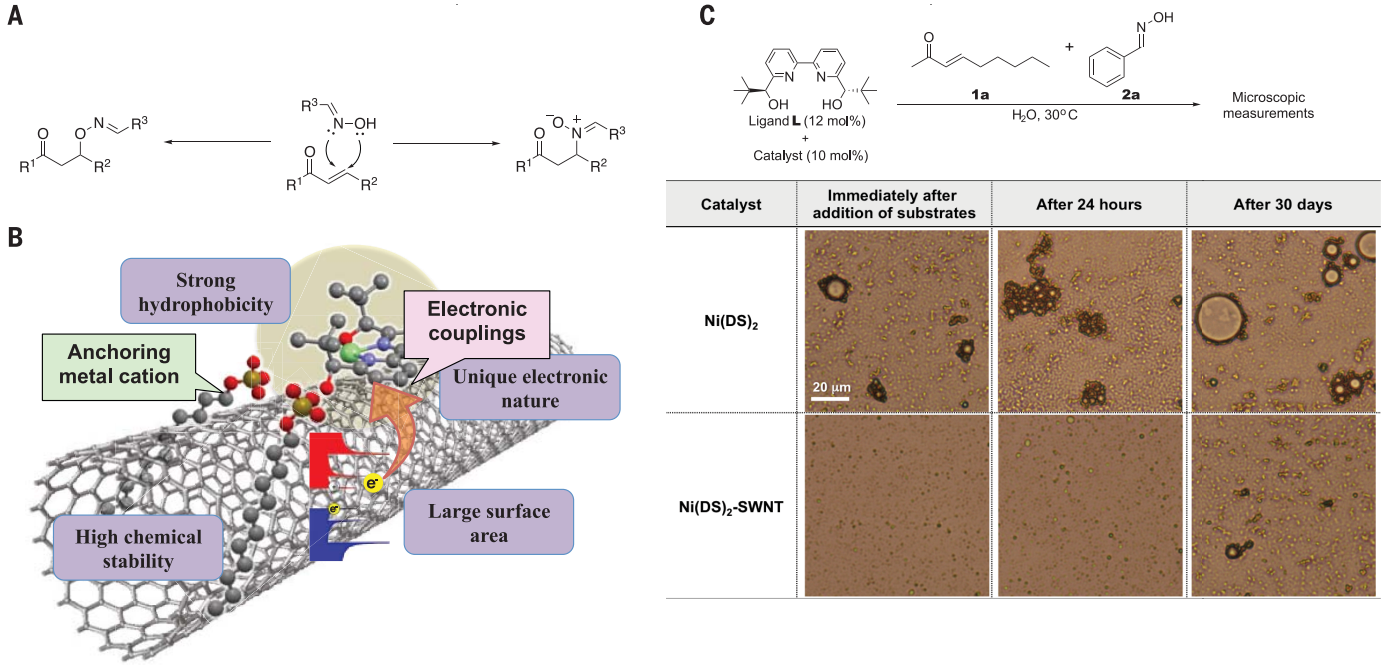


Fig. 1. Asymmetric 1,4-addition of aldoximes and catalyst design. (**A**) Two possible fates of the conjugate addition of aldoximes. R, alkyl or aromatic group. (**B**) Schematic image of Lewis acid–SWNT–integrated catalysts. e⁻, electron. (**C**) Microscopic images of $Ni(DS)_2$-catalyzed reaction mixtures with or without SWNTs.

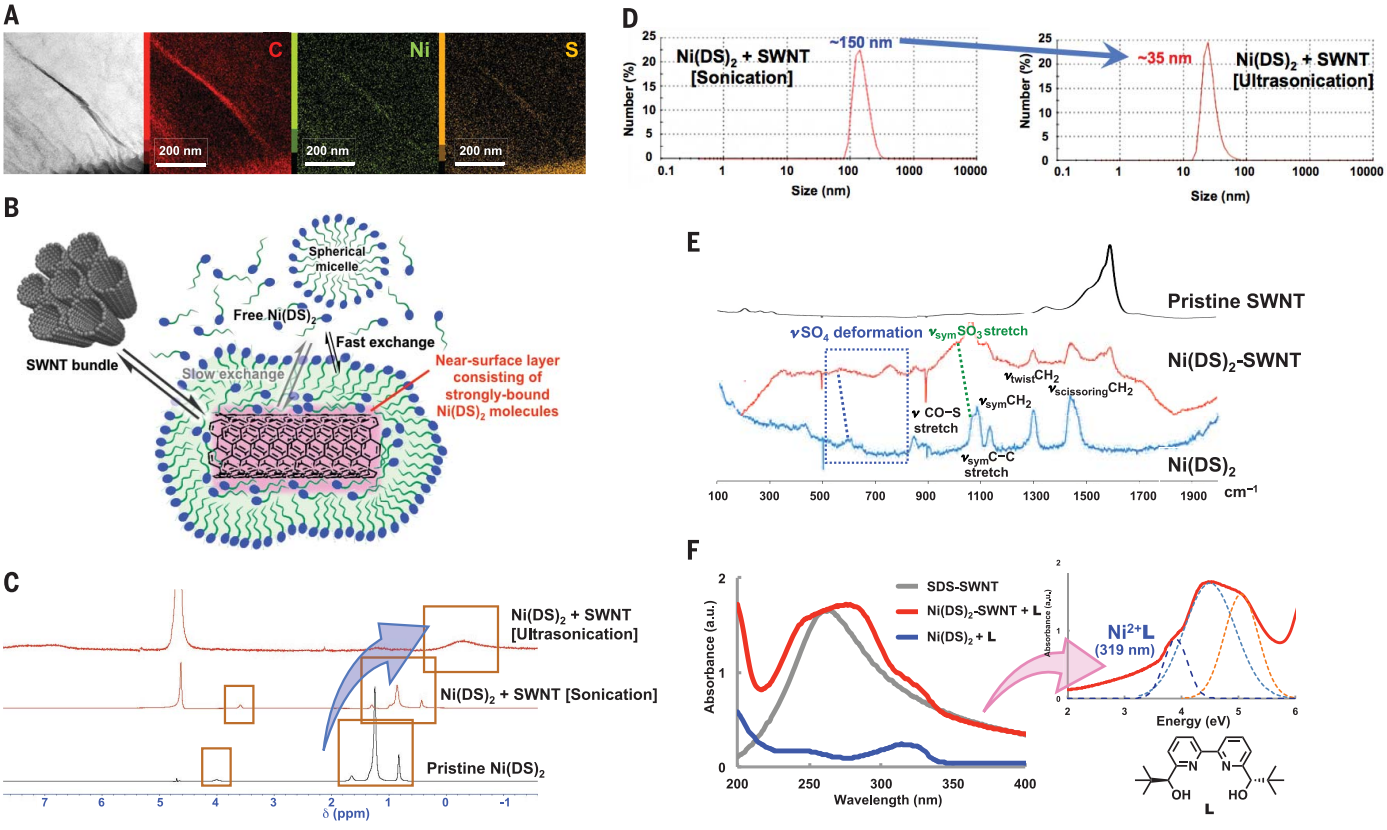


Fig. 2. Toward an activation mode of Lewis acid catalysis for asymmetric nitrone formation. (**A**) STEM images of the $Ni(DS)_2$–SWNT composite with elemental mapping. (**B**) Possible states of $Ni(DS)_2$ in the reaction mixture. The chiral ligand was omitted for simplicity. (**C**) NMR studies of the $Ni(DS)_2$–SWNT composite in deuterium oxide. δ, chemical shift; ppm, parts per million. (**D**) Size distribution of uniform dispersion of the $Ni(DS)_2$–SWNT composite. (**E**) Raman spectra of pristine SWNT (top), $Ni(DS)_2$–SWNT composite (middle), and $Ni(DS)_2$ (bottom). (**F**) UV/vis absorbance of the $Ni(DS)_2$–SWNT composite. a.u., arbitrary units.

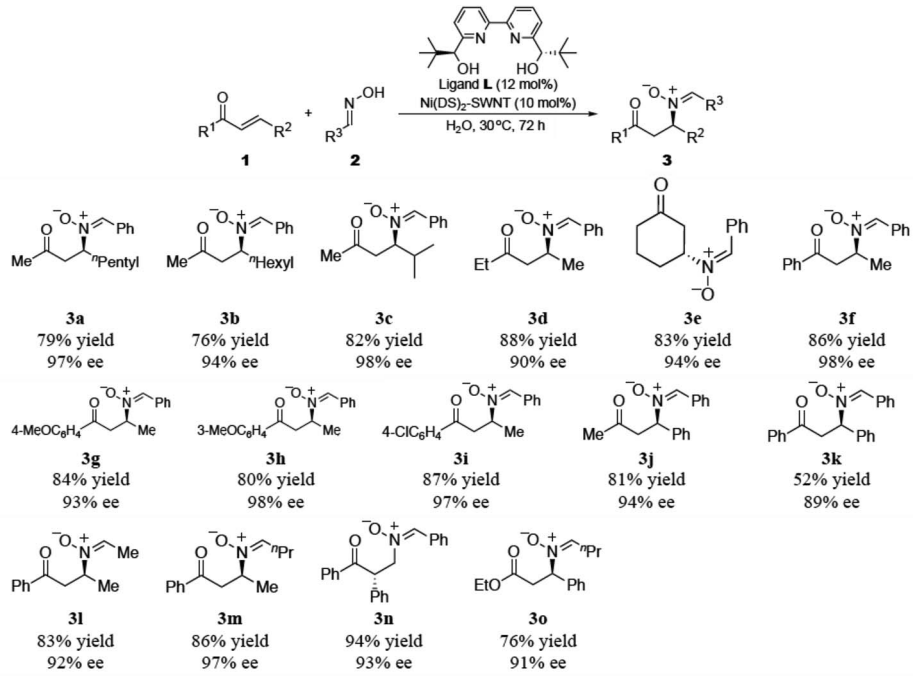


Fig. 3. Scope of the asymmetric 1,4-addition of aldoximes with electron-deficient olefins. Yields of isolated products are shown.

solvents. The calculated Racah parameter for six-coordinate SWNT-Ni^{2+}-**L** species in water was ~80 cm^{-1} smaller than that of the Ni^{2+}-**L** complex in water without SWNTs (figs. S5 and S6), which suggests the expansion of d orbitals due to mixing with the π orbitals of the ligand (SWNT) leading to delocalization. Besides the redshift of the π-π* transition of the 2,2′-bipyridine ring, this nephelauxetic effect by integration with SWNTs would provide a strong rationale for boosted catalytic performance of the Ni^{2+}-**L** complex. Although the UV/vis/NIR absorption spectroscopy has proven to be a powerful tool in SWNT characterization, it was found not to be suitable for elucidating the electronic properties of SWNTs, owing to their large diameters (figs. S8 and S9). When SWNTs with smaller diameter were employed to support the feasible interaction with Ni^{2+}, intensity attenuation of the semiconducting SWNT peaks (in particular, S_{11}) was recorded in UV/vis/NIR spectra (fig. S10), and the fluorescence of semiconducting SWNTs with diameters >1 nm was effectively quenched by $Ni(DS)_2$ molecules (in contrast with SDS molecules) in the PLE mapping (fig. S12). The fluorescence quenching by the $Ni(DS)_2$ molecules and the pronounced attenuation of the S_{11} peaks in the UV/vis/NIR spectra accord with our expectation of electronic couplings. Consequently, together with the destabilization of the Ni^{2+} e_g orbital by addition of SWNTs, these observations are consistent with the expected electron couplings between the semiconducting SWNTs and the chiral nickel complex.

With the optimal conditions established, the scope of the reaction with respect to the chiral nitrone formation process was examined (Fig. 3). Substrates that underwent highly enantioselective reactions encompassed both acyclic and cyclic α,β-unsaturated ketones with either aromatic or aliphatic substituents (**3a** to **3k**). The rate of the reaction was reduced for the diphenyl enone that produced **3k**. Aliphatic aldoximes were also effective substrates in the reaction (**3l** and **3m**). The enantiofacial differentiation during the protonation process was also found to be highly selective (**3n**). To our surprise, the reaction with the less electrophilic α,β-unsaturated ester underwent nucleophilic addition of the aldoxime to afford the desired product with high enantioselectivity (**3o**).

Considering recent developments in Lewis acid catalysis, which have enabled a series of asymmetric reactions, the application of this system to various Lewis acid–catalyzed reactions is expected to open up a variety of opportunities and distinct modes of catalysis.

REFERENCES AND NOTES

1. U. M. Lindrom, *Organic Reactions in Water: Principles, Strategies and Applications* (Wiley, 2007).
2. S. Kobayashi, T. Wakabayashi, *Tetrahedron Lett.* **39**, 5389–5392 (1998).
3. K. Manabe, Y. Mori, T. Wakabayashi, S. Nagayama, S. Kobayashi, *J. Am. Chem. Soc.* **122**, 7202–7207 (2000).
4. T. Kitanosono, S. Kobayashi, "Water-Compatible Chiral Lewis Acids" in *Chiral Lewis Acids in Organic Synthesis*, J. Mlynarski, Ed. (Wiley-VCH, 2017), pp. 299–344.
5. C. D. Vanderwal, E. N. Jacobsen, *J. Am. Chem. Soc.* **126**, 14724–14725 (2004).
6. S. Bertelsen, P. Dinér, R. L. Johansen, K. A. Jørgensen, *J. Am. Chem. Soc.* **129**, 1536–1537 (2007).
7. P. Melchiorre *et al.*, *Eur. J. Org. Chem.* 5492–5495 (2007).
8. K. Nakama, S. Seki, S. Kanemasa, *Tetrahedron Lett.* **43**, 829–832 (2002).
9. K. Nakama, S. Seki, S. Kanemasa, *Tetrahedron Lett.* **42**, 6719–6722 (2001).
10. K. Koçak, S. Malkondu, S. Kurbanli, *Russ. J. Org. Chem.* **45**, 591–595 (2009).
11. W. Oppolzer, J. Deerberg, O. Tamura, *Helv. Chim. Acta* **77**, 554–560 (1994).
12. S. Iijima, *Nature* **354**, 56–58 (1991).
13. R. J. White, R. Luque, V. L. Budarin, J. H. Clark, D. J. Macquarrie, *Chem. Soc. Rev.* **38**, 481–494 (2009).
14. Y. Zhang, N. W. Franklin, R. J. Chen, H. Dai, *Chem. Phys. Lett.* **331**, 35–41 (2000).
15. Y. Li *et al.*, *J. Phys. Chem. B* **105**, 11424–11431 (2001).
16. W.-Q. Han, A. Zettl, *Nano Lett.* **3**, 681–683 (2003).
17. P. M. Ajayan *et al.*, *J. Am. Chem. Soc.* **116**, 7935–7936 (1994).
18. E. T. Thostenson, Z. Ren, T.-W. Chou, *Compos. Sci. Technol.* **61**, 1899–1912 (2001).
19. D. Tasis, N. Tagmatarchis, A. Bianco, M. Prato, *Chem. Rev.* **106**, 1105–1136 (2006).
20. M. J. O'Connell *et al.*, *Science* **297**, 593–596 (2002).
21. V. C. Moore *et al.*, *Nano Lett.* **3**, 1379–1382 (2003).
22. H. Liu, D. Nishide, T. Tanaka, H. Kataura, *Nat. Commun.* **2**, 309 (2011).
23. A. J. Blanch, C. E. Lenehan, J. S. Quinton, *J. Phys. Chem. B* **114**, 9805–9811 (2010).
24. C. Bolm, M. Ewald, M. Felder, G. Schlingloff, *Chem. Ber.* **125**, 1169–1190 (1992).
25. S. Ishikawa, T. Hamada, K. Manabe, S. Kobayashi, *Synthesis* 2176–2182 (2005).
26. S. Kobayashi, S. Nagayama, T. Busujima, *J. Am. Chem. Soc.* **120**, 8287–8288 (1998).
27. J. Xu *et al.*, *Langmuir* **33**, 5006–5014 (2017).
28. M. Monajjemi, V. S. Lee, in *Electronic Properties of Carbon Nanotubes*, J. M. Marulanda, Ed. (InTech, 2011), pp. 575–602.
29. S. Badaire, P. Poulin, M. Maugey, C. Zakri, *Langmuir* **20**, 10367–10370 (2004).
30. T. A. Shastry, A. J. Morris-Cohen, E. A. Weiss, M. C. Hersam, *J. Am. Chem. Soc.* **135**, 6750–6753 (2013).
31. A. I. Frolov, A. G. Rozhin, M. V. Fedorov, *ChemPhysChem* **11**, 2612–2616 (2010).
32. S. Guerini, A. G. S. Filho, J. M. Filho, O. L. Alves, S. B. Fagan, *Phys. Rev. B* **72**, 233401 (2005).
33. W. Cheung *et al.*, *Chem. Sci* **7**, 5192–5199 (2016).
34. Y. Kato, Y. Niidome, N. Nakashima, *Angew. Chem. Int. Ed.* **48**, 5435–5438 (2009).
35. V. B. Kartha, L. C. Leitch, H. H. Mantsch, *Can. J. Chem.* **62**, 128–132 (1984).
36. R. Andrés, M. Gruselle, B. Malézieux, M. Verdaguer, J. Vaissermann, *Inorg. Chem.* **38**, 4637–4646 (1999).

ACKNOWLEDGMENTS

We thank T. Saito (AIST) for providing the SWNTs; T. Saito and M. Tange (AIST) for the UV/vis/NIR spectroscopy studies; M. Tange and T. Okazaki (AIST) for the PLE spectroscopy; S. Maruyama and S. Chiashi (University of Tokyo) for the Raman spectroscopy studies; E. Nakamura and K. Harano (University of Tokyo) for the DLS analyses; N. Kuramitsu and T. Maki (University of Tokyo) for the STEM, EDS, and TEM analyses; and Y. Saito (University of Tokyo) for the x-ray crystallography. **Funding:** This work was partially supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) under the Global COE Program, with support from the University of Tokyo; the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan; and the Japan Science Technology Agency (JST). **Author contributions:** T.K. and P.X. designed and performed the experiments. T.K. and S.K. conceived of and designed the investigations. S.K. directed the investigations and wrote the manuscript, with revisions provided by T.K. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Metrical parameters for the structure of **3g** are available from the Cambridge Crystallographic Data Centre under reference number CCDC-1858210.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/311/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S5
^1H and ^{13}C Spectra
References (*37*–*47*)

25 August 2017; resubmitted 27 May 2018
Accepted 22 August 2018
10.1126/science.aap7883

OPTICAL METAMATERIALS

Hierarchically porous polymer coatings for highly efficient passive daytime radiative cooling

Jyotirmoy Mandal[1], Yanke Fu[1], Adam C. Overvig[1], Mingxin Jia[2], Kerui Sun[1], Norman N. Shi[1], Hua Zhou[3,4], Xianghui Xiao[3,4], Nanfang Yu[1]*, Yuan Yang[1]*

Passive daytime radiative cooling (PDRC) involves spontaneously cooling a surface by reflecting sunlight and radiating heat to the cold outer space. Current PDRC designs are promising alternatives to electrical cooling but are either inefficient or have limited applicability. We present a simple, inexpensive, and scalable phase inversion–based method for fabricating hierarchically porous poly(vinylidene fluoride-co-hexafluoropropene) [P(VdF-HFP)$_{HP}$] coatings with excellent PDRC capability. High, substrate-independent hemispherical solar reflectances (0.96 ± 0.03) and long-wave infrared emittances (0.97 ± 0.02) allow for subambient temperature drops of ~6°C and cooling powers of ~96 watts per square meter (W m^{-2}) under solar intensities of 890 and 750 W m^{-2}, respectively. The performance equals or surpasses those of state-of-the-art PDRC designs, and the technique offers a paint-like simplicity.

Cooling human-made structures, such as buildings, is a widespread necessity faced by humans today (*1*). However, compression-based cooling systems that are prevalently used (e.g., air conditioners) consume substantial amounts of energy, have a net heating effect, require ready access to electricity, and often require coolants that are ozone-depleting or have a strong greenhouse effect (*2*, *3*). Therefore, inexpensive, eco-friendly approaches with net cooling capability are desirable for reducing energy costs, operation time, and associated ozone-depleting and CO_2 emissions from traditional cooling systems, or providing relief where electrical cooling is not available. One alternative to energy-intensive cooling methods is passive daytime radiative cooling (PDRC)—a phenomenon where a surface spontaneously cools by reflecting sunlight [wavelengths (λ) ~0.3 to 2.5 μm] and radiating heat to the cold outer space through the atmosphere's long-wave infrared (LWIR) transmission window (λ ~ 8 to 13 μm). PDRC is most effective if a surface has a high, hemispherical solar reflectance ($\bar{R}_{solar}$) that minimizes solar heat gain and a high, hemispherical, LWIR thermal emittance ($\bar{\epsilon}_{LWIR}$) that maximizes radiative heat loss (*4*). If $\bar{R}_{solar}$ and $\bar{\epsilon}_{LWIR}$ are sufficiently high, a net radiative heat loss can occur, even under sunlight. The passive nature of this effect makes PDRC highly appealing, as cooling occurs without the need for electricity, refrigerants, or mechanical pumps that require maintenance.

Research in recent decades has yielded a variety of PDRC designs comprising sophisticated emissive coatings such as photonic structures, dielectrics, polymers, and polymer-dielectric composites on metal mirrors (*5*–*11*). Although efficient, these designs are costly and susceptible to corrosion. Furthermore, unlike paints, they are prefabricated and cannot be directly applied to existing roofs or walls, which have various compositions, textures, and geometries (*7*, *9*, *10*). Therefore, cool-roof paints (CRPs), which combine a modest optical performance with easy applicability and inexpensiveness, remain the benchmark for PDRC (*12*–*15*). However, CRPs, which consist of dielectric pigments (e.g., titania and zinc oxide) embedded in a polymer matrix, have a low solar reflectance (typically ~0.85) due to the pigments' ultraviolet (UV) absorptance and the low near-to-short–wavelength infrared (NIR-to-SWIR, λ ~0.7 to 2.5 μm) reflectance (*13*). We realized that replacing the pigments in CRPs with light-scattering air voids can not only eliminate this problem and increase the optical performance to state-of-the-art levels, but also avoid the material, processing, and environmental costs associated with pigments (*14*, *16*). Inspired by this idea, we report a simple, scalable, and inexpensive phase inversion–based process for fabricating hierarchically porous polymer coatings that exhibit excellent $\bar{R}_{solar}$ and $\bar{\epsilon}_{LWIR}$. Specifically, we achieved substrate-independent hemispherical $\bar{R}_{solar} = 0.96 \pm 0.03$ and $\bar{\epsilon}_{LWIR} = 0.97 \pm 0.02$ with hierarchically porous poly(vinylidene fluoride-co-hexafluoropropene) [P(VdF-HFP)$_{HP}$]. The values result in a superb PDRC capability, exemplified by a subambient cooling of ~6°C and an average cooling power of ~96 W m^{-2} under solar intensities of 890 and 750 W m^{-2}, respectively. The performance is on par with or exceeds those in previous reports (*7*, *9*, *10*). Because the fabrication technique is room-temperature– and solution-based, the porous polymer coatings can be applied by conventional approaches like painting and spraying to diverse surfaces such as plastics, metal, and wood. Moreover, it can incorporate dyes to achieve a desirable balance between color and cooling performance. The performance of the coatings and the paint-like convenience of the technique make it promising as a viable way to achieve high-performance PDRC.

Our phase inversion–based method for making hierarchically porous polymers starts with the preparation of a precursor solution of P(VdF-HFP) (polymer) and water (nonsolvent) in acetone (solvent) (Fig. 1A). We apply a film onto a substrate and dry it in air. The rapid evaporation of the volatile acetone causes the P(VdF-HFP) to phase-separate from the water, which forms micro- and nanodroplets. The P(VdF-HFP)$_{HP}$ coating is formed (Fig. 1, A and B) after the water evaporates. The micro- and nanopores in the coating efficiently backscatter sunlight and enhance thermal emittance (Fig. 1C). Consequently, P(VdF-HFP)$_{HP}$ films with ~50% porosity and thickness $\gtrsim$300 μm have an exceptional, substrate-independent hemispherical $\bar{R}_{solar}$ of 0.96 and $\bar{\epsilon}_{LWIR}$ of 0.97 (Fig. 1, D to F). At thicknesses $\gtrsim$500 μm, $\bar{R}_{solar}$ $\gtrsim$0.98 is achieved (figs. S2 and S15). The high $\bar{R}_{solar}(\theta)$ ensures excellent reflection of sunlight from all incidences (Fig. 1E) and eliminates the need for silver reflectors used in previous designs (*7*, *9*, *10*), while the high $\bar{\epsilon}_{LWIR}(\theta)$ leads to a hemispherical $\bar{\epsilon}_{LWIR}$ that is >10% higher than previously reported values (Fig. 1F) (*7*, *9*). The precursor's paint-like applicability makes P(VdF-HFP)$_{HP}$ attractive for practical use.

P(VdF-HFP) (Fig. 2A) has ideal intrinsic electromagnetic properties for PDRC applications. First, it has a negligible extinction coefficient across the solar wavelengths (λ = 0.3 to 2.5 μm) (Fig. 2B), unlike dielectric pigments of paints (fig. S10) and silver, which both absorb UV light. This keeps solar heating to a minimum. Second, the polymer has multiple extinction peaks in the thermal wavelengths, including 14 in the LWIR window, which arise from the different vibrational modes of its molecular structure (Fig. 2B) (*17*–*19*). Consequently, P(VdF-HFP) efficiently radiates heat in the LWIR window, where peak blackbody emissions from terrestrial surfaces and a high atmospheric transmittance into space coincide.

When structured by the phase-inversion technique into a hierarchical form consisting of ~2- to 10-μm micropores partitioned by a nanoporous phase (Fig. 1B and fig. S4), P(VdF-HFP) exhibits high $\bar{R}_{solar}$ and $\bar{\epsilon}_{LWIR}$. Pore-size measurements indicate that the pore sizes are bimodally distributed, with broad distributions centered at ~0.2 and ~5.5 μm for the nano- and micropores, respectively (Fig. 2C) (*4*). As corroborated by finite-difference time-domain simulations (*4*), the abundant micropores with sizes ~5 μm efficiently scatter sunlight of all wavelengths (Fig. 2D). This is further enhanced by the nanopores with sizes ~50 to 500 nm, which strongly scatter shorter, visible wavelengths (Fig. 2D). The simulations are also experimentally substantiated by diffuse transmission characterizations, which yield a photon mean free path (l_f) of ~6 μm for the blue wavelengths and ~10 μm for the NIR wavelengths (fig. S3). In the absence of any intrinsic absorption,

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA. [2]Department of Mechanical Engineering, Columbia University, New York, NY 10027, USA. [3]Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA. [4]National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973, USA.
*Corresponding author. Email: yy2664@columbia.edu (Y.Y.); ny2214@columbia.edu (N.Y.)

this results in a high optical backscattering of sunlight and thus a matte, white appearance. Furthermore, the unoriented pores result in a high, diffuse $\bar{R}_{solar}(\theta)$ regardless of the angle of incidence (Fig. 1E). In the thermal wavelengths, the emittance is likely enhanced across the LWIR window by the broadening of the extinction peaks (Fig. 2B) due to impurities (e.g., moisture), polymer chain deformation, and amorphousness in the nanoporous polymer (*20*–*22*). We attribute the high $\bar{\epsilon}_{LWIR}(\theta)$ for a wide, angular range (Fig. 1F) to the open, porous surface (Fig. 1B) and the effective medium behavior of the nanoporous P(VdF-HFP)$_{HP}$ coating at large wavelengths (*4*). A combination of these two features provides a gradual refractive index transition across polymer-air boundaries. Therefore, emitted radiation is not hindered at the surface and $\bar{\epsilon}_{LWIR}(\theta)$ is high regardless of the angle (Fig. 2, E and F).

The high $\bar{R}_{solar}$ and $\bar{\epsilon}_{LWIR}$ allow P(VdF-HFP)$_{HP}$ coatings to achieve exceptional daytime subambient cooling under widely different skies of Phoenix (USA), New York (USA), and Chattogram (Bangladesh) (Fig. 3). For instance, under a peak solar intensity I_{solar} of ~890 W m^{-2} and a clear sky with low humidity in Phoenix, P(VdF-HFP)$_{HP}$ coatings without any convection shields achieved a subambient temperature drop (ΔT) of ~6°C. Promisingly, ΔT ~3°C was also observed in Chattogram, where fog and haze impeded radiative heat loss into the sky. P(VdF-HFP)$_{HP}$ coatings also attained exceptional cooling powers ($P_{cooling}$) (*4*) of 96 and 83 W m^{-2} in Phoenix and New York, respectively. The values are consistent with theoretical calculations (table S1) and indicate P(VdF-HFP)$_{HP}$'s potential to reduce air-conditioning costs of buildings. We cannot directly compare the performance to earlier results, as $P_{cooling}$ depends heavily on experimental design, geography, and meteorological variables (table S2) (*23*–*25*). Nevertheless, the high performance without convection shields in different climates indicate that P(VdF-HFP)$_{HP}$'s PDRC capability is better than or on par with high-performance PDRC results in the literature (*7*, *9*, *10*).

The excellent optical performance of P(VdF-HFP)$_{HP}$ is complemented by a paint-like applicability, which is crucial for direct application on structures. We can paint, dip-coat, or spray P(VdF-HFP)$_{HP}$ onto diverse substrates like metal, plastics, and wood (Fig. 4, A to C). Furthermore, P(VdF-HFP)$_{HP}$ can also be made into strong, recyclable sheets (Fig. 4D and figs. S12 to S13). We also conducted accelerated thermal aging and moisture testing that showed the durability of the coatings and sheets (table S3). P(VdF-HFP) is intrinsically resistant to weathering, fouling, and UV radiation (*21*, *26*). When made porous, it hydrophobically repels waterborne dirt (fig. S14). During accelerated aging and monthlong outdoor exposure tests, these properties enabled P(VdF-HFP)$_{HP}$ to retain its optical performance at near-pristine levels (table S3 and fig. S15). For instance, a monthlong outdoor exposure in New York City only changed $\bar{R}_{solar}/\bar{\epsilon}_{LWIR}$ from 0.94/0.93 to 0.93/0.93.

An often-unstated but important practical requirement for PDRC coatings is compatibility with colors. To minimize solar heating, colored PDRC coatings should maximize the reflection of NIR-to-SWIR wavelengths (0.7 to 2.5 μm), which contain ~51% of solar energy but are invisible to the human eye. However, paints typically have low reflectances ($\bar{R}_{NIR-SWIR}$) in the NIR-to-SWIR wavelengths (fig. S10) (*13*, *27*). By contrast, porous P(VdF-HFP)$_{HP}$ coatings containing blue, yellow, and black dyes and with thickness ~350 μm efficiently backscatter sunlight not absorbed by the dyes to yield correspondingly colored coatings with high $\bar{R}_{NIR-SWIR}$ of 0.73, 0.89, and 0.62, respectively (Fig. 4, E and F). These values were measured with black substrates but exceed reflectances of thin films (~25 μm) of similarly colored "IR-reflective" pigments on reflective substrates (Fig. 4, F and G) and of the same pigments on black substrates by a large margin (*27*, *28*). Dyed P(VdF-HFP)$_{HP}$ coatings may thus address the long-standing problem of achieving PDRC with colored coatings, greatly widening their scope of use.

The phase inversion–based technique, shown here for P(VdF-HFP)$_{HP}$, is compatible with a wide variety of polymers. The method thus allows optically suitable polymers to be easily structured into PDRC coatings with other potential benefits. For instance, poly(methyl methacrylate) yields glossy coatings, ethyl cellulose provides biocompatibility and enables use of eco-friendly solvents, and polystyrene enables operation at temperatures >200°C (fig. S16). The diverse possibilities makes

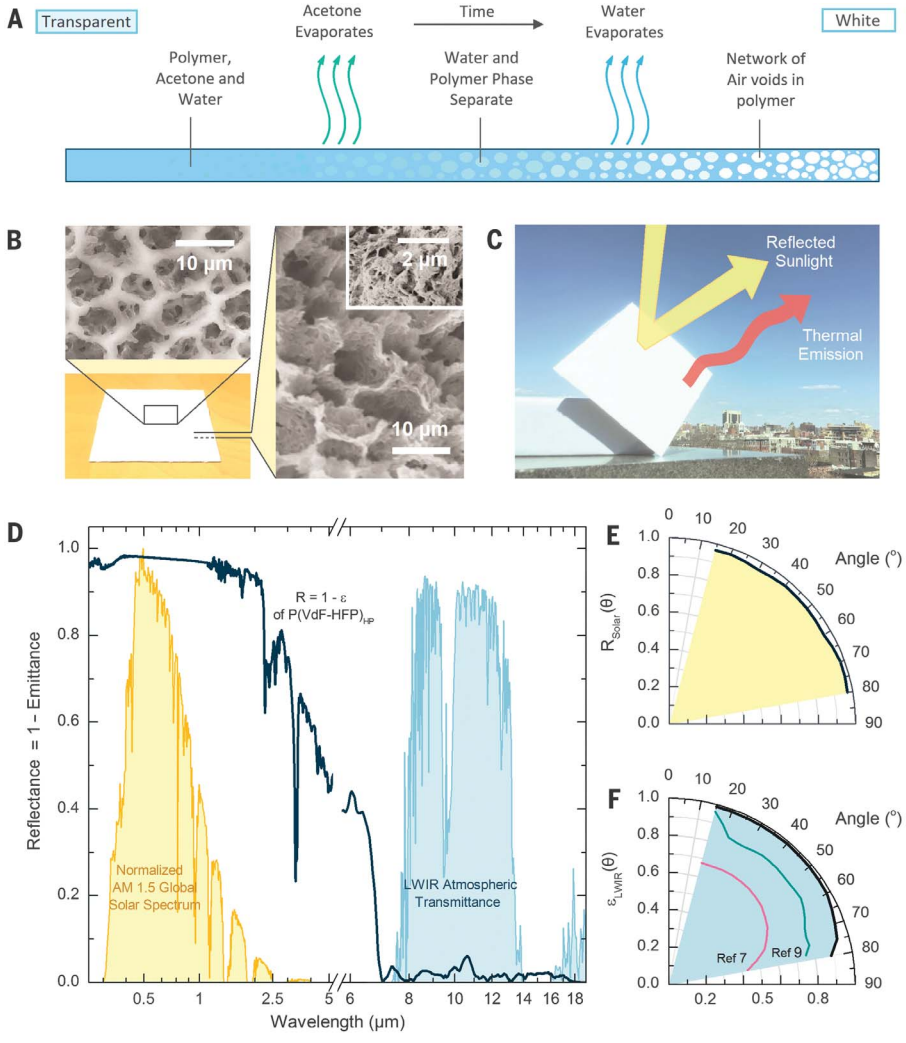


Fig. 1. The formation and optical properties of P(VdF-HFP)$_{HP}$. (**A**) Schematic of the phase inversion process, showing the formation of a hierarchically porous polymer coating from a solution of acetone (solvent), water (nonsolvent), and P(VdF-HFP) (polymer). (**B**) Micrographs showing top and cross-section views of P(VdF-HFP)$_{HP}$. Inset shows the nanoporous features. (**C**) Photograph superimposed with schematics to show that high $\bar{R}_{solar}$ and $\bar{\epsilon}_{LWIR}$ enable a net radiative loss and PDRC. (**D**) Spectral reflectance [$R(\lambda) = 1 - \epsilon(\lambda)$] of a 300-μm-thick P(VdF-HFP)$_{HP}$ coating presented against normalized ASTM G173 Global solar spectrum and the LWIR atmospheric transparency window. $\bar{R}_{solar}$ (0.96) and $\bar{\epsilon}_{LWIR}$ (0.97) are exceptionally high, especially given that they are achieved on a black selective solar absorber (fig. S2) (*29*). (**E**) P(VdF-HFP)$_{HP}$'s high $\bar{R}_{solar}(\theta)$ and (**F**) $\bar{\epsilon}_{LWIR}(\theta)$ across angles result in excellent hemispherical $\bar{R}_{solar}$ and $\bar{\epsilon}_{LWIR}$.

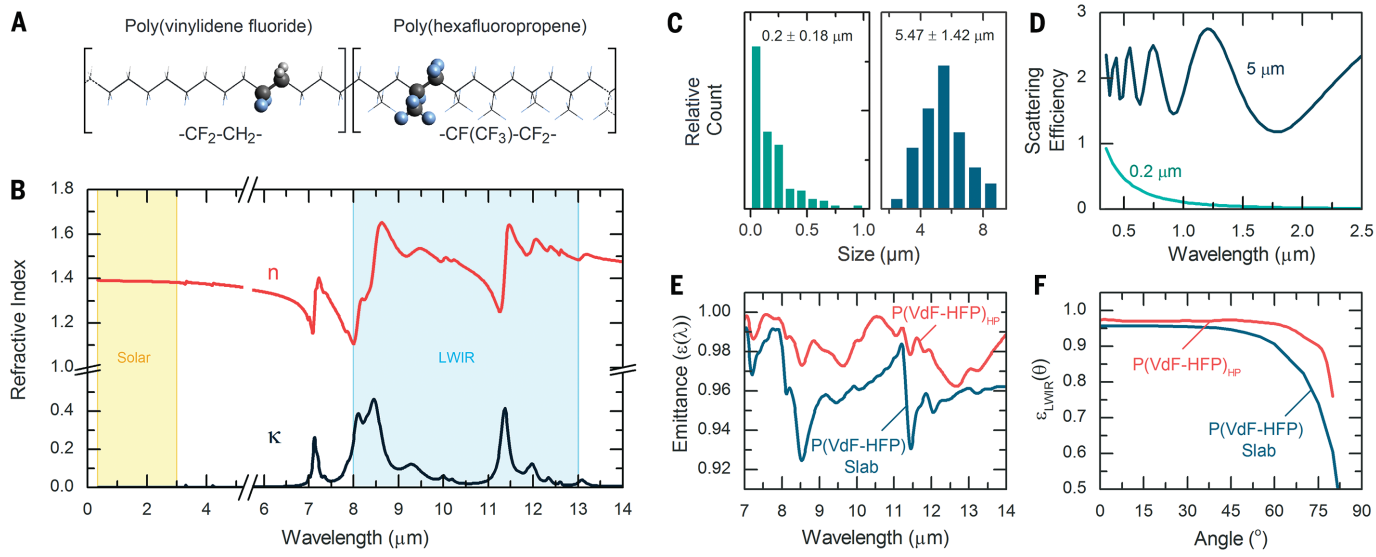


Fig. 2. The optical properties of P(VdF-HFP)$_{HP}$. (**A**) A wireframe showing the structure of P(VdF-HFP), with the VdF and HFP repeating units shown. (**B**) Experimental complex spectral refractive index ($n + i\kappa$) of P(VdF-HFP), showing negligible absorptivity in the solar, and high emissivity in the LWIR, wavelengths. The peaks in κ correspond to the vibrational modes of different molecular components (e.g., CF_3, CF_2, CF, C-C, CH_2, C-H, and carbon backbone) (*17*–*19*). (**C**) Size distributions of nano- and micropores in P(VdF-HFP)$_{HP}$ showing number-weighted mean pore sizes of ~0.2 μm for nanopores and ~5.5 μm for micropores. (**D**) Simulated scattering cross-section spectra of circular micro- and nanovoids in P(VdF-HFP)$_{HP}$. Voids of different sizes collectively scatter all solar wavelengths, resulting in a high $\bar{R}_{solar}$. (**E**) Spectral LWIR emittance and (**F**) $\bar{\epsilon}_{LWIR}(\theta)$ of P(VdF-HFP)$_{HP}$ compared to a solid P(VdF-HFP) slab of the same volume. As evident, the former has a higher spectral and angular emittance. Further details are provided in the supplementary materials (*4*).

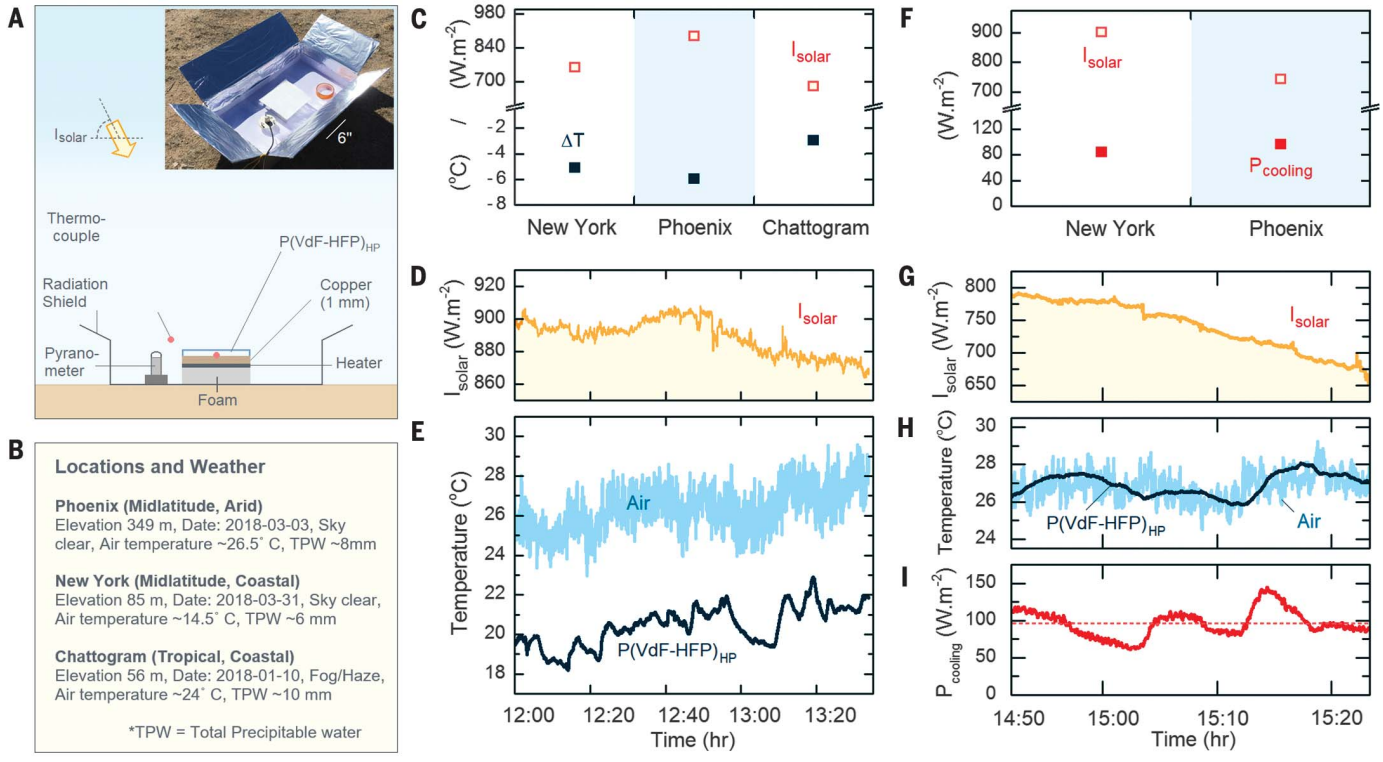


Fig. 3. Passive daytime radiative cooling performance of P(VdF-HFP)$_{HP}$. (**A**) Schematic of the setup for testing performance under sunlight. (**B**) Topographic and meteorological information of the test locations. (**C**) Average solar intensity (I_{solar}) and subambient temperature drops (ΔT) of P(VdF-HFP)$_{HP}$ coatings in New York, Phoenix, and Chattogram. (**D**) Detailed I_{solar} and (**E**) temperature data of the result for Phoenix in (C). (**F**) I_{solar} and cooling powers ($P_{cooling}$) of P(VdF-HFP)$_{HP}$ coatings measured in New York and Phoenix. (**G**) Detailed I_{solar}, (**H**) temperature tracking, and (**I**) $P_{cooling}$ data of the result for Phoenix in (F). Dotted line in (I) indicates average $P_{cooling}$ over the duration of the experiment. Additional information is provided in the supplementary materials (*4*).

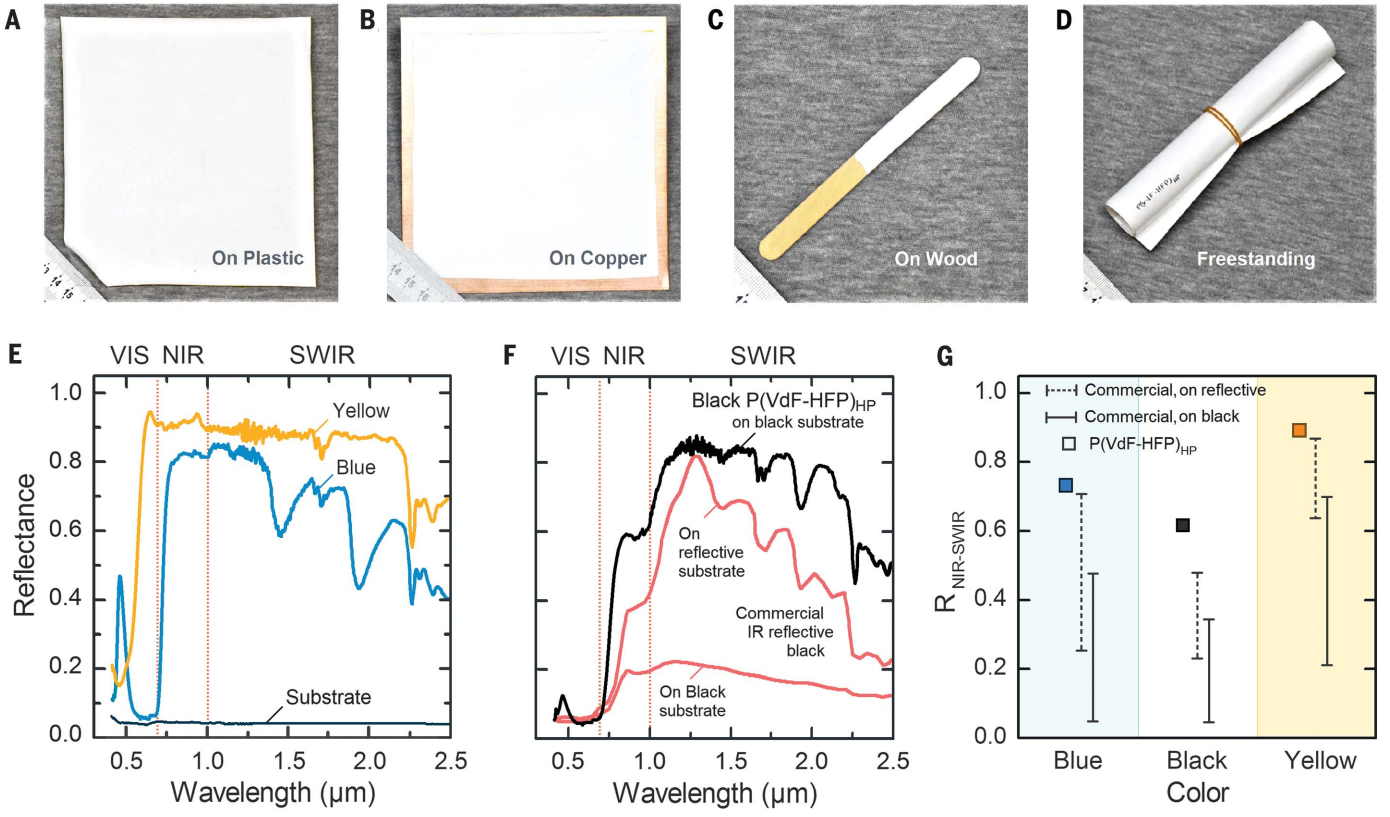


Fig. 4. Versatility of P(VdF-HFP)$_{HP}$ coatings. P(VdF-HFP)$_{HP}$ can be (**A**) painted onto plastics, (**B**) spray-coated on copper, (**C**) dip-coated on wood, and (**D**) made into strong, flexible, and freestanding sheets for tarpaulin-like designs. (**E**) Spectral reflectances of ~350-μm-thick blue and yellow P(VdF-HFP)$_{HP}$ coatings and (**F**) of a black P(VdF-HFP)$_{HP}$ coating compared to a commercial black pigment on reflective and black substrates. (**G**) Despite being on black substrates, their $\bar{R}_{NIR-SWIR}$ surpasses those of similarly colored "IR-reflective" pigments (~25-μm-thick films) on both black and reflective substrates.

the phase inversion–based technique a viable pathway for making both generic and specialized PDRC coatings.

REFERENCES AND NOTES

1. M. Burke, S. M. Hsiang, E. Miguel, *Nature* **527**, 235–239 (2015).
2. World Bank, (2017). *Access to Electricity*. Retrieved from https://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS.
3. United States Environment Protection Agency, (2018). *Phaseout of Ozone-Depleting Substances (ODS)*. Retrieved from https://www.epa.gov/ods-phaseout.
4. Materials and methods are available in the supplementary materials.
5. T. S. Eriksson, S. J. Jiang, C. G. Granqvist, *Sol. Energy Mater.* **12**, 319–325 (1985).
6. C. G. Granqvist, A. Hjortsberg, *Appl. Phys. Lett.* **36**, 139–141 (1980).
7. A. P. Raman, M. A. Anoma, L. Zhu, E. Rephaeli, S. Fan, *Nature* **515**, 540–544 (2014).
8. A. R. Gentle, G. B. Smith, *Nano Lett.* **10**, 373–379 (2010).
9. A. R. Gentle, G. B. Smith, *Adv. Sci.* **2**, 1500119 (2015).
10. Y. Zhai *et al.*, *Science* **355**, 1062–1066 (2017).
11. N. Yu, J. Mandal, A. C. Overvig, N. N. Shi, Patent WO 2016205717A1 (2016).
12. C. S. Wojtysiak, J. W. Butler, Patent PCT/AU02/00695 (2009).
13. J. Song *et al.*, *Sol. Energy Mater. Sol. Cells* **130**, 42–50 (2014).
14. G. T. J. Bentley, *Introduction to Paint Chemistry* (Springer Science, Bristol, UK, ed. 4, 1998).
15. Cool Roof Rating Council, *Rated Products Directory*. Retrieved from https://coolroofs.org/directory.
16. A. Al-Kattan *et al.*, *Environ. Sci. Technol.* **48**, 6710–6718 (2014).
17. Y. Bormashenko, R. Pogreb, O. Stanevsky, E. Bormashenko, *Polym. Test.* **23**, 791–796 (2004).
18. M. Kobayashi, K. Tashiro, H. Tadokoro, *Macromolecules* **8**, 158–171 (1975).
19. L. N. Sim, S. R. Majid, A. K. Arof, *Vib. Spectrosc.* **58**, 57–66 (2012).
20. H. Hagemann, R. G. Snyder, A. J. Peacock, L. Mandelkern, *Macromolecules* **22**, 3600–3606 (1989).
21. Arkema (2017). *KYNAR® & KYNAR FLEX® PVDF Performance Characteristics and Data*. Retrieved from www.extremematerials-arkema.com/en/product-families/kynar-pvdf-family/download-performance-characteristics-data-brochure.
22. R. S. Bretzlaff, R. P. Wool, *Macromolecules* **16**, 1907–1917 (1983).
23. T. S. Eriksson, C. G. Granqvist, *Appl. Opt.* **21**, 4381–4388 (1982).
24. A. W. Harrison, M. R. Walton, *Sol. Energy* **20**, 185–188 (1978).
25. C. J. Chen, in *Physics of Solar Energy* (Wiley, 2011), pp. 77–104.
26. Arkema (2008). *New fluoropolymer latex technology for cool materials solutions across an expanded color space*. Retrieved from http://coolcolors.lbl.gov/assets/docs/PAC-2008-03-06/Arkema-slides.pdf.
27. BASF, (2017) *Ultra-cool - The new heat reflective coatings from BASF*. Retrieved from www2.basf.us/pdfs/ULTRA-Cool.pdf.
28. R. Levinson, P. Berdahl, H. Akbari, *Sol. Energy Mater. Sol. Cells* **89**, 351–389 (2005).
29. J. Mandal *et al.*, *Adv. Mater.* **29**, 201702156 (2017).

ACKNOWLEDGMENTS

We thank K. K. Mandal and A. Krejci for their help on this study. We thank G. Chen and Y. Huang at MIT, A. Sobel at Columbia University, X. B. Zuo and J. Ilavsky at Argonne National Laboratory for help in measurements and discussions, and Arkema for providing P(VdF-HFP) samples. Research was carried out in part at the Center for Functional Nanomaterials in Brookhaven National Laboratory and the Photonics Spectroscopy Facility in CUNY Advanced Science Research Center. **Funding:** This work was supported by startup funding from Columbia University, the NSF MRSEC program through Columbia University's Center for Precision Assembly of Superstratic and Superatomic Solids (Y.Y. DMR-1420634), AFOSR (Y.Y. grant no. FA9550-18-1-0410), AFOSR MURI (Multidisciplinary University Research Initiative) program (N.Y. grant no. FA9550-14-1-0389), AFOSR DURIP (Defense University Research Instrumentation Program) (N.Y. grant no. FA9550-16-1-0322), and the National Science Foundation (N.Y. grant no. ECCS-1307948). A.C.O. acknowledges support from the NSF IGERT program (no. DGE-1069240). We acknowledge support from the Advanced Photon Source in Argonne National Laboratory (under Contract no. DE-AC02-06CH11357). This research used resources of the Full-Field X-ray Imaging Beamline (FXI) at 18-ID of the National Synchrotron Light Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under contract no. DE-AC02-98CH10886. **Author contributions:** J.M., Y.Y., and N.Y. conceived the concept and designed experiments. All authors contributed to experiments and data analysis. A.C.O. and J.M. performed the simulations. J.M., A.C.O., Y.Y., and N.Y. wrote the manuscript. **Competing interests:** A patent (PCT/US2016/038190) has been granted related to this work. A provisional patent (U.S. 62/596,145) has been filed related to this work. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Information requests should be directed to the corresponding authors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/315/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S3
References (*30*–*36*)

23 May 2018; accepted 28 August 2018
Published online 27 September 2018
10.1126/science.aat9513

CHEMICAL SENSING

Aptamer–field-effect transistors overcome Debye length limitations for small-molecule sensing

Nako Nakatsuka[1,2], Kyung-Ae Yang[3], John M. Abendroth[1,2], Kevin M. Cheung[1,2], Xiaobin Xu[1,2], Hongyan Yang[4], Chuanzhen Zhao[1,2], Bowen Zhu[1,5], You Seung Rim[1,5]*, Yang Yang[1,5], Paul S. Weiss[1,2,5]†, Milan N. Stojanović[3,6]†, Anne M. Andrews[1,2,4]†

Detection of analytes by means of field-effect transistors bearing ligand-specific receptors is fundamentally limited by the shielding created by the electrical double layer (the "Debye length" limitation). We detected small molecules under physiological high–ionic strength conditions by modifying printed ultrathin metal-oxide field-effect transistor arrays with deoxyribonucleotide aptamers selected to bind their targets adaptively. Target-induced conformational changes of negatively charged aptamer phosphodiester backbones in close proximity to semiconductor channels gated conductance in physiological buffers, resulting in highly sensitive detection. Sensing of charged and electroneutral targets (serotonin, dopamine, glucose, and sphingosine-1-phosphate) was enabled by specifically isolated aptameric stem-loop receptors.

Field-effect transistors (FETs) modified with target-specific receptors could enable direct electronic target detection (*1*, *2*). Signal transduction and amplification in FET-based sensors is based on electrostatic gating of thin-film semiconductor channels by target-receptor interactions, such that even low receptor occupancy measurably affects transconductance (*3*). However, receptor-modified FETs must overcome two fundamental limitations to become more widely adopted: (i) In solutions containing ions, the electrical double layer shields semiconductor charge carriers to limit gating in response to recognition events. The extent of shielding (i.e., the effective sensing distance) is characterized by the Debye length, which in physiological fluids is <1 nm (table S1) (*4*). (ii) Small target molecules with few or no charges have minimal impact on semiconductor transconductance unless they trigger conformational changes in charged receptors within or near the Debye length, or otherwise affect surface potentials (*5*).

We overcame both of these obstacles by combining highly sensitive FETs with a specific type of oligonucleotide stem-loop receptor selected for adaptive target recognition (Fig. 1A). We fabricated nanometer-thin In_2O_3 FETs (Fig. 1B) using methods that facilitate micro- and nanoscale patterning and are readily scalable for producing large numbers of devices (*2*, *6*). Sensing with FETs is inherently nonlinear (*5*), which enables target detection over larger and lower concentration ranges than can be achieved with equilibrium-based sensors (*7*). Although aptamers have been used as receptors for FET devices (*2*, *8*), it proved critical to combine ligand-induced stem-loop conformational rearrangements and close proximity to the surfaces of quasi–two-dimensional FETs. Changes in conformation of negatively charged phosphodiester backbones enabled signal transduction and amplification under biologically relevant conditions with low-charge and neutral targets.

Solution-phase selection of aptamers circumvented tethering of small-molecule targets and was based on stem-loop closing with appropriate counterselection against interferents (Fig. 1C) (*9*, *10*). This approach yielded aptamers characterized by adaptive-loop binding. Strategies and details of the selections and counterselections are given in fig. S1 and tables S2 and S3. We isolated original receptors for dopamine, serotonin, glucose, and sphingosine-1-phosphate (S1P) (Fig. 1, D to G, and table S4). Dopamine was targeted because we had constructed FET devices using a previously reported dopamine aptamer (*11*), but these required dilute ion concentrations for sensing (*2*). Serotonin was pursued as another important neurotransmitter target (*12*) having no reported aptamer sequences. Ultimately, we aim to distinguish serotonin from dopamine and other similarly structured molecules in measurements of interneuronal signaling (*13*–*15*). Glucose was selected as an example of an important neutral target. Aptamers interacting directly with glucose have not been reported (although compare with aptamers for glucose sensors) (*10*). The lipid S1P (critical micellar concentration <10 μM), which prevents chemotherapy-associated apoptosis (*16*), was chosen as an example of a zwitterionic target.

Fluorescence assays were used to characterize aptamer-target dissociation constants (K_d) (fig. S2A). Selection led to high-affinity aptamers for dopamine (150 nM) and serotonin (30 nM) (fig. S2, B and C). Counterselection minimized interactions with other neurotransmitters and metabolites (Fig. 1, H and I) critical for sensing in the presence of high concentrations of similarly structured countertargets in vivo. Notably, our dopamine aptamer did not recognize norepinephrine, in contrast to cross-reactivity of a previously reported dopamine aptamer (*2*, *11*). Poor selectivity has also been problematic for fast-scan cyclic voltammetry, the most common method for sensing dopamine (*13*, *17*). The affinity of the glucose aptamer (~10 mM) (fig. S2D) and selectivity with respect to analogs (Fig. 1J and fig. S3) were consistent with the receptor recognizing hydrophobic surfaces of glucose (*18*). The affinity of the S1P aptamer was 180 nM (Fig. 1K and fig. S2E), which was not as high as a reported mirror-image aptamer (4 nM) (*19*).

We covalently modified thin-film In_2O_3 FETs with dopamine or serotonin aptamers using silane chemistry (fig. S4) to investigate electronic small-molecule detection (Fig. 1A). Despite subnanometer Debye screening lengths, aptamer-FETs responded to wide ranges of target concentrations (10^{-14} to 10^{-9} M) in undiluted (i.e., physiological) phosphate-buffered saline (1× PBS; Fig. 2A and fig. S5A) or artificial cerebrospinal fluid (aCSF; Fig. 2, B and C) with response times on the order of seconds (fig. S6). Scrambled aptamer sequences (table S5) produced negligible responses (Fig. 2, A to C, and fig. S5A), as did FETs lacking aptamers (fig. S5, B and C). Even at physiological ion concentrations (and hence substantially reduced Debye lengths), FET responses for our dopamine aptamer were more than three orders of magnitude greater than those of the previously reported dopamine aptamer (*11*) in 0.1× PBS (Fig. 2A) because of the designed positioning of recognition regions capable of adaptive conformational changes in the new aptamer.

Dopamine aptamer–FETs were selective for dopamine versus serotonin, norepinephrine, tyramine, and dopamine metabolites (Fig. 2D and fig. S7A). Serotonin aptamer–FETs were selective for serotonin versus dopamine, norepinephrine, histamine, other biogenic amines, and indole metabolites (Fig. 2E and fig. S7B). Aptamer-FET selectivity was further investigated with surface-enhanced Raman spectroscopy (SERS; fig. S8, A and B). Raman signatures were enhanced only in close proximity to metal surfaces because of the short range of evanescent fields, with the strongest enhancement within ~1 nm of surfaces (similar to the physiological Debye length) (*20*). After dopamine or serotonin were introduced, SERS spectra exhibited complex pattern changes that were not evident with nontarget compounds (fig. S8, C and D).

Concentration sensitivity ranges could be "tuned" by altering the numbers of serotonin aptamers on FET surfaces (Fig. 2F). To evaluate sensing in an undiluted biological matrix, we

[1]California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA. [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. [3]Center for Innovative Diagnostic and Therapeutic Approaches, Department of Medicine, Columbia University, New York, NY 10032, USA. [4]Department of Psychiatry and Biobehavioral Science, Semel Institute for Neuroscience and Human Behavior, and Hatos Center for Neuropharmacology, University of California, Los Angeles, CA 90095, USA. [5]Department of Materials Science and Engineering, University of California, Los Angeles, CA 90095, USA. [6]Departments of Biomedical Engineering and Systems Biology, Columbia University, New York, NY 10032, USA.
*Present address: School of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Republic of Korea.
†Corresponding author. Email: aandrews@mednet.ucla.edu (A.M.A.); mns18@cumc.columbia.edu (M.N.S.); psw@cnsi.ucla.edu (P.S.W.)

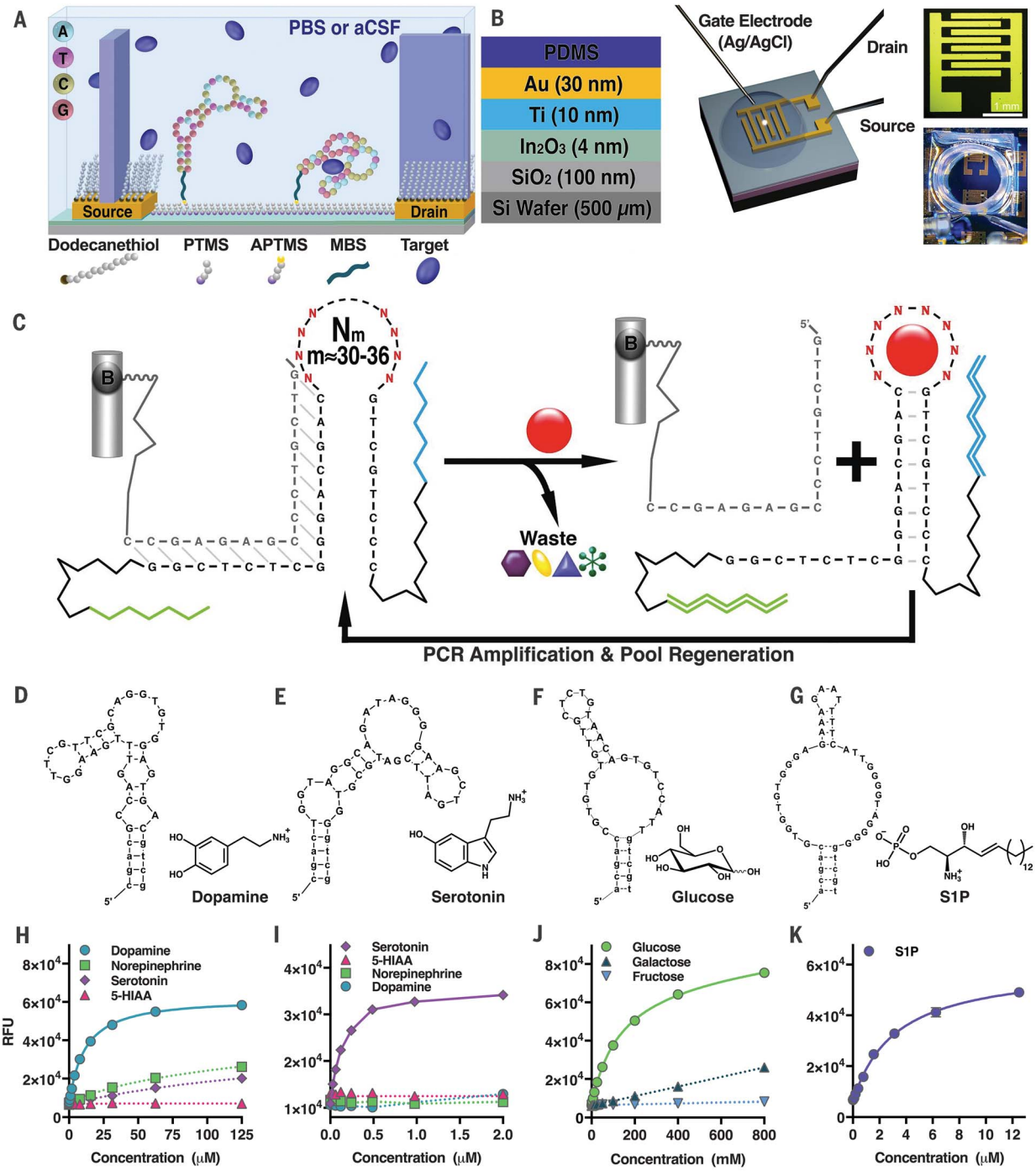


Fig. 1. Isolation of stem-loop aptamer receptors. (**A**) Schematic of FET surface chemistry. PTMS, trimethoxy(propyl)silane; APTMS, (3-aminopropyl)trimethoxysilane; MBS, 3-maleimidobenzoic acid *N*-hydroxysuccinimide ester. (**B**) Layer-by-layer composition of FETs, FET microscope image, and photograph of experimental setup. PDMS, polydimethylsiloxane. (**C**) Oligonucleotide libraries (N_m, with random regions m from 30 to 36 nucleotides, flanked by constant regions and oligonucleotide primer regions for polymerase chain reaction amplification) were attached to agarose-streptavidin columns via biotinylated (B) complementary sequences. Exposure to targets (red sphere) causes elution of aptamers in which stems are stabilized. These sequences are preferentially amplified. Exposure to countertargets (alternative shapes) eliminates cross-reactive sequences. (**D** to **K**) Aptamers for dopamine (K_d = 150 nM) (D), serotonin (K_d = 30 nM) (E), glucose (K_d = 10 mM) (F), and S1P (K_d = 180 nM) (G) were isolated. Solution-phase SELEX (i.e., systematic evolution of ligands by exponential enrichment) was used to identify aptamers that were directly converted to sensors. The complementary oligonucleotide was labeled with a quencher instead of biotin, whereas the aptamer was labeled with a fluorophore (table S4), leading to adaptive binding sensors with responses shown in (H) to (K). Fluorescence responses indicate selectivities of dopamine, serotonin, and glucose aptamers in the presence of specific versus nonspecific targets. Fluorescence-concentration curves (RFU, relative fluorescence units) were the result of *N* = 3 measurements with SEMs too small to be visualized in the graphs shown.

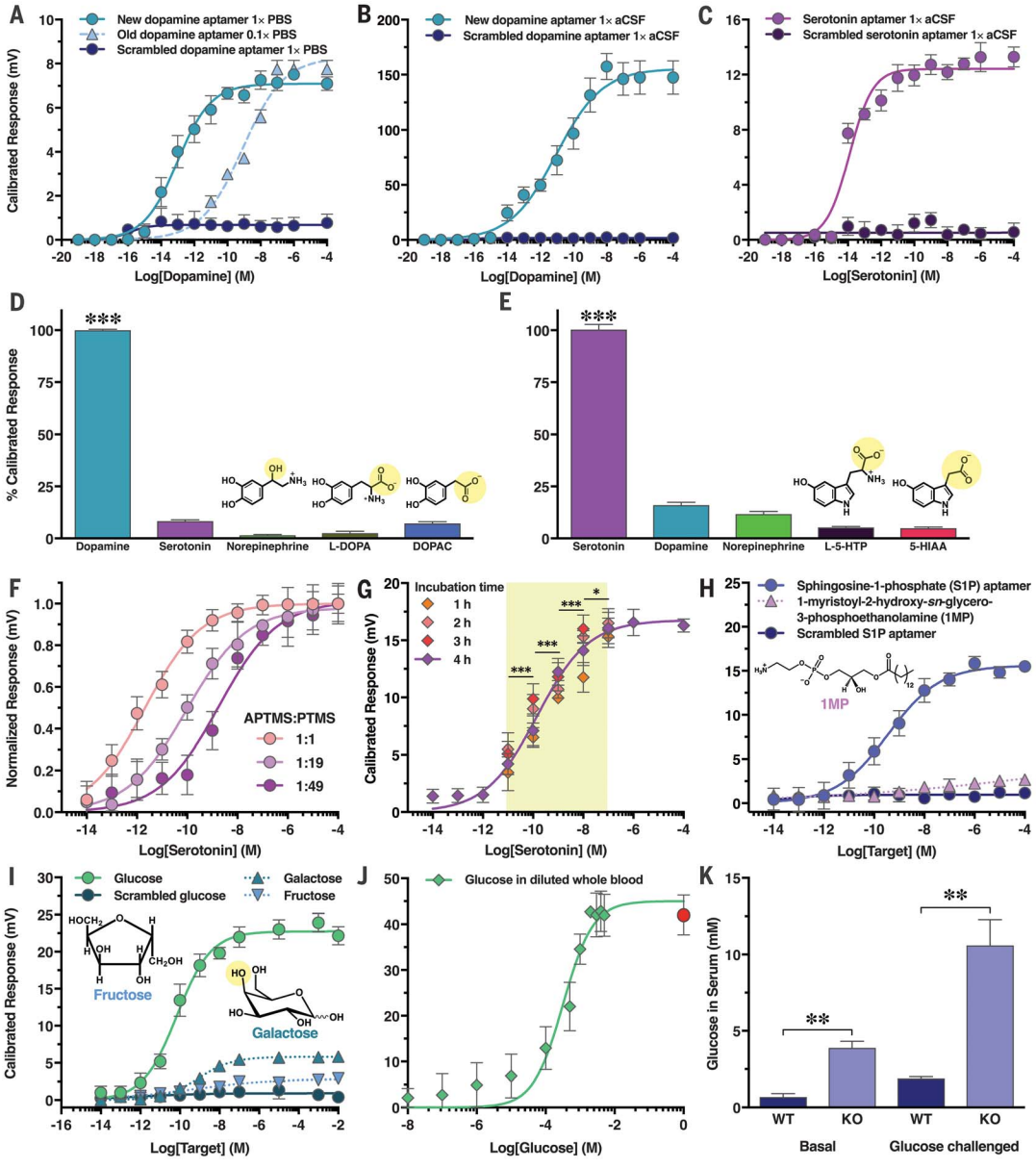


Fig. 2. Electronic small-molecule detection using aptamer-functionalized FET sensors. (**A**) Responses of FET sensors functionalized with the new dopamine aptamer [K_d = 150 nM, full-strength PBS (1× PBS)] or its scrambled sequence, compared to FET responses with a previously reported dopamine aptamer (K_d = 1 μM, 0.1× PBS) (*2*). (**B**) New and scrambled dopamine aptamer–FET responses to dopamine in 1× aCSF. (**C**) For serotonin aptamer–FETs, serotonin in 1× aCSF led to concentration-dependent responses, whereas scrambled serotonin sequences showed negligible responses. (**D**) New dopamine aptamer–FET responses to 100 μM norepinephrine, serotonin, L-3,4-dihydroxyphenylalanine (L-DOPA), and 3,4-dihydroxyphenylacetic acid (DOPAC) were negligible relative to dopamine (10 nM). (**E**) Serotonin aptamer–FET responses to 100 μM dopamine, norepinephrine, L-5-hydroxytryptophan (L-5-HTP), or 5-hydroxyindoleacetic acid (5-HIAA) were negligible relative to serotonin (10 nM). (**F**) Serotonin aptamer–FET sensitivities were shifted by altering ratios of amine-terminated/methyl-terminated silanes for surface tethering. (**G**) Serotonin aptamer–FETs after 1 to 4 hours of incubation in serotonin-free brain tissue followed by addition of serotonin exhibited reproducible responses with differentiable physiological concentrations. (**H**) S1P aptamer–FETs showed concentration-dependent responses to S1P but not to a phospholipid with similar epitopes or a scrambled sequence in 1× HEPES. (**I**) Glucose sensing in 1× Ringer's buffer. Responses of glucose aptamer–FETs were minimal or negligible for galactose, fructose, and a scrambled sequence. (**J**) Glucose aptamer–FET responses in mouse whole blood diluted in Ringer's to construct a concentration curve. The red circle represents response in undiluted whole blood. (**K**) Glucose aptamer–FETs enabled differentiation of hyperglycemia in serotonin transporter–deficient (KO) mice versus wild-type (WT) mice by measuring glucose levels in diluted serum under basal and glucose-challenged conditions. All calibrated responses were at gate voltage V_G = 100 mV. Error bars are ± SEM with N = 6 [(A) to (C), (H), (I), and (K)] or N = 3 samples per group [(D) to (G) and (J)]. In (D) and (E), ***P < 0.001 versus countertargets; in (G), ***P < 0.001, *P < 0.05 versus different serotonin concentrations (10 pM to 100 nM); in (K), **P < 0.01 KO versus WT. **P < 0.01 KO versus WT.

added serotonin to brain tissue from mice lacking neuronal serotonin (i.e., *Tph2* null mice) (Fig. 2G) (*21*). Electronic FET responses differentiated physiologically relevant serotonin concentrations (10 pM to 100 nM) (*14*). Sensor responses to dopamine or to the serotonin metabolite 5-hydroxyindoleacetic acid in tissue were negligible (fig. S9A). The high sensitivity of aptamer-FETs offsets losses often encountered in biological environments, and sensitivity for modest changes in target concentrations was observed despite large sensing ranges. Sensor performance in tissue was reproducible when repeated 12 hours later (fig. S9B). Moreover, continuous exposure of serotonin aptamer–FETs to brain tissue for 1 to 4 hours produced stable concentration-dependent conductance responses and was another indication of sensor stability (Fig. 2G).

Aptamer-FET responses to the zwitterionic lipid S1P were recorded at concentrations ranging from 10 pM to 100 nM. A nontarget lipid (1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphoethanolamine) bearing similar epitopes (Fig. 2H) exhibited negligible responses, as did a scrambled S1P sequence (table S5). Glucose aptamer–FETs exhibited concentration-dependent responses to glucose (10 pM to 10 nM). The FET responses to other monosaccharides (e.g., galactose and fructose) were minimal, as were responses when a scrambled glucose sequence was used (Fig. 2I and table S5). Experiments with SERS corroborated target-specific recognition in close proximity to substrates for S1P and glucose aptamers (fig. S8, E and F).

We detected glucose in whole blood diluted with Ringer's buffer (10 μM to 1 mM; Fig. 2J). We also measured glucose levels in diluted serum from mice lacking serotonin transporter expression characterized by hyperglycemia (*22*). Elevations in serum glucose in basal and glucose-challenged states were observed using glucose aptamer–FETs (Fig. 2K); glucose concentrations were similar to those determined in whole blood using a glucometer (fig. S10). These findings demonstrated the feasibility of aptamer-FET sensing in diluted, yet full–ionic strength, blood or serum and the ability to differentiate modest yet physiologically relevant differences in neutral target concentrations.

Aptamer-FET sensing enabled observations suggestive of mechanism. In addition to FET responses at subthreshold-regime gate voltages (Fig. 2), we examined characteristics of FET transfer curves [i.e., source-drain currents (I_{DS}) versus source-gate voltage sweeps (V_{GS})]. Transfer curves for increasing target concentrations diverged for dopamine aptamer– and glucose aptamer–FETs versus serotonin aptamer– and S1P aptamer–FETs (Fig. 3, A to D). Dopamine and serotonin each have one positive charge at physiological pH. Transfer curve divergence for these molecules enables us to conclude that signal transduction mechanisms based exclusively on target charge, as has been proposed (*23*), are incorrect and preclude detection of neutral targets. The divergence of *I-V* curves also suggests different conformational changes upon target binding. For dopamine and glucose, transfer curves were consistent with aptamer reorientations occurring such that substantial portions of the negatively charged backbones moved closer to n-type semiconductor channels, thereby increasing electrostatic repulsion of charge carriers (band bending) and decreasing transconductance, measured as target-related current responses (Fig. 3E). In contrast, we hypothesized that serotonin and S1P aptamers moved predominantly away from channel surfaces upon target capture, thereby increasing transconductance (Fig. 3F).

We used circular dichroism (CD) spectroscopy to gain additional insight (*24*, *25*). For dopamine and serotonin, large changes in CD peak positions and relative intensities suggested shifts away from predominant duplex signals (maxima at ~280 nm) and formation of new target-induced structural motifs. A parallel (or mixed) G-quadruplex (maximum shifted to 260 nm) (*26*) was suggested for dopamine-aptamer complexes (Fig. 4A), whereas an antiparallel G-quadruplex (maximum shifted to 290 nm) was indicated for serotonin-aptamer

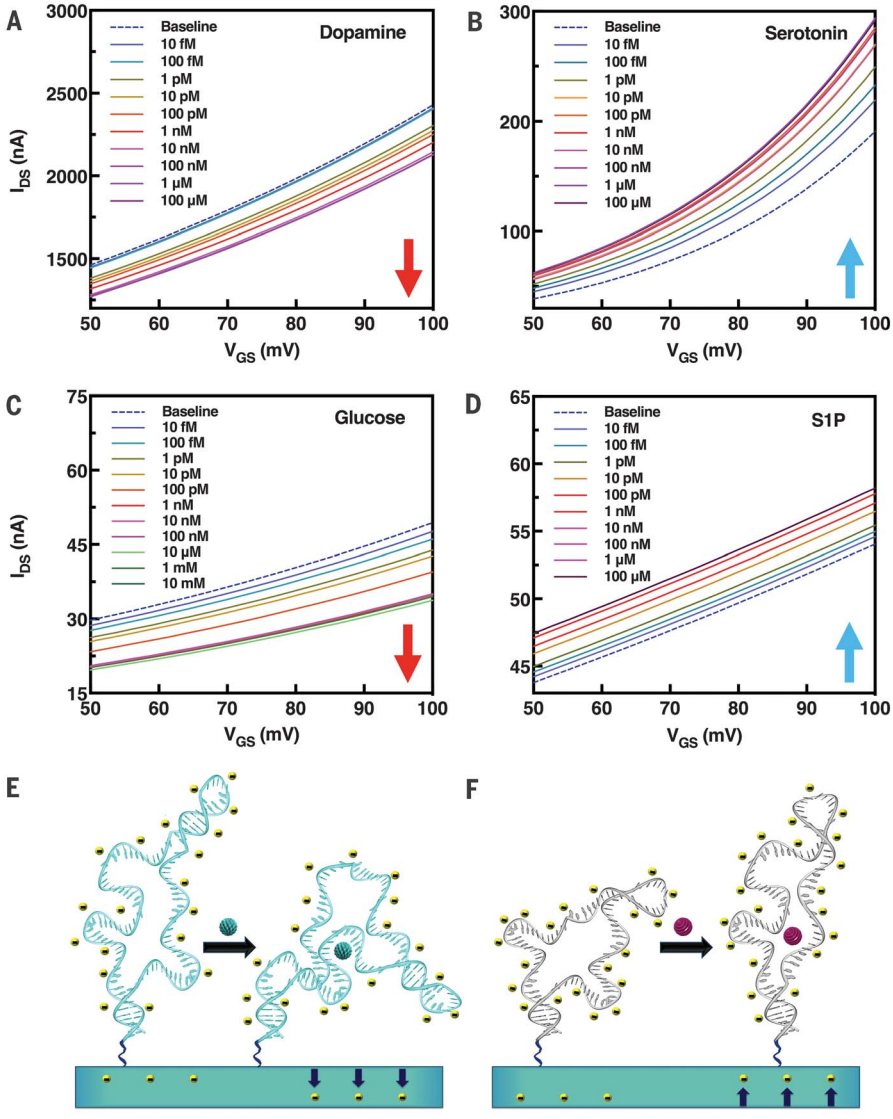


Fig. 3. Aptamer-functionalized FET mechanisms. (**A**) Exposure of dopamine aptamer–FETs to dopamine (1× aCSF) led to concentration-dependent reductions in source-drain currents. (**B**) For serotonin aptamer–FETs, increasing concentrations of serotonin (1× aCSF) produced increases in source-drain currents. (**C**) Exposure of glucose aptamer–FETs to glucose (1× Ringer's) led to reductions in source-drain currents. (**D**) The S1P aptamer–FET transfer curves (1× HEPES) increased in response to target concentrations. Transfer curves shown are representative of $N = 6$ individual measurements. (**E** and **F**) Hypothesized mechanism of stem-loop aptamer target-induced reorientations in close proximity to semiconductor channels and within or near the Debye length. In (E), aptamers reorient closer to FETs to deplete channels electrostatically (e.g., dopamine, glucose). In (F), aptamer stem-loops reorient away from semiconductor channels, thereby increasing transconductance (e.g., serotonin, S1P). Schematics are idealized and do not reflect individual aptamer secondary structural motifs.

complexes (Fig. 4B). As with fluorescence, FET, and SERS data, CD indicated selectivity of dopamine and serotonin aptamers for their targets versus similarly structured countertargets (fig. S11, A and B). Although fluorescence, FET, and SERS findings specified target recognition for glucose and S1P aptamers, changes in CD spectra were not observed for these aptamers (fig. S11, C and D). Thus, for glucose and S1P aptamers, all major DNA domains (i.e., G-quartets, helices, and single-stranded regions) were formed prior to target binding, and adaptive binding occurred through spatial rearrangement of existing secondary structures and companion ions (*27*).

We used Förster resonance energy transfer (FRET) to investigate changes in aptamer backbone distances during target-induced conformational changes. We identified FRET sensors for serotonin and glucose aptamers (table S6). For serotonin, the decrease in FRET (Fig. 4C and fig. S12A) was consistent with a substantial fraction of the longest loop in the G-quadruplex moving away from the semiconductor surface, and hence with the upward shifts in FET transfer curves (Fig. 3B). For glucose, FRET results (Fig. 4D and fig. S12B) supported movement of the second stem in the aptamer toward the semiconductor surface, consistent with downward shifts in FET transfer curves (Fig. 3C). For the glucose aptamer, we increased the stem lengths for attachment to FET surfaces (Fig. 4E). Conductance responses decreased with additional base pairs (Fig. 4F), which suggested that recognition occurred farther away from FETs as the attachment stems became longer. This strategy might be used to tune sensitivity ranges of sensor array elements, thereby extending the ranges of arrays.

Together, all mechanistic findings are consistent with aptamer conformational changes enabling FET sensing under physiological conditions and without aptamer labeling or additional surface chemistries [compare with (*28*)]. Note that because of the aptamer selection strategy, target-specific aptamer reorientations occur in close proximity to semiconductor surfaces, and in some cases, even in the absence of formation of new secondary structural motifs. General aptamer reorientation can be inferred from FET gate-voltage sweeps, with additional FET mechanisms possibly contributing for specific sensors (e.g., band bending) and permittivity and mobility changes. Unlike large protein receptors (e.g., antibodies), highly selective, chemically synthesized, compact nucleic acid receptors identified through in vitro selection are amenable to affinity tuning (*29*, *30*) and targeting of a wide variety of small (and large) molecules for electronic sensing (*23*).

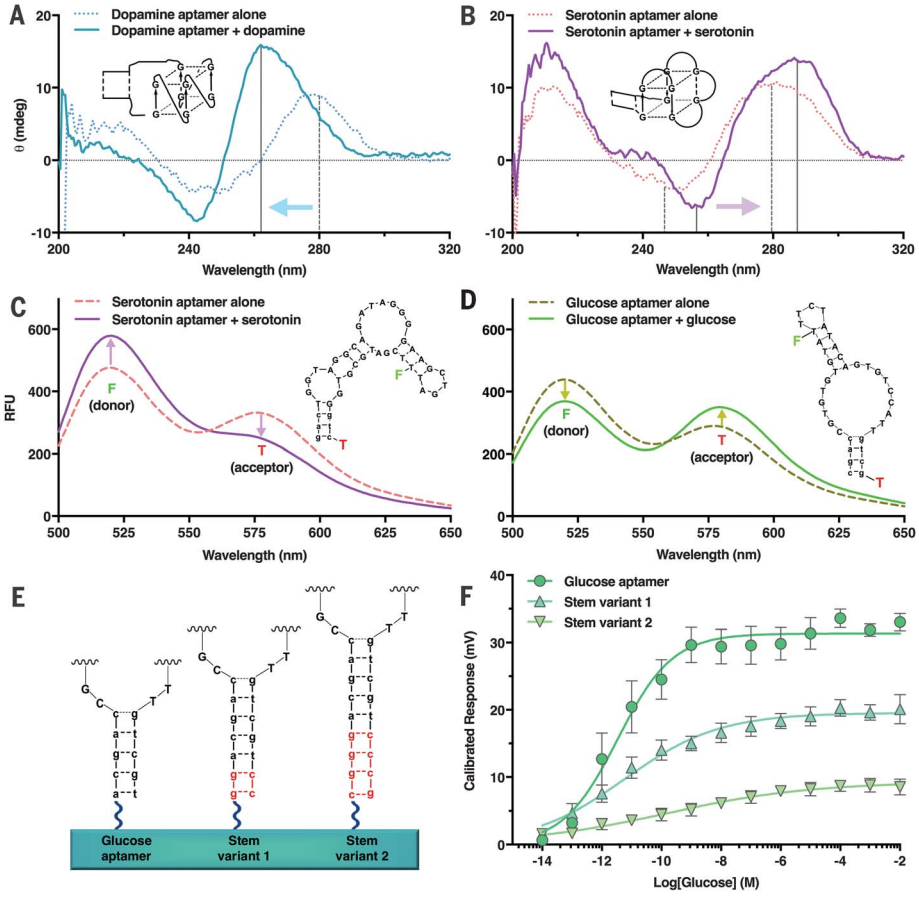


Fig. 4. Changes in aptamer secondary structures upon adaptive binding to small-molecule targets. (**A**) Circular dichroism spectroscopy of the dopamine aptamer upon target capture showed spectral shifts indicating formation of a parallel G-quadruplex (1× aCSF). (**B**) By contrast, the serotonin aptamer showed shifts in peak positions indicating formation of an antiparallel G-quadruplex. (**C** and **D**) FRET between donor- [fluorescein (F), excited at 470 nm] and acceptor- [5-carboxytetramethylrhodamine (T)] labeled aptamers was monitored before and after target incubation. For serotonin aptamers (C), donor fluorescence increased while acceptor emission decreased upon serotonin incubation, suggesting that fluorophores move farther away from each other upon target exposure. Conversely, for glucose aptamers (D), the emission spectra for the acceptor increased while donor fluorescence decreased upon glucose exposure, indicative of the acceptor moving closer to the donor, thereby enabling increased energy transfer. Stem-loop movement directions indicated by FRET for glucose versus serotonin aptamers are consistent with their divergent FET transfer curve directions in Fig. 3. (**E** and **F**) For glucose aptamer–FETs with rigid double-stranded attachment stems [(E), left], increasing distances from semiconductor surfaces by increasing the stem lengths [(E), stem variants, right] resulted in length-associated decreases in FET calibrated responses (1× Ringer's solution) (F). Spectra shown in (A) to (D) are representative of $N = 2$ samples per condition; error bars in (F) are ±SEM with $N = 3$ samples per group.

REFERENCES AND NOTES

1. Y. Cui, Q. Wei, H. Park, C. M. Lieber, *Science* **293**, 1289–1292 (2001).
2. J. Kim *et al.*, *ACS Nano* **9**, 4572–4582 (2015).
3. P. Casal *et al.*, *Philos. Trans. R. Soc. A* **370**, 2474–2488 (2012).
4. A. Vacic *et al.*, *J. Am. Chem. Soc.* **133**, 13886–13889 (2011).
5. P. S. Weiss, P. L. Trevor, M. J. Cardillo, *J. Chem. Phys.* **90**, 5146–5153 (1989).
6. W. S. Liao *et al.*, *Science* **337**, 1517–1521 (2012).
7. K. Shoorideh, C. O. Chui, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5111–5116 (2014).
8. M. P. Landry *et al.*, *Nat. Nanotechnol.* **12**, 368–377 (2017).
9. R. Nutiu, Y. Li, *Angew. Chem. Int. Ed.* **44**, 1061–1065 (2005).
10. K. A. Yang *et al.*, *Nat. Chem.* **6**, 1003–1008 (2014).
11. R. Walsh, M. C. DeRosa, *Biochem. Biophys. Res. Commun.* **388**, 732–735 (2009).
12. S. C. Altieri *et al.*, *Neuropsychopharmacology* **40**, 1456–1470 (2015).
13. N. Nakatsuka, A. M. Andrews, *ACS Chem. Neurosci.* **8**, 218–220 (2017).
14. H. Yang, A. B. Thompson, B. J. McIntosh, S. C. Altieri, A. M. Andrews, *ACS Chem. Neurosci.* **4**, 790–798 (2013).
15. A. P. Alivisatos *et al.*, *ACS Nano* **7**, 1850–1866 (2013).
16. G. I. Perez, C. M. Knudson, L. Leykin, S. J. Korsmeyer, J. L. Tilly, *Nat. Med.* **3**, 1228–1232 (1997).
17. N. T. Rodeberg, S. G. Sandberg, J. A. Johnson, P. E. M. Phillips, R. M. Wightman, *ACS Chem. Neurosci.* **8**, 221–234 (2017).
18. C. Ke, H. Destecroix, M. P. Crump, A. P. Davis, *Nat. Chem.* **4**, 718–723 (2012).
19. W. G. Purschke *et al.*, *Biochem. J.* **462**, 153–162 (2014).
20. S. S. Masango *et al.*, *Nano Lett.* **16**, 4251–4259 (2016).
21. M. Angoa-Pérez *et al.*, *ACS Chem. Neurosci.* **5**, 908–919 (2014).
22. X. Chen, K. J. Margolis, M. D. Gershon, G. J. Schwartz, J. Y. Sze, *PLOS ONE* **7**, e32511 (2012).
23. N. Nakatsuka, P. S. Weiss, A. M. Andrews, *Chem. Rev.* (2018).
24. W. Liu *et al.*, *J. Phys. Chem. B* **115**, 13051–13056 (2011).
25. S. Nagatoishi, Y. Tanaka, K. Tsumoto, *Biochem. Biophys. Res. Commun.* **352**, 812–817 (2007).
26. J. Kypr, I. Kejnovská, D. Renčiuk, M. Vorlícková, *Nucleic Acids Res.* **37**, 1713–1725 (2009).
27. Y. Nagai *et al.*, *Langmuir* **29**, 9951–9957 (2013).
28. N. Gao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14633–14638 (2016).
29. F. Ricci, A. Vallée-Bélisle, A. J. Simon, A. Porchetta, K. W. Plaxco, *Acc. Chem. Res.* **49**, 1884–1892 (2016).
30. R. E. Armstrong, G. F. Strouse, *Bioconjug. Chem.* **25**, 1769–1776 (2014).

ACKNOWLEDGMENTS

We thank D. Kuhn and H. Chen for resources and assistance, and K. Gothelf for helpful discussion. **Funding:** Supported by NIH (DA045550, CA199849, GM104960), NSF (CCF1518715, 1509794, CMMI-1636136), CalBrain, NantWorks, Hewlett Packard, the Merkin Family Foundation, and the China Scholarship Council. **Author contributions:** N.N., K.A.Y., P.S.W., M.N.S., and A.M.A. conceived experiments; N.N. and K.M.C. performed aptamer-FET measurements; K.A.Y. and M.N.S. designed and isolated stem-loop aptamer receptors and conducted FRET measurements; N.N., J.M.A., and X.X. conducted spectroscopy experiments; H.Y. carried out mouse experiments; C.Z., B.Z., Y.S.R., and Y.Y. designed and fabricated thin-film transistors; and N.N., P.S.W., M.N.S., and A.M.A. wrote the manuscript. **Competing interests:** N.N., K.A.Y., P.S.W., M.N.S., and A.M.A. have filed a patent on stem-loop receptor-based field effect sensor devices for sensing at physiological salt concentrations, U.S. application no. 504901225. M.N.S. has patent applications, a start-up company, and consulting income for work on small-molecule aptamers. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are presented in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/319/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S9
References (*31*–*55*)

15 August 2017; resubmitted 30 April 2018
Accepted 6 August 2018
Published online 6 September 2018
10.1126/science.aao6750

GRAPHENE

A family of finite-temperature electronic phase transitions in graphene multilayers

Youngwoo Nam[1,2], Dong-Keun Ki[1,3], David Soler-Delgado[1], Alberto F. Morpurgo[1]*

Suspended Bernal-stacked graphene multilayers up to an unexpectedly large thickness exhibit a broken-symmetry ground state whose origin remains to be understood. We show that a finite-temperature second-order phase transition occurs in multilayers whose critical temperature (T_c) increases from 12 kelvins (K) in bilayers to 100 K in heptalayers. A comparison of the data with a phenomenological model inspired by a mean-field approach suggests that the transition is associated with the appearance of a self-consistent valley- and spin-dependent staggered potential that changes sign from one layer to the next, appearing at T_c and increasing upon cooling. The systematic evolution with thickness of several measured quantities imposes constraints on any microscopic theory aiming to analyze the nature of electronic correlations in this system.

Clean two-dimensional (2D) conductors in the presence of a large perpendicular magnetic field are strongly correlated systems. Their ground states are determined by Coulomb repulsion between electrons and are characterized by broken symmetries and nontrivial topological invariants that depend sensitively on electron density (n) and applied field (B) (*1*–*4*). Accordingly, a series of quantum phase transitions occurs upon varying n and B, which manifest themselves in the very rich phenomenology ubiquitously observed in magneto-transport measurements (*5*–*13*). These considerations hold true irrespective of the specific 2D conductor considered, because they rely almost exclusively on the formation of Landau levels that quench the electron kinetic energy and allow electron-electron (*e*-*e*) interactions to dominate. In the absence of Landau levels, however, *e*-*e* interactions play a much less prominent role.

Multiple recent experiments indicate that in graphene multilayers this is not the case (*14*–*19*). A gapped insulating state at $B = 0$, first reported in Bernal-stacked bilayers (*14*–*17*), has been recently observed in all even Bernal-stacked multilayers up to octalayer graphene (8LG) (*18*, *19*). By contrast, odd Bernal-stacked multilayers (so far mono- and trilayer have been studied) remain conducting at low temperatures (*20*–*23*). These findings defy common expectations—namely, that the behavior of graphene multilayers should approach that of graphite as thickness is increased. No direct information about the nature of the insulating state in even multilayers could be obtained before now, because the phenomenon is observed only in suspended graphene devices of the highest quality (*14*–*23*), which makes measurements other than transport challenging. Because the resistance of even multilayers increases upon cooling without showing abrupt changes at any specific temperature, it is unclear even whether the insulating state results from a quantum phase transition (with a gap present at all temperatures) or whether a phase transition occurs at a critical temperature (T_c) with the gap vanishing for temperatures $T > T_c$ (*16*, *17*, *24*–*28*). To address these issues, we studied ultraclean, suspended Bernal-stacked multilayers of graphene near charge neutrality and have shown that at $B = 0$ these systems unambiguously exhibit an interaction-driven finite-temperature phase transition to a broken-symmetry state at a critical temperature T_c that depends on thickness.

To draw these conclusions, we focused on the density of electrons present in the conduction band of charge-neutral multilayers, $n_{th}(T)$, whose temperature dependence is determined by the low-energy density of states (DOS). At sufficiently low T, $n_{th}(T)$ should show an exponential increase in the presence of a gap—given that at charge neutrality electrons in the conduction band are thermally activated from the valence band—or stay constant if an overlap between the valence and conduction bands is present. If the multilayers are zero-band gap semiconductors, $n_{th}(T)$ is expected to increase with increasing T, with a specific dependence determined entirely by the low-energy DOS. Quantitative information can be obtained by comparing experimental data with the dependence of $n_{th}(T)$ calculated from a chosen theoretical expression for the low-energy DOS, as we illustrate for the case of bilayers. Under the assumption that only the nearest-neighbor in-plane (γ_0) and out-of-plane (γ_1) hopping terms are relevant (the so-called minimal tight-binding model), the low-energy quadratic dispersion relation of electrons in bilayer graphene (Fig. 1A) leads to an energy-independent DOS (Fig. 1B). The resulting density $n_{th}(T)$ of electrons in the conduction band then increases linearly with temperature (Fig. 1C). However, if a gap Δ opens (Fig. 1D), the DOS is modified (Fig. 1E), and so is the temperature dependence of $n_{th}(T)$. Figure 1F shows $n_{th}(T)$ expected for a gap Δ exhibiting a mean-field temperature dependence and vanishing at T_c [see the inset and (*29*)]. A transition becomes visible at $T = T_c$, below which $n_{th}(T)$ is pronouncedly suppressed compared with the noninteracting case.

What makes these considerations useful is that the very small charge inhomogeneity present in suspended graphene multilayers enables the density $n_{th}(T)$ of thermally excited electrons to be determined experimentally over a broad range of temperatures. $n_{th}(T)$ can be extracted from the density dependence of the conductance $G(n)$ near charge neutrality. To understand why and how, it is useful to look at the double logarithmic plot of $G(n)$ (inset of Fig. 1G). We see that there is a range of n over which $\log(G)$ is constant and that $\log(G)$ starts increasing substantially only when n becomes larger than a threshold [which we denote $n^*(T)$; $n^*(T)$ increases at higher T]. The physical reason is clear: The square conductance remains virtually unchanged as long as the density n of electrons accumulated by the gate voltage is much smaller than the density of electrons $n_{th}(T)$ already present owing to thermal activation from the valence band. Finding that the square conductance increases substantially only when n exceeds $n^*(T)$ therefore implies that $n_{th}(T) \sim n^*(T)$, as discussed previously for mono- and bilayer graphene (*21*, *30*–*32*). In practice, the exact value of $n^*(T)$ depends on the criterion used for its definition: Here, we define $n^*(T)$ as the value of n for which the conductivity $\sigma(n,T) = 1.67\sigma(n = 0,T)$, a relation that we can validate for bilayers and that is approximately correct for thicker multilayers (*29*). We emphasize, however, that none of our key conclusions depend on the precise criterion used (*29*).

The temperature dependence of $n^*(T)$ extracted from measurements on four different bilayer devices is shown in Fig. 1G. For all devices, a critical temperature $T_c \cong 12$ K is found, below which $n^*(T)$ is suppressed compared with the gapless case. The red continuous curve in Fig. 1G, which represents $n_{th}(T)$ expected for a gap having a mean-field temperature dependence and $\Delta_0 = \Delta(T = 0) \cong 1.9$ meV, reproduces the data very satisfactorily. The presence of a clear transition below which a finite Δ appears, the value of T_c, and the shape of $n^*(T)$ are very robust against all aspects of the data analysis. They can be extracted directly from the data without any assumption about the DOS and do not depend on the criterion used to extract $n^*(T)$ (only the quantitative determination of Δ requires the DOS to be specified) [see (*29*)]. Notably, we found the same value for the critical temperature in all the different devices measured, despite the sample-to-sample fluctuations in the absolute value of the low-temperature resistance and in how pronounced the insulating

[1]Department of Quantum Matter Physics (DQMP) and Group of Applied Physics (GAP), University of Geneva, 24 Quai Ernest-Ansermet, CH1211 Genéve 4, Switzerland. [2]Department of Physics, Gyeongsang National University, Jinju-daero 501, Jinju-si, South Korea. [3]Department of Physics, The University of Hong Kong, Hong Kong, China.
*Corresponding author. Email: alberto.morpurgo@unige.ch

state was at the lowest temperature of our measurements [because of these fluctuations, even the occurrence of an insulating state driven by *e-e* interaction has been questioned in some past experimental work (*14*–*17*, *30*, *33*)]. The reason for this high degree of reproducibility is that our measurements effectively probe the DOS averaged over the entire device area, whereas in the insulating state the absolute value of the resistance is strongly affected by any percolating conducting path [e.g., at edges (*34*, *35*)] that occupies a negligible fraction of the total area, giving a negligible contribution to the DOS. The highly reproducible behavior of $n^*(T)$ allows us to establish unambiguously the occurrence of an insulating, gapped state near charge neutrality, which is entered through a second-order phase transition at T_c = 12 K.

The same strategy outlined for bilayers can be applied to thicker even multilayers, in which multiple conduction and valence bands are present (*29*, *36*–*39*). Recent work (specifically, the quantization of the Hall effect in low-magnetic-field and magneto-Raman experiments) has provided evidence that in suspended devices at low energy these bands are well described by including only nearest-neighbor in-plane (γ_0) and out-of-plane (γ_1) hopping terms (*18*, *19*, *40*, *41*). According to this description, all conduction and valence bands in even multilayers are quadratic at low energy (with different effective masses), so that the DOS in the noninteracting case is again energy independent [see (*29*)]. The same argument used for bilayers implies that $n_{th}(T)$ increases linearly with temperature in the absence of interactions and that the opening of a gap causes a suppression for $T < T_c$. Figure 2, A and C, illustrates the results of experiments performed on tetralayer (4LG) and hexalayer (6LG) graphene devices: $n^*(T)$ depends linearly on temperature at sufficiently high temperatures, but below a critical temperature T_c (38 K for 4LG and 90 K for 6LG), $n^*(T)$ is suppressed, just as for bilayers (Fig. 1, F and G). For 4LG and 6LG, the data are excellently reproduced by assuming that a gap Δ with a mean-field temperature dependence opens simultaneously at T_c on all quadratic bands (red line; $\Delta_0 \cong 5.2$ meV for 4LG and $\Delta_0 \cong 13$ meV for 6LG). As in the bilayer case, the temperature dependence of the resistance at charge neutrality (Fig. 2, B and D) demonstrates the insulating nature of thicker even multilayers but provides no specific feature allowing the identification of a critical temperature.

The analysis of the measured temperature dependence of $n^*(T)$ and the comparison with the calculated expression for $n_{th}(T)$ can also be performed and interpreted in the same way for odd Bernal multilayers. Within the minimal tight-binding model mentioned above, odd multilayers contain multiple conduction and valence bands, all having a quadratic dispersion, except for one that is a linear Dirac band (*29*, *36*–*39*). The DOS associated with the linear band vanishes at charge neutrality and contributes at most a few percent of the total DOS in the temperature and density ranges relevant for our work (*29*). It is therefore negligible in practice, so that the same considerations made for even multilayers hold true for odd ones. In Fig. 3, data measured on odd Bernal-stacked multilayers confirm that this is the case for both trilayer (3LG) (Fig. 3A) and heptalayer (7LG) graphene (Fig. 3C): $n^*(T)$ increases linearly with temperature at high temperatures, exhibiting a pronounced suppression for $T < T_c$ (with T_c = 33 K for 3LG and 100 K for 7LG), in complete analogy to the case of even multilayers. Like the data for even multilayers, the data for odd multilayers agree quantitatively with the behavior expected if a gap Δ opens simultaneously at T_c in all quadratic bands, with a mean-field temperature dependence and T = 0 values $\Delta_0 \cong 5$ meV for 3LG and $\Delta_0 \cong 13$ meV for 7LG. For odd multilayers, the low-temperature conductivity remains finite, $\sigma \sim e^2/h$ (where h is Planck's constant), indicating that the linear band does not gap out (Fig. 3, B and D).

In summary, we find that a phase transition occurs in all the Bernal-stacked multilayers investigated, irrespective of whether they are even or odd, with only the even ones becoming insulating at low temperatures. Data for multilayers of all

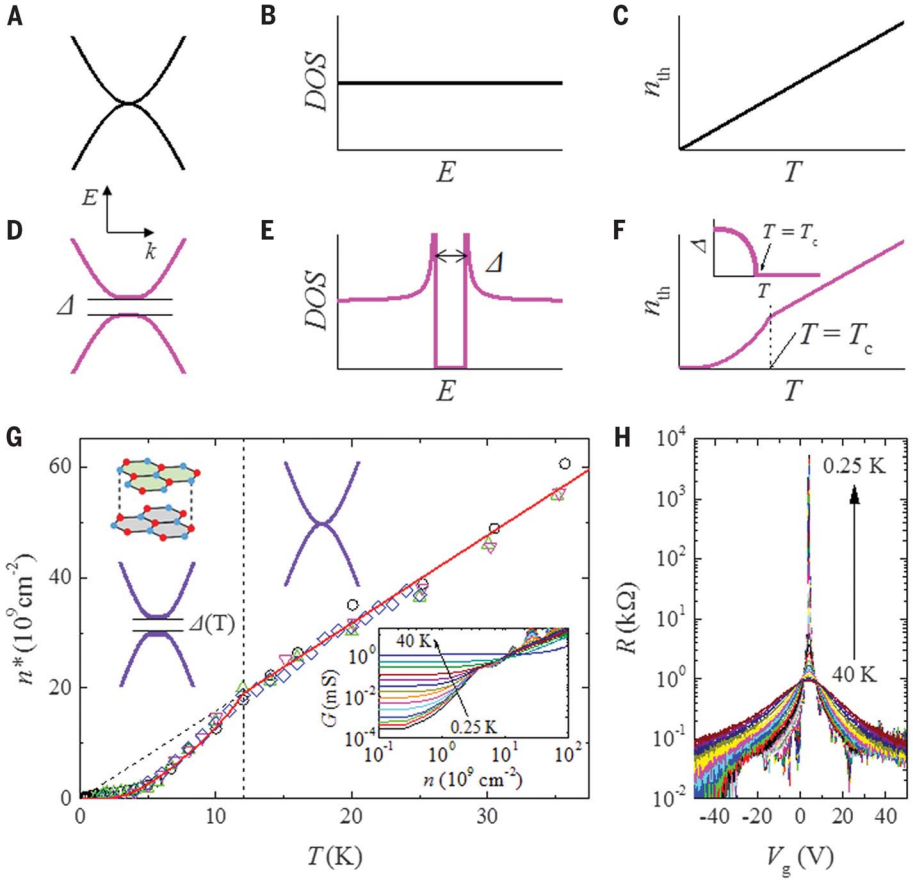


Fig. 1. Electronic phase transition in bilayer graphene. (**A**) In the absence of interactions, Bernal-stacked graphene bilayers are zero-gap semiconductors with valence and conduction bands dispersing quadratically at low energy (E), resulting in (**B**) a constant DOS. (**C**) The density of electrons thermally excited from the conduction band to the valence band, $n_{th}(T)$, increases linearly with temperature. (**D**) The opening of a gap (Δ) resulting from interactions leads to (**E**) a modified DOS and causes (**F**) a suppression of $n_{th}(T)$ [inset, $\Delta(T)$ expected from a mean-field description, used to calculate $n_{th}(T)$]. Each symbol in (**G**) represents $n^*(T)$ [$\sim n_{th}(T)$ as discussed in the text and (*29*)] measured on a distinct bilayer device, showing a transition at T_c = 12 K in all cases. Above T_c, the value of the slope of $n^*(T)$ matches the one expected from the known DOS of 2LG (i.e., the effective mass). The red line is a fit to the expression for $n_{th}(T)$ calculated with a mean-field temperature dependence of $\Delta(T)$ (*29*). (Bottom inset) The double logarithmic plot of $G(n)$ measured at different values of T (from bottom to top, 0.27, 0.33, 0.43, 0.53, 0.78, 1.1, and 1.4 K; 2.0 to 5.0 K in 1-K steps; and 6.7, 10, and 40 K), from which $n^*(T)$ is extracted. (**H**) The gate-voltage (V_g) dependence of the resistance (R) of this same device for the same range of temperatures (from top to bottom, 0.25, 0.27, 0.29, 0.33, 0.38, 0.43, 0.48, 0.53, 0.62, 0.78, 0.94, 1.1, 1.2, 1.4, 1.6, and 1.7 K; 2.0 to 5.0 K in 0.5-K steps; 5.0, 5.7, and 6.7 K; 8.0 to 16 K in 2-K steps; and 20 to 40 K in 5-K steps) as in (G) exhibits a pronounced insulating behavior around charge neutrality but no indications of a phase transition.

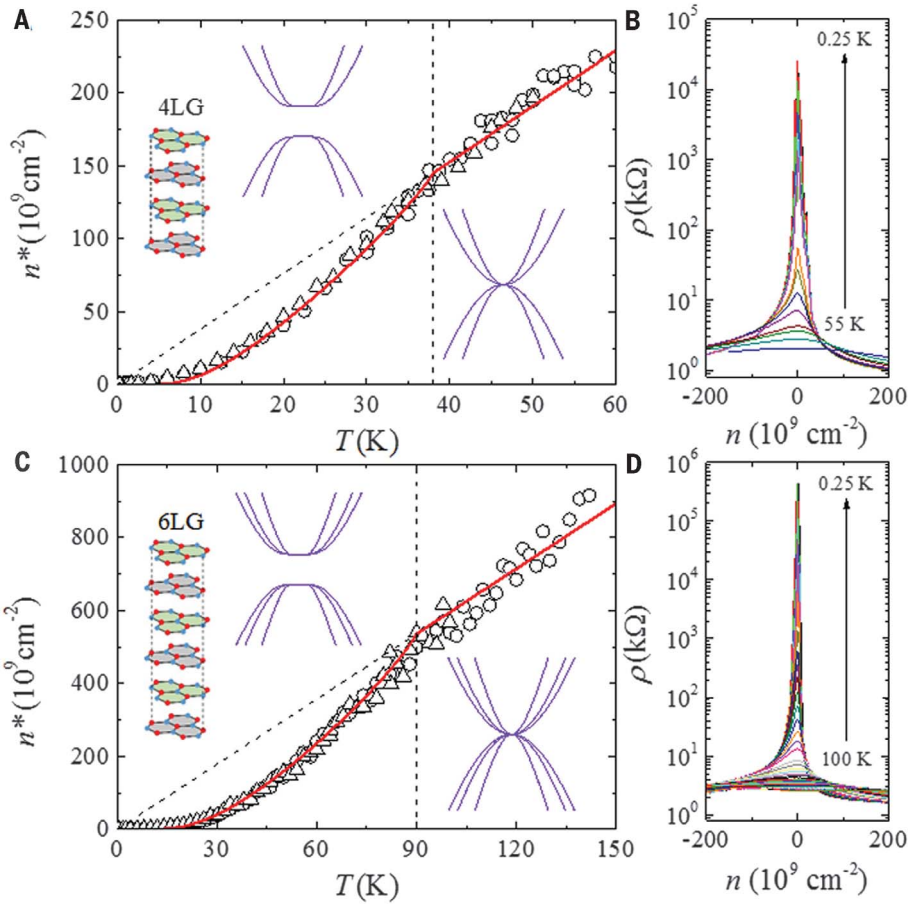


Fig. 2. Electronic phase transition in even Bernal-stacked multilayers. Shown are the data for (**A** and **B**) 4LG and (**C** and **D**) 6LG graphene. (A) and (C) show that at sufficiently high temperatures, the measured $n^*(T)$ depends linearly on T. Below the critical temperature T_c (38 K for 4LG and 90 K for 6LG), $n^*(T)$ is suppressed, as found in bilayers (Fig. 1). The red continuous curves in (A) and (C), which correspond to the dependence of $n_{th}(T)$ expected for a gap having a mean-field temperature dependence, reproduce the data. In (A) and (C), the circles and triangles represent data obtained from different devices, demonstrating the high reproducibility of our observations. [(B) and (D)] Charge carrier density dependence of the resistivity (ρ) at different temperatures [from top to bottom, 0.25 to 1.75 K in 0.3-K steps and 10 to 55 K in 5-K steps in (B) and 0.25 K, 0.4 to 1.7 K in 0.3-K steps, 2.5 K, and 4.0 to 100 K in 2-K steps in (D)], showing an insulating state around charge neutrality at low temperatures. The asymmetry visible for positive and negative values of n is caused by the formation of a *pn* junction at the contacts, as typically observed in two-terminal suspended graphene devices. In all multilayers for which four-terminal devices (*49*) could be realized, the resistivity is nearly perfectly symmetric upon changing of the sign of n (see, e.g., Figs. 1H and 3B).

thicknesses are shown in Fig. 4A, each with the corresponding fit to the calculated temperature dependence of the density of thermally activated electrons $n_{th}(T)$. Notably, all data collapse on top of one another if $n^*(T)/n^*(T_c)$ is plotted versus T/T_c. (Fig. 4A, inset). T_c increases linearly with increasing thickness (Fig. 4B), and so does the gap Δ_0 that is proportional to T_c (Fig. 4C): The best linear fit (the continuous line in Fig. 4C) is close to the expected mean-field value, $\Delta_0/k_BT_c = 1.76$ (where k_B is Boltzmann's constant) (the dashed line in Fig. 4C) (*29*).

The experimental findings consistently provide substantial evidence about the microscopic nature of the broken-symmetry state, namely, that the effect of interactions is well described by a valley- and spin-dependent mean-field staggered potential that changes sign from one layer to the next, acting as an order parameter (*18*, *19*). A scenario based on the minimal tight-binding model augmented with a self-consistent staggered potential explains that at low temperatures, even multilayers become fully insulating, whereas in odd multilayers a finite conductivity of $\sim e^2/h$ persists (see Figs. 2, B and D, and 3, B and D); that from 1LG to 8LG, the first quantum Hall plateau appears systematically at filling factor $\nu = 2N$, with $\sigma_{xy} = 2Ne^2/h$, where N is the multilayer thickness [see (*18*) and (*19*)]; that the transition occurs in odd multilayers; why the gap in the different quadratic bands has the same magnitude; and why the gap opens simultaneously in all bands at the same critical temperature (inasmuch as the latter two points are concerned, a generic mechanism could gap each band independently from the others, resulting in multiple transitions with different values of T_c and Δ) (*42*, *43*). In the scenario that we propose, with the staggered potential acting as an order parameter, the broken symmetry is discrete, which is why the occurrence of finite-temperature phase transitions in Bernal multilayers does not conflict with the Mermin-Wagner theorem (*44*), despite all multilayers being effectively 2D.

Despite the success of the proposed model, in the absence of a comprehensive theoretical analysis, important questions remain. A key question is the validity of the minimal tight-binding model because the vast existing literature on graphite suggests that more tight-binding parameters should be included. A recent theoretical analysis of *e-e* interaction at the Hartree-Fock level (*43*) shows that adding hopping terms present in graphite—most notably, the parameter γ_2, responsible for the semimetallic behavior of graphite ($\gamma_2 \sim -20$ meV, usually)—would make the systematic behavior observed in multilayers impossible to reproduce theoretically. At the simplest level, the same holds true for other hopping parameters, such as γ_3, responsible for trigonal warping. Why some hopping terms that must be included to describe graphite do not appear in thin multilayers requires an explanation. A key difference may be that our experiments probe charge-neutral multilayers (i.e., $n \sim 0$), whereas in graphite approximately $n \sim 3 \times 10^{11}$ to 4×10^{11} electrons/cm^2 are present in each individual layer (*45*). It has been established theoretically and experimentally that, as n approaches charge neutrality, large renormalization effects drastically change the hopping parameters in graphene: In monolayers, for instance, the Fermi velocity (and hence, γ_0) is predicted to diverge as $n \to 0$ (*22*, *46*–*48*). Although a thorough analysis of this renormalization process for all hopping parameters is lacking, what is known at this stage implies that there is no compelling reason why the hopping terms used to describe graphite and charge-neutral thin multilayers should be the same.

Irrespective of these details, what is most notable in the experimental findings is the occurrence of a phase transition with increasingly large T_c in graphene multilayers that, until now, may have been expected to behave as bulk graphite. Why precisely T_c increases with increasing thickness or at which thickness this trend breaks down and the behavior of bulk graphite is recovered remains to be understood. At this stage, a thorough microscopic theoretical analysis is required not only to understand in detail the properties of this family of electronic systems but also to disclose its full potential to reveal subtle phenomena emerging from the physics of correlated electrons.

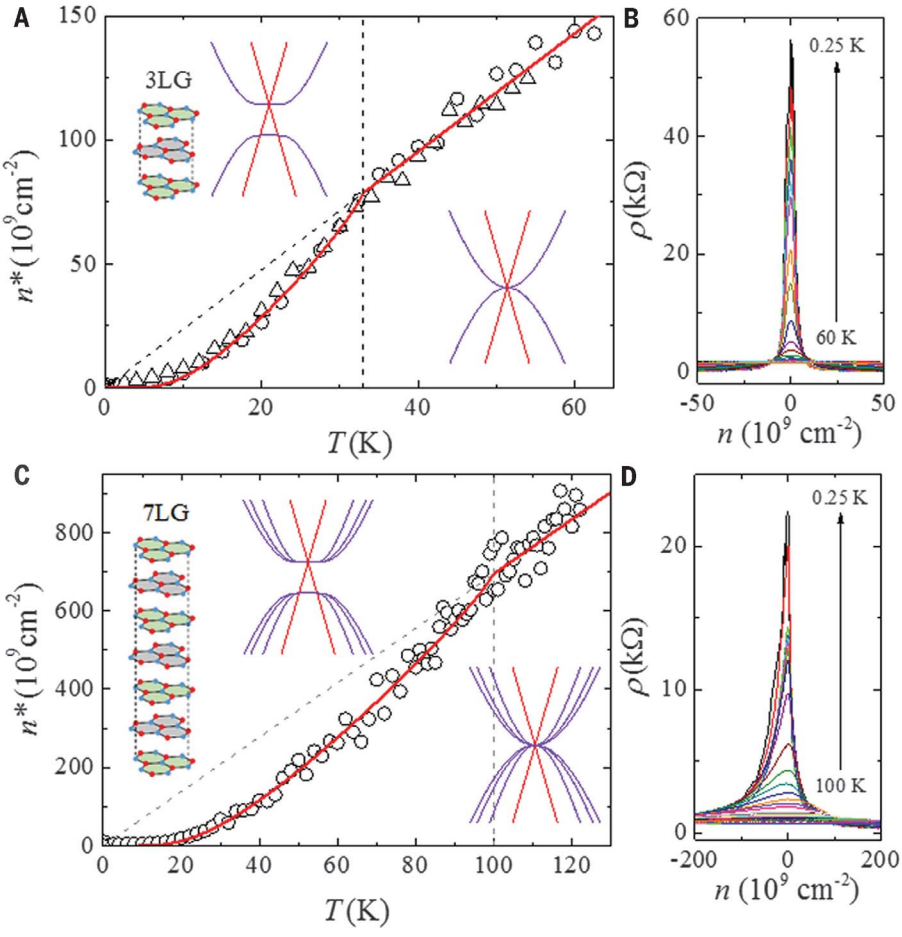


Fig. 3. Electronic phase transition in odd Bernal-stacked multilayers. Shown are the data for (**A** and **B**) 3LG and (**C** and **D**) 7LG graphene. (A) and (C) show that the temperature dependence of $n^*(T)$ is linear at sufficiently high T, just as observed for even layers, because the DOS associated with the Dirac linear band present in odd layers is negligible in the energy range probed by the experiments. For temperatures below T_c (33 K for 3LG and 100 K for 7LG), $n^*(T)$ is suppressed, demonstrating that all the quadratic bands gap out for $T < T_c$. The red continuous curves were calculated by assuming a mean-field gap. In (A), the circles and the triangles represent data obtained from two different devices. (B) and (D) indicate the charge carrier density dependence of the resistivity at different temperatures [from top to bottom, 0.25 to 1.5 K in 0.25-K steps, 2.5 K, 4.0 to 8.0 K in 2-K steps, and 10 to 60 K in 2.5-K steps in (B) and 0.24, 1.64, and 2.5 K and 4.0 to 100 K in 2-K steps in (D)], showing a finite low-temperature conductivity of ~e^2/h around charge neutrality, caused by the presence of the linear band that remains ungapped. The 3LG data in (B), measured on a four-terminal device, are nearly perfectly symmetric in n; for the 7LG device in (D), the asymmetry stems from the two-terminal device configuration.

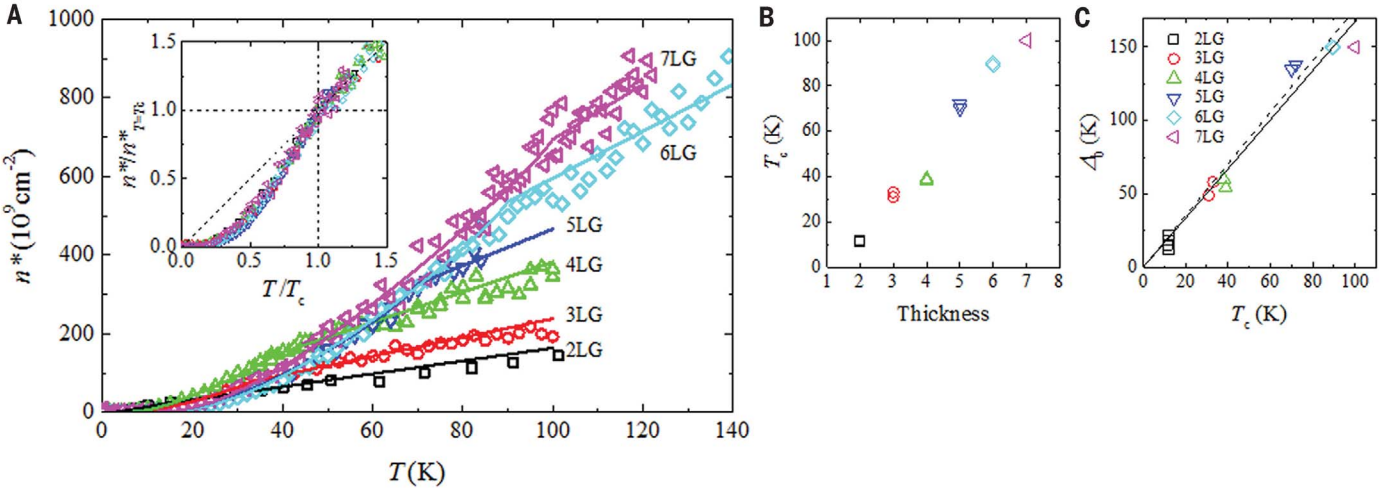


Fig. 4. Evolution with thickness of the phase transition to the broken-symmetry state. (**A**) n^* versus temperature for Bernal-stacked multilayers of all thicknesses, from bi- to heptalayer graphene. The continuous lines of the corresponding color are obtained by fitting the theoretical expression for the temperature dependence of the density of electrons thermally excited from the valence band to the conduction band, $n_{th}(T)$, calculated by using a mean-field temperature dependence for the gap. (Inset) For normalized quantities [i.e., $n^*(T)/n^*(T_c)$ plotted versus T/T_c], all curves collapse on top of one another. For every thickness, only one critical temperature is observed and only one value of the gap is needed to reproduce the data, indicating that a gap opens simultaneously in all quadratic bands. (**B**) T_c increases linearly with increasing thickness, and the gap Δ_0 is proportional to T_c (**C**). The best linear fit (continuous line) is close to the relation expected from mean-field theory, $\Delta_0/k_BT_c = 1.76$ (dashed line).

REFERENCES AND NOTES

1. D. Yoshioka, *The Quantum Hall Effect* (Springer, 2002).
2. J. K. Jain, *Composite Fermions* (Cambridge Univ. Press, 2007).
3. Z. F. Ezawa, *Quantum Hall Effects: Field Theoretical Approach and Related Topics* (World Scientific, 2008).
4. E. Fradkin, *Field Theories of Condensed Matter Physics* (Cambridge Univ. Press, 2013).
5. D. C. Tsui, H. L. Stormer, A. C. Gossard, *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
6. R. Willett *et al*., *Phys. Rev. Lett.* **59**, 1776–1779 (1987).
7. J. P. Eisenstein, A. H. Macdonald, *Nature* **432**, 691–694 (2004).
8. P. Maher *et al*., *Nat. Phys.* **9**, 154–158 (2013).
9. A. F. Young *et al*., *Nature* **505**, 528–532 (2014).
10. D.-K. Ki, V. I. Fal'ko, D. A. Abanin, A. F. Morpurgo, *Nano Lett.* **14**, 2135–2139 (2014).
11. J. Falson *et al*., *Nat. Phys.* **11**, 347–351 (2015).
12. A. A. Zibrov *et al*., *Nature* **549**, 360–364 (2017).
13. J. I. A. Li *et al*., *Science* **358**, 648–652 (2017).
14. R. T. Weitz, M. T. Allen, B. E. Feldman, J. Martin, A. Yacoby, *Science* **330**, 812–816 (2010).
15. F. Freitag, J. Trbovic, M. Weiss, C. Schönenberger, *Phys. Rev. Lett.* **108**, 076602 (2012).
16. J. Velasco Jr. *et al*., *Nat. Nanotechnol.* **7**, 156–160 (2012).
17. W. Bao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10802–10805 (2012).
18. A. L. Grushina *et al*., *Nat. Commun.* **6**, 6419 (2015).
19. Y. Nam, D.-K. Ki, M. Koshino, E. McCann, A. F. Morpurgo, *2D Mater.* **3**, 045014 (2016).
20. K. I. Bolotin *et al*., *Solid State Commun.* **146**, 351–355 (2008).
21. X. Du, I. Skachko, A. Barker, E. Y. Andrei, *Nat. Nanotechnol.* **3**, 491–495 (2008).
22. D. C. Elias *et al*., *Nat. Phys.* **7**, 701–704 (2011).
23. W. Bao *et al*., *Nat. Phys.* **7**, 948–952 (2011).
24. R. Nandkishore, L. Levitov, *Phys. Rev. Lett.* **104**, 156803 (2010).
25. Y. Lemonik, I. L. Aleiner, C. Toke, V. I. Fal'ko, *Phys. Rev. B* **82**, 201408 (2010).
26. F. Zhang, H. Min, M. Polini, A. H. MacDonald, *Phys. Rev. B* **81**, 041402 (2010).
27. F. Zhang, J. Jung, G. A. Fiete, Q. Niu, A. H. MacDonald, *Phys. Rev. Lett.* **106**, 156801 (2011).
28. F. Zhang, A. H. MacDonald, *Phys. Rev. Lett.* **108**, 186804 (2012).
29. See the supplementary materials.
30. A. S. Mayorov *et al*., *Science* **333**, 860–863 (2011).
31. A. S. Mayorov *et al*., *Nano Lett.* **12**, 4629–4634 (2012).
32. J. Crossno *et al*., *Science* **351**, 1058–1061 (2016).
33. P. San-Jose, R. V. Gorbachev, A. K. Geim, K. S. Novoselov, F. Guinea, *Nano Lett.* **14**, 2052–2057 (2014).
34. J. Li, I. Martin, M. Buttiker, A. F. Morpurgo, *Nat. Phys.* **7**, 38–42 (2011).
35. M. J. Zhu *et al*., *Nat. Commun.* **8**, 14552 (2017).
36. F. Guinea, A. H. Castro Neto, N. M. R. Peres, *Phys. Rev. B* **73**, 245426 (2006).
37. B. Partoens, F. M. Peeters, *Phys. Rev. B* **75**, 193402 (2007).
38. M. Koshino, T. Ando, *Phys. Rev. B* **76**, 085425 (2007).
39. M. Koshino, E. McCann, *Phys. Rev. B* **81**, 115315 (2010).
40. S. Berciaud, M. Potemski, C. Faugeras, *Nano Lett.* **14**, 4548–4553 (2014).
41. Y. Henni *et al*., *Nano Lett.* **16**, 3710–3716 (2016).
42. C. Yoon, Y. Jang, J. Jung, H. Min, *2D Mater.* **4**, 021025 (2017).
43. M. Koshino, K. Sugisawa, E. McCann, *Phys. Rev. B* **95**, 235311 (2017).
44. N. D. Mermin, H. Wagner, *Phys. Rev. Lett.* **17**, 1133–1136 (1966).
45. N. B. Brandt, S. M. Chudinov, Y. G. Ponomarev, *Semimetals 1: Graphite and Its Compounds* (Elsevier, 1988).
46. J. González, F. Guinea, M. A. H. Vozmediano, *Nucl. Phys. B* **424**, 595–618 (1994).
47. S. Das Sarma, E. H. Hwang, W.-K. Tse, *Phys. Rev. B* **75**, 121406 (2007).
48. M. Polini, R. Asgari, Y. Barlas, T. Pereg-Barnea, A. H. MacDonald, *Solid State Commun.* **143**, 58–62 (2007).
49. D.-K. Ki, A. F. Morpurgo, *Nano Lett.* **13**, 5165–5170 (2013).
50. A. Morpurgo, Y. Nam, D. Ki, D. Soler-Delgado, A family of finite-temperature electronic phase transitions in graphene multilayers, Version 1, Harvard Dataverse (2018).

ACKNOWLEDGMENTS

We gratefully acknowledge A. Ferreira for continued technical support of the experiments and T. Giamarchi for useful discussions. **Funding:** Financial support from the Swiss National Science Foundation, the NCCR QSIT, and the EU Graphene Flagship Project is also gratefully acknowledged. **Author contributions:** Y.N., D.-K.K., and D.S.-D. fabricated the devices and performed the electrical measurements. Y.N. and D.-K.K. analyzed the data under the supervision of A.F.M. All authors discussed experimental results and their interpretation. A.F.M. wrote the manuscript with input from all authors. **Competing interests:** There are no competing financial interests. **Data and materials availability:** The dataset used in this paper is available at Harvard Dataverse (*50*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/324/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (*51*–*62*)

19 December 2017; accepted 28 August 2018
10.1126/science.aar6855

GEOPHYSICS

Shear properties of Earth's inner core constrained by a detection of *J* waves in global correlation wavefield

Hrvoje Tkalčić* and Thanh-Son Phạm

Seismic *J* waves, shear waves that traverse Earth's inner core, provide direct constraints on the inner core's solidity and shear properties. However, these waves have been elusive in the direct seismic wavefield because of their small amplitudes. We devised a new method to detect *J* waves in the earthquake coda correlation wavefield. They manifest through the similarity with other compressional core-sensitive signals. The inner core is solid, but relatively soft, with shear-wave speeds and shear moduli of 3.42 ± 0.02 kilometers per second and 149.0 ± 1.6 gigapascals (GPa) near the inner core boundary and 3.58 ± 0.02 kilometers per second and 167.4 ± 1.6 GPa in Earth's center. The values are 2.5% lower than the widely used Preliminary Earth Reference Model. This provides new constraints on the dynamical interpretation of Earth's inner core.

Earth's inner core was proposed more than 80 years ago to explain the arrival of compressional waves in places that were impossible to predict on the basis of the assumption of single-layered core (*1*). The hypothesis of a solid inner core is the result of the liquid-solid phase change in iron at high pressure (*2*), which implies the existence of shear waves in the inner core (seismic *J* phase). Previous claims of body-wave observations (*3*–*7*) featured bursts of energy around the time predicted by the spherically symmetric Preliminary Reference Earth Model (PREM) (*8*). The rigidity of the inner core in PREM is mainly constrained by Earth's free oscillations (*9*). However, *J* waves have remained elusive, because it was demonstrated that routine observations of *PKJKP* waves are extremely unlikely in the seismic wavefield at periods greater than 10 s (*10*). We used advances in earthquake coda cross-correlation (*11*) and identified the presence of the *J* phase. We identified arrivals of *PKJKP* waves in a correlation pair with another core phase, *PKIKPPKIKP* (hereafter referred to as *I2*), over a range of angular distances. This allowed us to determine, with high precision, the shear-wave speed reduction of 2.5 ± 0.5% relative to the reference model PREM, which corresponds to a shear-wave speed of 3.42 ± 0.02 km/s and shear modulus of 149.0 ± 1.6 GPa near the inner core boundary and 3.58 ± 0.02 km/s and 167.4 ± 1.6 GPa in Earth's center. This is evidence for a soft inner core and explains the absence of *PKJKP* waves in the seismic wavefield. Thus, our findings have a range of implications for the structure and dynamics of the inner core (*12*, *13*).

We have recently shown that the features emerging in the stacked cross-correlations (also known as Earth's cross-correlogram or global correlogram) do not correspond to structural Green's functions (response of Earth's structure between two receivers) but rather emerge owing to similarities of two or more seismic phases that arrive at recorders with the same slowness and share a common subset of propagation legs (*11*). Earth's correlation wavefield is thus a manifestation of similarity among seismic phases in the direct seismic wavefield (*14*). The similarity between the weak signals is a more efficient mechanism of detection than the identification of weak signals in the noisy seismic wavefield. Consequently, the correlation wavefield features "exotic seismic phases" that traverse Earth's inner core and deep parts of Earth, for example, multiple inner-core phases: *I3*, *I4*, and *I5*. Additionally, there are correlation phases that do not have correspondences in the direct seismic wavefield, for example, the phase *cS-cP*, which provides additional insights into the seismic wavefield (*11*, *14*).

We applied the above principles using global data (fig. S1) to conduct a systematic search for the presence of a *PKJKP* signal over a number of correlation pairs of seismic phases in which one of the phases is *PKJKP* (figs. S3 and S4). The rationale for this search is that the similarity of the reticent *PKJKP* signals with other, more prominent phases such as *PKiKP*, *PKIKP*, or *PKIKPPKIKP* (Fig. 1) will manifest itself as one of the features in the global correlogram. We successfully fit the travel-time prediction of three correlation phases: *PKIKPPKIKP-PKJKP* (*I2-PKJKP*), *PKIKP-PKJKP*, and *PKiKP-PKJKP*, with features in the global correlogram synthetized by the same Earth model (*14*). Because we observed *I2-PKJKP* near the angular distance 0°, the prominence occurs because of contributions from all azimuths (*15*). *PKIKP-PKJKP* and *PKiKP-PKJKP* are weaker because there is no such focusing effect in the 120° to 150° angular distance range (Fig. 2C).

Research School of Earth Sciences, The Australian National University, Canberra, ACT 2601, Australia.
***Corresponding author. Email: hrvoje.tkalcic@anu.edu.au**

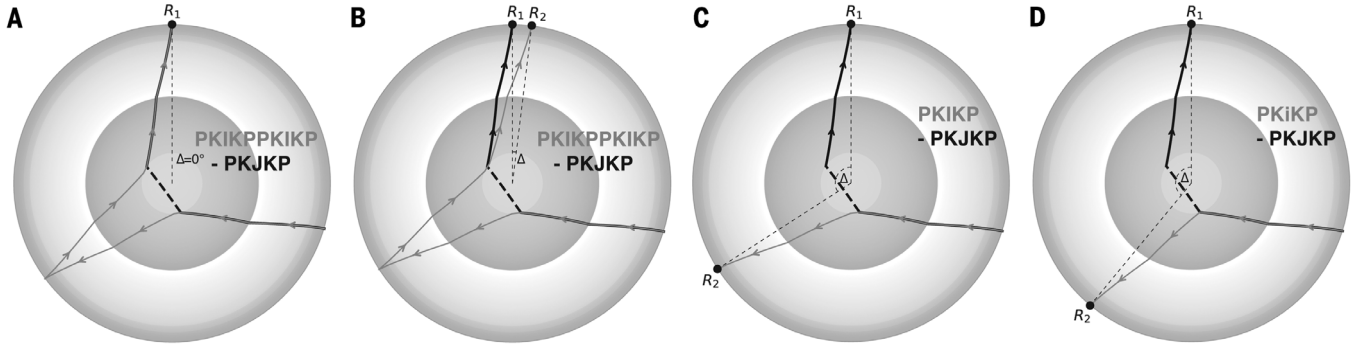


Fig. 1. Seismic phases, ray paths, and receiver geometry. (**A** to **D**) Cross sections of Earth illustrating the geometry of *PKJKP* waves and other compressional waves used to detect *PKJKP* in Earth's correlation wavefield. Shear-wave leg through the inner core is shown with a dashed line in all panels. The arrow indicates the direction of seismic waves. In (A), *PKJKP* is shown in black and *PKIKPPKIKP* in gray. *PKIKPPKIKP* is also referred to as *I2*. R_1 is the receiver at the angular distance Δ = 0° used to perform a cross-correlation. (B) is the same as (A), but the receivers are now separated by an angular distance Δ ≠ 0°. In (C), the two receivers are now separated by the angular distance Δ = 120°. PKJKP is shown in black and *PKIKP* in gray. (D) is similar to (C), but instead of *PKIKP* waves that traverse the inner core, the waves that correlate with *PKJKP* (black) are *PKiKP* waves that reflect from the inner core (gray). *PKJKP* is shown in black and *PKiKP* in gray.

We computed synthetic global correlograms for a crude parameter-space search over a range of shear-wave speed in the inner core (Fig. 2). Our synthesized global correlograms (Fig. 2) are, overall, insensitive to forced changes in the inner core shear-wave speed. We detected a number of prominent and stable features that correspond to the correlation phases (light blue curves). However, a clear move out (a shift in the time–angular distance domain) of *PKIKP-PKJKP*, *PKiKP-PKJKP*, and *I2-PKJKP* is evident with the changing inner core shear-wave speed. We compare these to the theoretically predicted arrivals that are based on the model PREM (orange curves). At the 10% reduction with respect to PREM (Fig. 2A), the *I2-PKJKP* cusp is visible, but there is a negative time offset with respect to the PREM prediction and the prominent phase *PcP** (*14*) that starts about a minute later at 15° of angular distance. With increased inner core shear-wave speed, the cusp shifted toward later times because *PKJKP* waves arrived earlier at the receiver, so that the time difference between *PKJKP* and later *I2* arrivals increases.

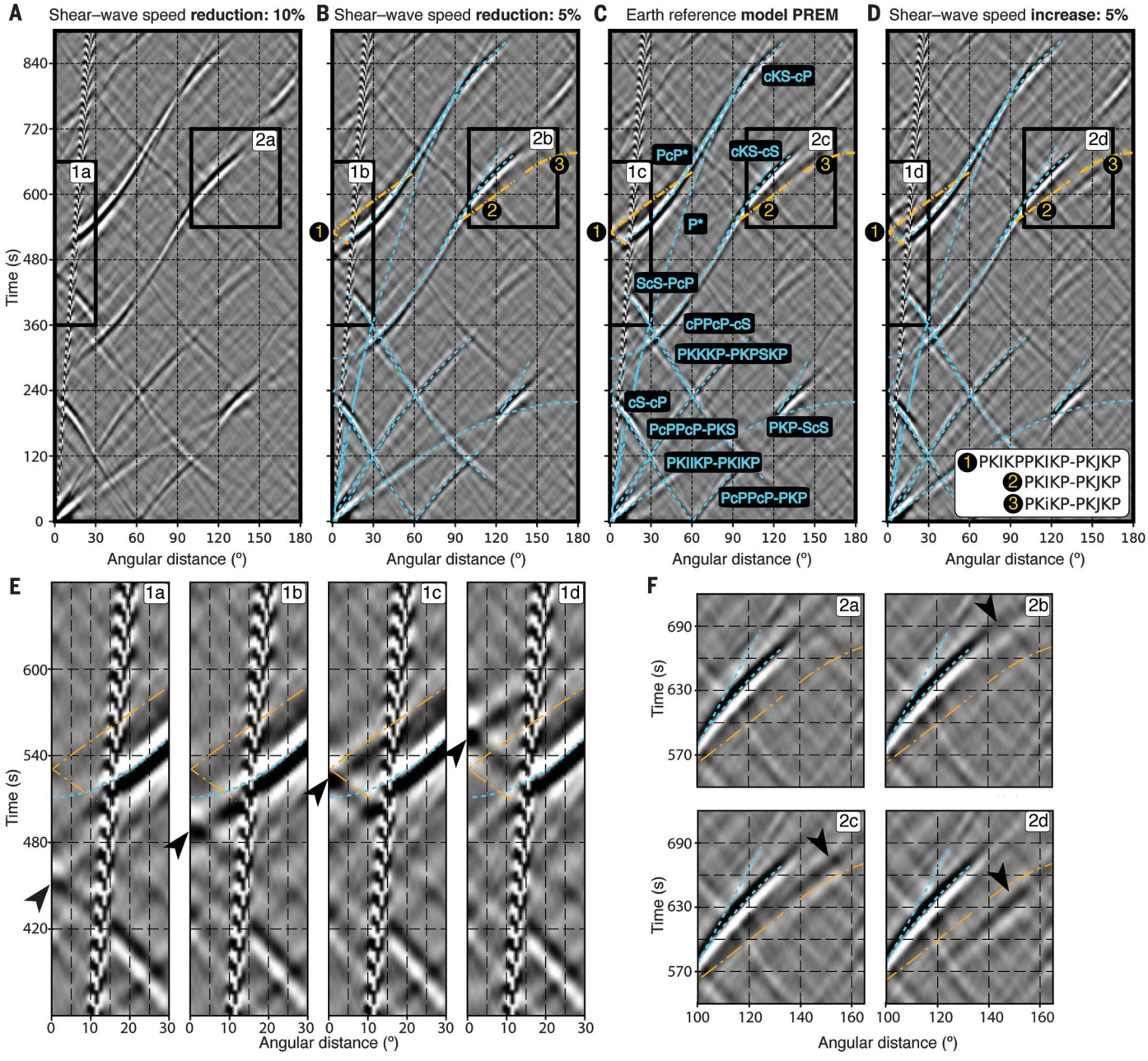


Fig. 2. Detection of *PKJKP* in synthetic correlation wavefield. (**A** to **F**) Synthetic global correlograms for different assumptions of shear-wave speed in Earth's inner core relative to the spherically symmetric Earth model PREM (*8*): (A) 10% reduction, (B) 5% reduction, (C) the PREM value (0% change), and (D) 5% increase. Windows 1 and 2 (indicated by the black rectangles) in (A) to (D) focus on the correlation phases of interest (indicated by yellow numbers 1, 2, and 3) and are enlarged in sections (E) and (F). Positive amplitudes are in white shades, and negative amplitudes are in black shades; the intensity of the black or white indicates the strength. Blue lines represent theoretically predicted features in the correlation wavefield that are insensitive to the inner core shear-wave speed; for clarity, they are labeled in (C) and omitted elsewhere. To enhance the clarity of all features, the theoretical curves are not shown in (A). Orange lines are theoretical predictions of *PKIKPPKIKP-PKJKP* (also called *I2-PKJKP*), *PKiKP-PKJKP*, and *PKIKP-PKJKP* according to the model PREM. For the ray geometry of these phases, see Fig. 1. Black arrows in (E) and (F) indicate the changing position of the features sensitive to the inner core shear-wave speed.

At the 5% increase, the cusp is more pronounced and is positioned upward with respect to the PREM prediction and the prominent phase *PcP** (Fig. 2D).

Our observed correlation wavefield is generally depleted in detail in comparison with the simulated correlation wavefield (*11*, *16*) (Fig. 3). However, the level of similarity that we reproduced through the seismic wavefield propagation in the spherically symmetric Earth model PREM in this frequency range (15 to 50 s) is notable. This attests to the generally well-constrained radial Earth model from long-period seismic observations. The *I2-PKJKP* cusp is a prominent feature that we identified in the observed correlation wavefield, whereas we cannot discern the *PKIKP-PKJKP* and *PKiKP-PKJKP*. We compared the *I2-PKJKP* cusp offset toward early times to the PREM prediction corrected by the dispersion relationship for the central period T = 23.1 s (*17*). Our fine parameter-space search simulations (*14*) yielded the best Earth model that fit the observed correlation wavefield, in which the inner core has a 2.5 ± 0.5% reduced shear-wave speed relative to the PREM values. We based our estimate on visual comparisons of the cusp positions and of negative peaks in the slant stacks produced from our observed and synthetized correlation wavefield (Fig. 3D). The width of the fringes of the features in the correlograms prevents a better estimate of their position at the present time. More-quantitative approaches included the computation of the difference between the observed and synthesized correlograms but did not yield a more conclusive result or a better resolution.

We compared our estimated shear-wave parameters with some previous estimates (Table 1). At the period T = 1 s and using the PREM parametric form, the shear-wave speed is 3.42 ± 0.02 km/s at the inner core boundary and 3.58 ± 0.02 km/s in the center of Earth. These values are in reasonable agreement with normal mode estimates and with some previous estimates from body-wave observations (Table 1). From these values and the relationship between the shear-wave speed and shear modulus and density, we derived the shear modulus of 149.0 ± 1.6 GPa at the inner core boundary and 167.4 ± 1.6 GPa in the center of Earth. If the compressional speed is taken from PREM (*8*), we can infer the inner core Poisson's ratio of about 0.45. The reduction from the inner core boundary (0.449) to the center (0.444) is smaller than what we can reliably determine with our current precision. In addition, we simulated the amplitudes in the observed correlogram by changing the shear attenuation or the quality factor (Q_μ, an inverse of shear attenuation) in the inner core (fig. S5). This yielded an interval of values for Q_μ, ranging from the PREM values down to the 50% reduction relative to PREM for the idealized and unrealistic extreme case in which the synthetic experiment does not capture the true effects of attenuation. The effect of changing the shear-wave speed and Q structure of the inner core on the inner-core sensitive modes (when the full coupling theory is considered) can be substantial and must be accounted for in resolving the existing normal mode theory-observation misfits. This is in addition to Earth's heterogeneous structure and anisotropy. Eigenfrequencies and Q values of the modes can be strongly affected, and some modes can even exchange their identity (*18*). Our newly estimated shear-wave speed reduction of 2.5% with respect to the PREM model and a hypothetical reduction of 25% in Q_μ would lower the eigenfrequency of the ${}_{10}S_2$ mode by about 3%.

One hypothesis proposed to explain the high attenuation and low shear-wave speed is inner core melt pockets (*19*–*21*). However, although compaction during solidification constrains the upper limit of volume fraction (*22*), theoretical considerations otherwise leave this value unconstrained. A melt volume fraction of more than 10% (*23*) explains the inner core seismological parameters reported in PREM (*8*). Following the same line of argument, our values likely require an even higher portion of melt. Thin films of melt might be consistent with our parameters as the geometry and distribution of melt also affects the shear and attenuation properties.

The hypothesis that premelting effects are prominent near the inner core boundary (*24*) makes determining whether the shear-wave speed reduction occurs only in the top or in the bulk of the inner core relevant. We experimented with the shear-wave speed profiles in Earth's inner core to determine whether we could simulate the timing of the *I2-PKJKP* cusp equally well by reducing the shear-wave speed only in the uppermost part of the inner core. We found that a shear-wave speed reduction of ~35% with respect to the PREM model confined to the uppermost 50 km of the inner core can explain the timing of the *I2-PKJKP* cusp (*14*). However, apart from making the *I2-PKJKP* cusp too weak, this substantial reduction in shear-wave speed cannot predict the correlation phase *PKIIKP-PKIKP* (labeled in supplementary animation S3 and in Fig. 2C), which is a prominent feature in our observed and the best-fit simulated correlograms (Fig. 3). Additionally, *PKIIKP* waves have been observed and their amplitudes studied in conjunction with the shear-wave speed properties of the inner core (*25*). On the basis of this simulation, we thus eliminated the possibility that the shear-wave speed reduction is necessary only in the uppermost inner core. The bulk of the inner core is required to have a reduced shear-wave speed of 2.5% on average. A radial variation of the shear-wave speed is possible, but the exact distribution is hampered by the resolution combined with the non-uniqueness of the problem.

In contrast to the hypotheses above, intrinsic properties of polycrystalline iron at high pressures and temperatures have been proposed to explain the properties of the inner core (*26*, *27*). The ab initio estimates of inner core shear-wave speed (*23*) are 30% larger than the seismological observations. However, when ab initio molecular dynamics calculations account for the shear-wave reducing effects of polycrystallinity, defects, and grain boundaries, they become much more similar to the seismological values. The diffusion of body-centered cubic phase of iron atoms in solid state has been proposed to explain low shear modulus (*28*). Although this study is based on calculations for pure iron, the diffusion mechanism makes the

Table 1. Summary of previous and current results. Shear-wave speed and rigidity of the inner core from different seismological studies. "ICB" indicates measurements at the inner core boundary, and "center" indicates measurements at the center of Earth. For the shear-wave speed data, values that do not have these labels are average values for the inner core in which the shear-wave speed was not explicitly parameterized in terms of Earth's radius. Blank cells indicate that shear-wave speed and rigidity were not explicitly determined in those studies.

Study and method	Shear-wave speed at T = 1 s (km/s)	Rigidity (shear modulus) (GPa)
Dziewonski and Anderson (1981) (*8*)	ICB: 3.4629	ICB: 156.7
Global Earth model PREM	Center: 3.6245	Center: 176.1
Kennett *et al*. (1995) (*36*)	ICB: 3.504	ICB: 156.0
Global Earth model ak135	Center: 3.668	Center: 175.1
Julian *et al*. (1972) (*3*)	2.95 ± 0.1	
Short period body waves	3.62 (reinterpreted) (*4*)	
Wookey and Helffrich (2008) (*7*)	Similar to PREM	
Short period body waves		
Okal and Cansi (1998) (*4*)	3.65	
Intermediate period body waves		
Cao *et al*. (2005) (*6*)	~1.5% faster than that for PREM	
Intermediate period body waves		
Deuss *et al*. (2000) (*5*)	3.6	
Long-period body waves		
This study	ICB: 3.42 ± 0.02	ICB: 149.0 ± 1.6
Earthquake coda cross-correlation	Center: 3.58 ± 0.02	Center: 167.4 ± 1.6

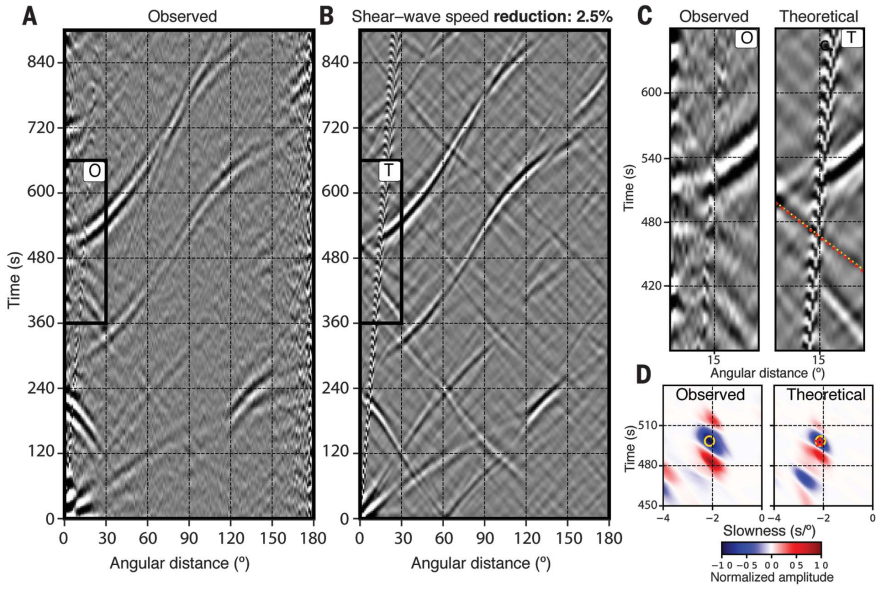


Fig. 3. Observed and synthetic correlograms and the *I2-PKJKP* cusp. (**A** and **B**) A comparison between (A) the observed global correlogram and (B) the best-fit simulated global correlogram for the central period of 23.1 s. The best-fit simulation uses the PREM model with shear-wave speed in the bulk of the inner core reduced by 2.5% [we used a smooth parameter-space search, best viewed as an animation (*14*)]. Compare with the cusp positions in Fig. 2. (**C**) Enlargements of windows O (observed) and T (theoretical) in (A) and (B), focused on the *I2-PKJKP* cusp. (**D**) The observed and simulated (theoretical) slant stacks. The yellow circle corresponds to the lower branch of the *I2-PKJKP* cusp in the slowness-time domain, and the yellow dotted line corresponds to the same in the travel-time domain. The red star in (D) and red dotted line in (C) are the values based on the best-fit Earth model with the 2.5% reduction of shear-wave speed in the inner core.

inner core soft with a very low resistance to shear without requiring the existence of trapped melt to explain seismological observations. Although our study does not exclude any of the above possibilities, it provides seismological evidence for a soft inner core (a less stiff inner core, in terms of its elastic behavior) from a different class of observations.

The relationship between our seismically obtained shear-wave speed and the viscosity is not straightforward. Viscosity depends both on material properties and the dynamics. The mineral physics estimates of viscosity (*29*) are lower than the geodynamic ones (*30*). The mineralogical models generally do not include the effects of microstructures (*31*), which may account for the discrepancy. Our values are consistent with a soft inner core. A dynamically soft core helps explain the lack of gravitational torque–driven length-of-day changes that would be expected with a rigid core combined with the fluctuation in rotation speed inferred from core flow and seismological detection of core rotation (*13*). Moreover, a soft inner core allows for deformation in both the polar and equatorial directions, consistent with observations related to polar motion or nutations (*32*).

Addition of data from local and regional networks will increase the number of station pairs at short angular distances and thus improve the quality of the global correlogram (fig. S1B). Further proliferation of seismic recorders in remote areas of Earth will increase the number of station pairs at antipodal distances. This, in turn, will improve the estimates of the time and amplitude of *J* phase–related features. Determining the frequency dependence of shear wave speeds will also place more constraints on attenuation and viscosity in the inner core. We expect that a shift toward higher frequencies is possible by considering early coda time windows in which the seismic phases are relatively less attenuated owing to their shorter paths through Earth. Detection of *J* waves confirms that Earth's inner core is solid, although elastically less stiff than previous estimates. This inference represents an advance in our understanding of structure and dynamics of the inner core—Earth's deepest time capsule that has been probed by the global correlation wavefield.

REFERENCES AND NOTES

1. I. Lehman, *Publications du Bureau Central séismologique international* **A14**, 87–115 (1936).
2. A. F. Birch, *Am. J. Sci.* **238**, 192–211 (1940).
3. B. R. Julian, D. Davies, R. M. Sheppard, *Nature* **235**, 317–318 (1972).
4. E. A. Okal, Y. Cansi, *Earth Planet. Sci. Lett.* **164**, 23–30 (1998).
5. A. Deuss, J. H. Woodhouse, H. Paulssen, J. Trampert, *Geophys. J. Int.* **142**, 67–73 (2000).
6. A. Cao, B. Romanowicz, N. Takeuchi, *Science* **308**, 1453–1455 (2005).
7. J. Wookey, G. Helffrich, *Nature* **454**, 873–876 (2008).
8. A. M. Dziewonski, D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
9. A. M. Dziewonski, F. Gilbert, *Nature* **234**, 465–466 (1971).
10. P. M. Shearer, C. A. Rychert, Q. Liu, *Geophys. J. Int.* **185**, 1379–1383 (2011).
11. T.-S. Phạm, H. Tkalčić, M. Sambridge, B. L. N. Kennett, *Geophys. Res. Lett.* **45**, 3035–3042 (2018).
12. R. Deguen, *Earth Planet. Sci. Lett.* **333–334**, 211–225 (2012).
13. H. Tkalčić, *The Earth's Inner Core: Revealed by Observational Seismology* (Cambridge Univ. Press, 2017).
14. Materials and methods are available as supplementary materials.
15. C. Sens-Schönfelder, R. Snieder, S. C. Stähler, *Geophys. Res. Lett.* **42**, 7483–7489 (2015).
16. P. Boué, P. Poli, M. Campillo, P. Roux, *Earth Planet. Sci. Lett.* **391**, 137–145 (2014).
17. H. Kanamori, D. L. Anderson, *Rev. Geophys.* **15**, 105–112 (1977).
18. J. Andrews, A. Deuss, J. Woodhouse, *Geophys. J. Int.* **167**, 204–212 (2006).
19. D. J. Doornbos, *Geophys. J. Int.* **38**, 397–415 (1974).
20. D. R. Fearn, D. E. Loper, P. H. Roberts, *Nature* **292**, 232–233 (1981).
21. S. C. Singh, M. A. Taylor, J. P. Montagner, *Science* **287**, 2471–2474 (2000).
22. I. Sumita, S. Yoshida, M. Kumazawa, Y. Hamano, *Geophys. J. Int.* **124**, 502–524 (1996).
23. L. Vočadlo, in *New Frontiers in Integrated Solid Earth Sciences*, S. A. P. L. Cloetingh, J. Negendank, Eds. (Springer, 2009), pp. 397–412.
24. B. Martorell, L. Vočadlo, J. Brodholt, I. G. Wood, *Science* **342**, 466–468 (2013).
25. V. F. Cormier, *Geophys. Res. Lett.* **42**, 7459–7466 (2015).
26. I. Jackson, J. D. Fitz Gerald, H. Kokkonen, *J. Geophys. Res. Solid Earth* **105**, 23605–23634 (2000).
27. A. B. Belonoshko *et al.*, *Science* **316**, 1603–1605 (2007).
28. A. B. Belonoshko *et al.*, *Nat. Geosci.* **10**, 312–316 (2017).
29. J. A. Van Orman, *Geophys. Res. Lett.* **31**, L20606 (2004).
30. B. A. Buffett, *Nature* **388**, 571–573 (1997).
31. M. I. Bergman, J. Yu, D. J. Lewis, G. K. Parker, *J. Geophys. Res. Solid Earth* **123**, 189–203 (2018).
32. L. Koot, M. Dumberry, *Earth Planet. Sci. Lett.* **308**, 343–349 (2011).
33. Data are available from the IRIS/IDA Seismic Network (https://doi.org/10.7914/SN/II), operated by the Scripps Institution of Oceanography.
34. Data are available from the Global Seismograph Network (https://doi.org/10.7914/SN/IU), operated by the Albuquerque Seismological Laboratory and the USGS.
35. Data are available from the Australian National Seismograph Network (www.fdsn.org/networks/detail/AU/), operated by Geoscience Australia.
36. B. L. N. Kennett, E. R. Engdahl, R. Buland, *Geophys. J. Int.* **122**, 108–124 (1995).

ACKNOWLEDGMENTS

The authors thank three anonymous reviewers for their constructive suggestions on how to improve the manuscript. H.T. is grateful to J. Badro, A. Belonoshko, M. Bergman, G. Helffrich, J. Hernlund, M. Dumberry, I. Jackson, S.-I. Karato, and S. Pachhai for useful discussions. **Funding:** The synthetic experiment was performed on the ANU Terrawulf cluster, a computational facility developed with support from the AuScope initiative. AuScope Ltd. is funded under the National Collaborative Research Infrastructure Strategy (NCRIS), an Australian Commonwealth Government Program. H.T.'s travel was partially supported through the AuScope Excellence in Research Achievement travel bursary award. **Author contributions:** H.T. conceived and supervised the study. Both authors discussed the results and determined the course of the study based on methodological developments achieved during T.-S.P.'s Ph.D. work. The codes for data processing, figures, and data analysis were produced by T.-S.P.; and data analysis and modeling were undertaken by T.-S. P., initiated by and in consultation with H.T., as a part of T.-S.P.'s Ph.D. work. H.T. wrote the manuscript with some edits provided by T.-S.P., edited figures and supplementary material, and interpreted the results in the context of current intellectual framework surrounding the inner core research. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms, related metadata, and/or derived products used in this study. Waveform data used in this study are available from (*33–35*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/329/suppl/DC1
Materials and Methods
Figs. S1 to S5
References (*37–45*)
Animations S1 to S3

13 July 2018; accepted 10 September 2018
10.1126/science.aau7649

TOPOLOGICAL MATTER

Evidence for Majorana bound states in an iron-based superconductor

Dongfei Wang[1,2]*, Lingyuan Kong[1,2]*, Peng Fan[1,2]*, Hui Chen[1], Shiyu Zhu[1,2], Wenyao Liu[1,2], Lu Cao[1,2], Yujie Sun[1,3], Shixuan Du[1,3,4], John Schneeloch[5], Ruidan Zhong[5], Genda Gu[5], Liang Fu[6], Hong Ding[1,2,3,4]†, Hong-Jun Gao[1,2,3,4]†

The search for Majorana bound states (MBSs) has been fueled by the prospect of using their non-Abelian statistics for robust quantum computation. Two-dimensional superconducting topological materials have been predicted to host MBSs as zero-energy modes in vortex cores. By using scanning tunneling spectroscopy on the superconducting Dirac surface state of the iron-based superconductor $FeTe_{0.55}Se_{0.45}$, we observed a sharp zero-bias peak inside a vortex core that does not split when moving away from the vortex center. The evolution of the peak under varying magnetic field, temperature, and tunneling barrier is consistent with the tunneling to a nearly pure MBS, separated from nontopological bound states. This observation offers a potential platform for realizing and manipulating MBSs at a relatively high temperature.

Majorana bound states (MBSs) in condensed-matter systems have attracted tremendous interest owing to their non-Abelian statistics and potential applications in topological quantum computation (*1*, *2*). A MBS is theoretically predicted to emerge as a spatially localized zero-energy mode in certain *p*-wave topological superconductors in one and two dimensions (*3*, *4*). Although the material realization of such *p*-wave superconductors has remained elusive, other platforms for MBSs have recently been proposed, using heterostructures between conventional *s*-wave superconductors and topological insulators (*5*), nanowires (*6*–*8*), quantum anomalous Hall insulators (*9*), or atomic chains (*10*), where the proximity effect on a spin-nondegenerate band creates a superconducting (SC) topological state. Various experimental signatures of MBSs (*11*–*14*) or Majorana chiral modes (*15*) have been observed in these heterostructures, but clear detection and manipulation of MBSs are often hindered by the contribution of nontopological bound states and complications of material interface.

Very recently, using high-resolution angle-resolved photoemission spectroscopy (ARPES), a potential platform for MBSs was discovered in the bulk superconductor $FeTe_{0.55}Se_{0.45}$, with a SC transition temperature T_c = 14.5 K and a simple crystal structure (Fig. 1A). Because of the topological band inversion between the p_z and d_{xz}/d_{yz} bands around the $\bar{\Gamma}$ point (*16*, *17*) and the multiband nature (Fig. 1B), this single material naturally has a spin-helical Dirac surface state, with an induced full SC gap and a small Fermi energy (Fig. 1C) (*18*); these properties would create favorable conditions for observing a pure MBS (*5*) that is isolated from other nontopological Caroli–de Gennes–Matricon bound states (CBSs) (*19*, *20*). The combination of high-T_c superconductivity and Dirac surface states in a single material removes the challenging interface problems in previous proposals and offers clear advantages for the detection and manipulation of MBSs.

Motivated by the above considerations, we carried out a high-resolution scanning tunneling microscopy/spectroscopy (STM/S) experiment on the surface of $FeTe_{0.55}Se_{0.45}$, which has a good atomic resolution that reveals the lattice formed by Te/Se atoms on the surface (Fig. 1D). We started with a relatively low magnetic field of 0.5 T along the *c* axis at a low temperature of 0.55 K, with a clear observation of vortex cores in Fig. 1E. At the vortex center, we observed a strong zero-bias peak (ZBP) with a full width at half maximum (FWHM) of 0.3 meV and an amplitude of 2 relative to the intensity just outside the gapped region. Outside of the vortex core, we clearly observed a SC spectrum with multiple gap features, similar to the ones observed in previous STM studies on the same material (*21*, *22*). These different SC gaps correspond well with the SC gaps on different Fermi surfaces of this material observed in previous ARPES studies (table S1) (*23*, *24*). A similar ZBP was reported previously (*22*).

We next demonstrate in Fig. 2 and fig. S4 (*24*) that across a large range of magnetic fields, the observed ZBP does not split when moving away from a vortex center. It can be clearly seen from Fig. 2, A to D, that the ZBP remains at the zero energy, while its intensity fades away when moving away from the vortex center. The nonsplit ZBP contrasts sharply with the split ZBP originating from CBS observed in conventional superconductors

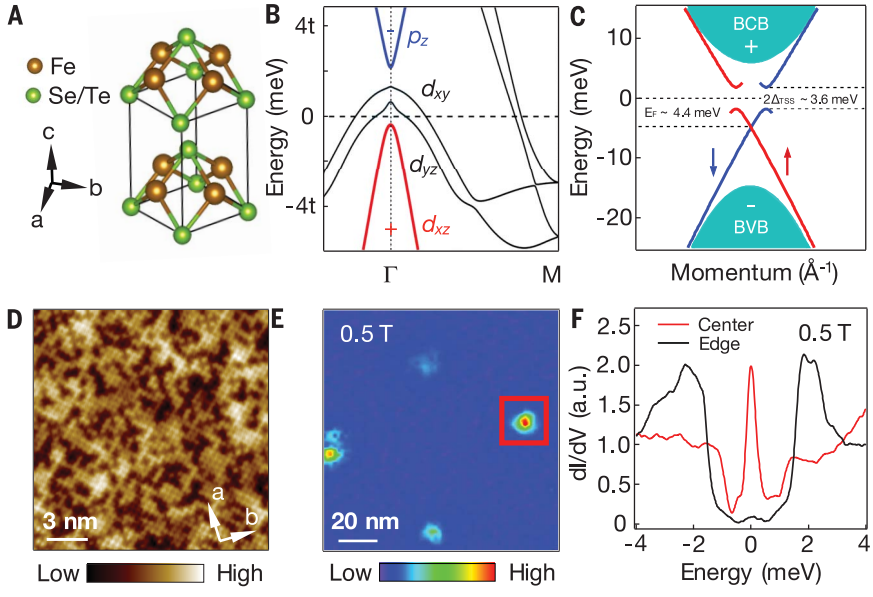


Fig. 1. Band structure and vortex cores of $FeTe_{0.55}Se_{0.45}$. (**A**) Crystal structure of $FeTe_{0.55}Se_{0.45}$. Axis a or b indicates one of the Fe–Fe bond directions. (**B**) A first-principle calculation of the band structure along the Γ-M direction. In the calculations, t = 100 meV, whereas t ~ 12 to 25 meV from ARPES experiments, largely depending on the bands (*23*). [Adapted from (*18*), figure 1C] (**C**) Summary of SC topological surface states on this material observed by ARPES from (*18*). (**D**) STM topography of $FeTe_{0.55}Se_{0.45}$ (scanning area, 17 nm by 17 nm). (**E**) Normalized zero-bias conductance (ZBC) map measured at a magnetic field of 0.5 T, with the area 120 nm by 120 nm. (**F**) A sharp ZBP in a *dI*/*dV* spectrum measured at the vortex core center indicated in the red box in (E). Settings are sample bias, V_s = −5 mV; tunneling current, I_t = 200 pA; and temperature, T = 0.55 K.

[1]Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences (CAS), Beijing 100190, China. [2]School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China. [3]CAS Center for Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China. [4]Collaborative Innovation Center of Quantum Matter, Beijing 100190, China. [5]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA. [6]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: dingh@iphy.ac.cn (H.D.); hjgao@iphy.ac.cn (H.-J.G.)

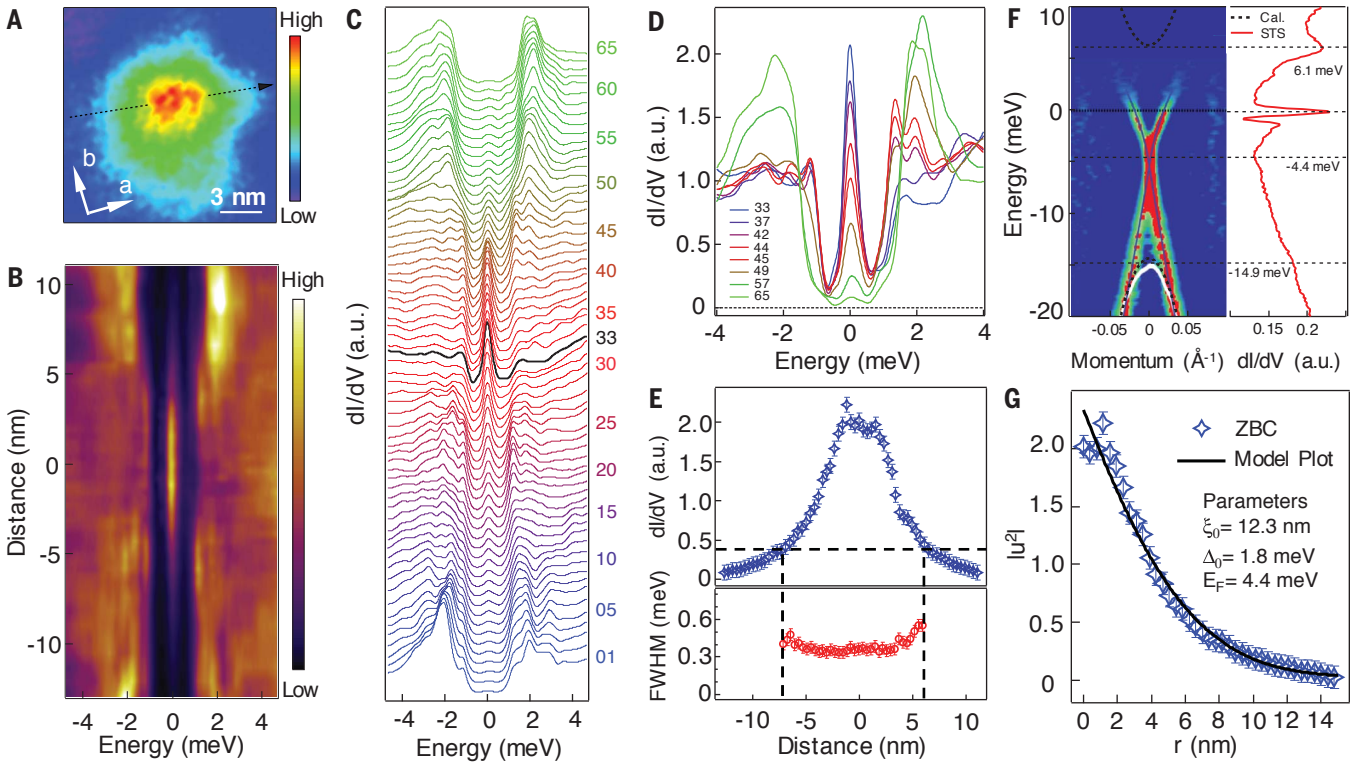


Fig. 2. Energetic and spatial profile of ZBPs. (**A**) A ZBC map (area, 15 nm by 15 nm) around vortex cores. (**B**) A line-cut intensity plot along the black dashed line indicated in (A). (**C**) A waterfall-like plot of (B) with 65 spectra, with the black curve corresponding to the one in the core center. (**D**) An overlapping display of eight *dI/dV* spectra selected from (C). (**E**) Spatial dependence of the height (top) and FWHM (bottom) of the ZBP. (**F**) Comparison between ARPES and STS results. (Left) ARPES results on the topological surface states. [Adapted from (*18*)] Black dashed curves are extracted from a first-principle calculation (*37*), with the calculated data rescaled to match the energy positions of the Dirac point and the top of the bulk valence band (BVB). (Right) A *dI/dV* spectrum measured from −20 to 10 meV. (**G**) Comparison between the measured ZBP peak intensity with a theoretical calculation of MBS spatial profile [(*24*), part VIII]. The data in (B) to (G) are normalized by the integrated area of each *dI/dV* spectrum. Settings are V_s = −5 mV, I_t = 200 pA, T = 0.55 K, and perpendicular magnetic field ($B_\perp$) = 0.5 T.

(*19*, *20*) and is consistent with tunneling into an isolated MBS in a vortex core of a SC topological material (*5*, *25*–*27*). We then extracted the position-dependent values of the ZBP height and width using simple Gaussian fits of the data in Fig. 2C and obtained the spatial profile shown in Fig. 2E; the decaying profile has a nearly constant line width of ~0.3 meV in the center, which is close to the total width (~0.28 meV) contributed from the STM energy resolution [~0.23 meV as shown in part I of (*24*)] and the thermal broadening [$3.5k_BT$ at 0.55 K ~ 0.17 meV, where k_B is the Boltzmann constant]. We further compared the observed ZBP height with a theoretical MBS spatial profile obtained by solving the Bogoliubov–de Gennes equation analytically (*5*, *25*) or numerically (*26*, *27*). By using the parameters of E_F = 4.4 meV, Δ_{sc} = 1.8 meV, and $\xi_0 = v_F/\Delta_{sc}$ = 12 nm, which are obtained directly from the topological surface state by our scanning tunneling spectroscopy (STS) and ARPES results (Fig. 2F) (*18*), the theoretical MBS profile matches well the experimental one (Fig. 2G).

The observation of a nonsplit ZBP, which is different from the split ZBP observed in a vortex of the $Bi_2Te_3/NbSe_2$ heterostructure (*13*, *28*, *29*), indicates that the MBS peak in our system is much less contaminated by nontopological CBS peaks, which is made possible by the large Δ_{sc}/E_F ratio in this system. In a usual topological insulator/superconductor heterostructure, this ratio is tiny, on the order of 10^{-3} to 10^{-2} (*28*). This has been shown to induce, in addition to the MBS at the zero energy, many CBSs, whose level spacing is proportional to Δ_{sc}^2/E_F. As a result, these CBSs were crowded together very close to the zero energy, making difficult a clean detection of MBS from the *dI/dV* spectra (*29*). However, on the surface of $FeTe_{0.55}Se_{0.45}$, the value of Δ_{sc}^2/E_F is ~0.74 meV, which is sufficiently large to push most CBSs away from the zero energy (*24*), leaving the MBS largely isolated and unspoiled. A large energy separation (0.7 meV) between the ZBP and the CBS was observed in fig. S3, E to H, which is in agreement with Δ_{sc}^2/E_F of the topological surface states [(*24*), part IV]. Also, all the bulk bands in this multiband material have fairly small values of E_F owing to large correlation-induced mass renormalization, ranging from a few to a few tens of millielectron volts; thus, their values of Δ_{sc}^2/E_F are also quite large (>0.2 meV) (table S1) (*24*). These large bulk ratios enlarge the energy-level spacing of CBSs inside the bulk vortex line, which helps reduce quasiparticle poisoning of the MBS at low temperature [(*24*), part II].

It has been predicted (*30*) that the width of the ZBP from tunneling into a single isolated MBS is determined by thermal smearing ($3.5k_BT$), tunneling broadening, and STM instrumentation resolution. We measured the tunneling barrier evolution of the ZBP (Fig. 3A). Robust ZBPs can be observed over two orders of magnitude in tunneling barrier conductance, with the width barely changing (Fig. 3B). Also, the line width of ZBPs is almost completely limited by the combined broadening of energy resolution and STM thermal effect, suggesting that the intrinsic width of the MBS is much smaller, and our measurements are within the weak tunneling regime.

However, we did observe some other ZBPs with a larger broadening (Fig. 3C). A larger ZBP broadening is usually accompanied with a softer SC gap, or the FWHM of ZBP increases with increasing subgap background conductance. The subgap background conductance, which is determined by factors such as the strength of scattering from disorder and quasiparticle interactions (*31*–*33*), introduces a gapless fermion bath that can poison the MBS, as explained previously (*34*). The effect of quasiparticle poisoning is to reduce the MBS amplitude and increase its width. This scenario is likely the origin of a

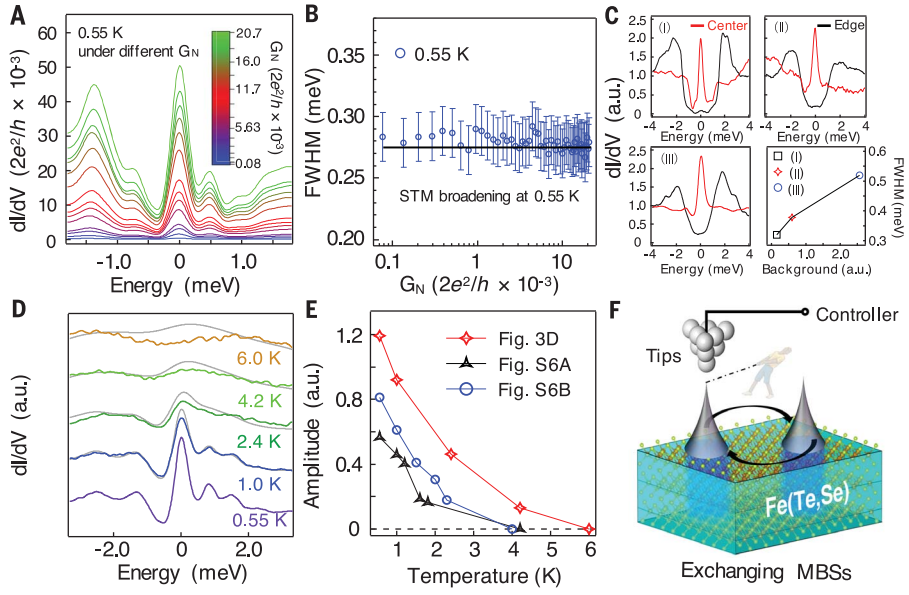


Fig. 3. Temperature and tunneling barrier evolution of ZBPs. (**A**) Evolution of ZBPs, with tunneling barrier measured at 0.55 K. $G_N \equiv I_t/V_s$, which corresponds to the energy-averaged conductance of normal states and represents the conductance of the tunneling barrier. I_t and V_s are the STS setpoint parameters. (**B**) FWHM of ZBPs at 0.55 K under different tunneling barriers. The black solid line is the combined effect of energy resolution (0.23 meV) (*24*) and tip thermal broadening ($3.5k_BT$) at 0.55 K. (**C**) FWHM of ZBP at the center of the vortex core is larger when the SC gap around the vortex core is softer. Background is defined as an integrated area from −1 to +1 meV of the spectra at the core edge. (**D**) Temperature evolution of ZBPs in a vortex core. The gray curves are numerically broadened 0.55 K data at each temperature. (**E**) Amplitude of the ZBPs shown in (D) and fig. S6 (*24*) under different temperatures. The amplitude is defined as the peak-valley difference of the ZBP. (**F**) Schematic of a possible way for realizing non-Abelian statistics in an ultralow-temperature STM experiment that may have an ability to exchange MBSs on the surface of Fe(Te, Se). (A) and (B) show the absolute value of conductance; $B_\perp$ = 2.5 T. In (D) and (E), the data are normalized by integrated area; V_s = −10 mV, I_t = 100 pA, T = 0.55 K, and $B_\perp$ = 4 T.

larger broadening of ZBP accompanied by a softer gap.

It has been pointed out by previous theoretical studies (*35*–*37*) that the condition of a bulk vortex line, such as its chemical potential, has substantial influences on the Majorana mode on the surface by the vortex phase transition. In order to further characterize the effects of bulk vortex lines, we have monitored the temperature evolution of a ZBP. As shown in Fig. 3D, the ZBP intensity measured at a vortex center decreases with increasing temperature and becomes extremely weak at 4.2 K and totally invisible at 6.0 K. A peak associated with a CBS would persist to higher temperatures and exhibit simple Fermi-Dirac broadening up to about $T_c/2$ (~8 K), below which the SC gap amplitude is almost constant, as observed in our previous ARPES measurement (*18*). Our observation (Fig. 3D) contradicts this expectation and indicates an additional suppression mechanism that is likely related to the poisoning of MBS by thermally excited quasiparticles. From the extraction of ZBP amplitude measured on several different vortices (three cases are shown in Fig. 3E), we found that most of the observed ZBPs vanish around 3 K, which is higher than the temperature in many previous Majorana platforms (*11*, *38*). This vanishing temperature is comparable with the energy level spacing of the bulk vortex line as discussed above; thus, the temperature dependence we found is consistent with a case of a MBS poisoned by thermally induced quasiparticles inside the bulk vortex line (*24*).

Our observations provide strong evidence for tunneling to an isolated MBS; many alternative trivial explanations [(*24*), part III] cannot account for all the observed features. It is technically possible to move a vortex by a STM tip, which in principle can be used to exchange MBSs inside vortices (Fig. 3F), consequently demonstrating non-Abelian statistics under a sufficiently low ($k_BT \ll \Delta_{sc}^2/E_F$) temperature (*2*). The high transition temperature and large SC gaps in this superconductor offer a promising platform to fabricate robust devices for topological quantum computation.

REFERENCES AND NOTES

1. A. Y. Kitaev, *Ann. Phys.* **303**, 2–30 (2003).
2. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
3. A. Y. Kitaev, *Phys. Uspekhi* **44** (10S), 131–136 (2001).
4. N. Read, D. Green, *Phys. Rev. B* **61**, 10267–10297 (2000).
5. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
6. R. M. Lutchyn, J. D. Sau, S. Das Sarma, *Phys. Rev. Lett.* **105**, 077001 (2010).
7. Y. Oreg, G. Refael, F. von Oppen, *Phys. Rev. Lett.* **105**, 177002 (2010).
8. A. C. Potter, P. A. Lee, *Phys. Rev. Lett.* **105**, 227003 (2010).
9. X.-L. Qi, T. L. Hughes, S.-C. Zhang, *Phys. Rev. B* **82**, 184516 (2010).
10. S. Nadj-Perge, I. K. Drozdov, B. A. Bernevig, A. Yazdani, *Phys. Rev. B* **88**, 020407 (2013).
11. V. Mourik *et al.*, *Science* **336**, 1003–1007 (2012).
12. S. Nadj-Perge *et al.*, *Science* **346**, 602–607 (2014).
13. H.-H. Sun *et al.*, *Phys. Rev. Lett.* **116**, 257003 (2016).
14. M. T. Deng *et al.*, *Science* **354**, 1557–1562 (2016).
15. Q.-L. He *et al.*, *Science* **357**, 294–299 (2017).
16. Z.-J. Wang *et al.*, *Phys. Rev. B* **92**, 115119 (2015).
17. X.-X. Wu, S. Qin, Y. Liang, H. Fan, J. Hu, *Phys. Rev. B* **93**, 115129 (2016).
18. P. Zhang *et al.*, *Science* **360**, 182–186 (2018).
19. C. Caroli, P. G. de Gennes, J. Matricon, *Phys. Lett.* **9**, 307–309 (1964).
20. H. F. Hess, R. B. Robinson, J. V. Waszczak, *Phys. Rev. Lett.* **64**, 2711–2714 (1990).
21. T. Hanaguri, S. Niitaka, K. Kuroki, H. Takagi, *Science* **328**, 474–476 (2010).
22. F. Massee *et al.*, *Sci. Adv.* **1**, e1500033 (2015).
23. H. Miao *et al.*, *Phys. Rev. B* **85**, 094506 (2012).
24. Materials and methods are available as supplementary materials.
25. Y. Wang, L. Fu, *Phys. Rev. Lett.* **119**, 187003 (2017).
26. C.-K. Chiu, M. J. Gilbert, T. L. Hughes, *Phys. Rev. B* **84**, 144507 (2011).
27. L.-H. Hu, C. Li, D.-H. Xu, Y. Zhou, F.-C. Zhang, *Phys. Rev. B* **94**, 224501 (2016).
28. J.-P. Xu *et al.*, *Phys. Rev. Lett.* **112**, 217001 (2014).
29. J.-P. Xu *et al.*, *Phys. Rev. Lett.* **114**, 017001 (2015).
30. F. Setiawan, C.-X. Liu, J. D. Sau, S. Das Sarma, *Phys. Rev. B* **96**, 184520 (2017).
31. S. Das Sarma, A. Nag, J. D. Sau, *Phys. Rev. B* **94**, 035143 (2016).
32. Y. Yin *et al.*, *Phys. Rev. Lett.* **102**, 097002 (2009).
33. C. Renner, A. D. Kent, P. Niedermann, O. Fischer, F. Lévy, *Phys. Rev. Lett.* **67**, 1650–1652 (1991).
34. J. R. Colbert, P. A. Lee, *Phys. Rev. B* **89**, 140505 (2014).
35. P. Hosur, P. Ghaemi, R. S. K. Mong, A. Vishwanath, *Phys. Rev. Lett.* **107**, 097001 (2011).
36. H.-H. Hung, P. Ghaemi, T. L. Hughes, M. J. Gilbert, *Phys. Rev. B* **87**, 035401 (2013).
37. G. Xu, B. Lian, P. Tang, X.-L. Qi, S.-C. Zhang, *Phys. Rev. Lett.* **117**, 047001 (2016).
38. F. Nichele *et al.*, *Phys. Rev. Lett.* **119**, 136803 (2017).

ACKNOWLEDGMENTS

We thank Q. Huan, H. Isobe, X. Lin, X. Wu, and K. Yang for technical assistance and P. A. Lee, T. K. Ng, S. H. Pan, G. Xu, J.-X. Yin, F. C. Zhang, and P. Zhang for useful discussions. **Funding:** This work at IOP is supported by grants from the Ministry of Science and Technology of China (2013CBA01600, 2015CB921000, 2015CB921300, and 2016YFA0202300), the National Natural Science Foundation of China (11234014, 11574371, and 61390501), and the Chinese Academy of Sciences (XDPB08-1, XDB07000000, and XDPB0601). L.F. and G.G. are supported by the U.S. Department of Energy (DOE) (DE-SC0010526 and DE-SC0012704, respectively). J.S. and R.Z. are supported by the Center for Emergent Superconductivity, an Energy Frontier Research Center funded by the U.S. DOE. **Author contributions:** H.D., H.-J.G., L.K., Y.S., and S.D. designed the experiments. D.W., L.K., P.F., H.C., S.Z., W.L., and L.C. performed the STM experiments. J.S., R.Z., and G.G. provided the samples. L.F. provided theoretical models and explanations. All the authors participated in analyzing the experimental data, plotting the figures, and writing the manuscript. H.D. and H.-J.G. supervised the project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data presented in this paper can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/333/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*39*–*69*)
Data File S1

22 June 2017; resubmitted 11 December 2017
Accepted 27 July 2018
Published online 16 August 2018
10.1126/science.aao1797

MAGNETIC RESONANCE

Hyperfine interaction of individual atoms on a surface

Philip Willke[1,2,3], Yujeong Bae[1,2,3], Kai Yang[1], Jose L. Lado[4,5], Alejandro Ferrón[6], Taeyoung Choi[2,3], Arzhang Ardavan[7], Joaquín Fernández-Rossier[4*], Andreas J. Heinrich[2,3]†, Christopher P. Lutz[1]†

Taking advantage of nuclear spins for electronic structure analysis, magnetic resonance imaging, and quantum devices hinges on knowledge and control of the surrounding atomic-scale environment. We measured and manipulated the hyperfine interaction of individual iron and titanium atoms placed on a magnesium oxide surface by using spin-polarized scanning tunneling microscopy in combination with single-atom electron spin resonance. Using atom manipulation to move single atoms, we found that the hyperfine interaction strongly depended on the binding configuration of the atom. We could extract atom- and position-dependent information about the electronic ground state, the state mixing with neighboring atoms, and properties of the nuclear spin. Thus, the hyperfine spectrum becomes a powerful probe of the chemical environment of individual atoms and nanostructures.

The hyperfine interaction between an electron and a nuclear spin provides insight into the electronic structure and chemical bonding of atoms, molecules, and solids. It is also key to realizing quantum operations of the nuclear spin (*1*–*3*). Although most experiments rely on ensemble measurement, detection and control of single nuclear spins is possible for diluted molecules in crystals (*4*), molecules in break junctions (*3*, *5*), ion traps (*2*), and defects in solids (*6*, *7*). Nevertheless, until now, no experimental tool allowed for imaging of the atomic-scale environment of the nucleus and simultaneous resolution of the hyperfine spectrum, which crucially depends on the local electronic structure.

Atomic resolution, along with atom manipulation, has been achieved by scanning probe methods, and they have been used to study the local electronic structure of atoms and molecules. Scanning probes allow chemical identification (*8*) and imaging of chemical bonds (*9*), and they provide access to the structure (*10*) and dynamics (*11*, *12*) of electron spins. The nuclear properties detected in scanning tunneling microscopy (STM) include changes in vibrational (*13*) and rotational (*14*) motion, but the hyperfine interaction has not previously been resolved owing to a limited resolution in energy (*15*, *16*).

Combining STM with electron spin resonance (ESR), as demonstrated for individual iron (Fe) (*17*, *18*) and titanium (Ti) atoms (*19*, *20*), potentially provides the required energy resolution. However, no hyperfine splitting was reported so far. For atoms that have low natural abundances of magnetic nuclei (such as Fe) or multiple magnetic isotopes (such as Ti), it is necessary to study a large number of individual atoms to establish the hyperfine physics. Here we present such a study, aided by a new understanding of the tunneling parameters (*21*) that improve the signal-to-noise ratio compared to previous ESR-STM experiments (*17*, *18*).

Individual Fe and Ti atoms from sources that have natural isotopic abundance were deposited on two atomic layers of MgO grown on silver (*22*). First, we examined Fe atoms adsorbed onto an oxygen binding site of the MgO (*23*) (Fig. 1A). We resonantly excited transitions between the electronic ground state (spin-up) and the first excited state (spin-down) of the Fe atom by using ESR. We used a magnetic field B = 0.9 T and a temperature T = 1.2 K, unless stated otherwise. This excitation led to a change in tunnel current ΔI when driven at the resonance frequency (*17*) (Fig. 1B). A magnetic tip was used for magnetoresistive readout, leading to different conductance for the two electronic states.

For 3.4% of Fe atoms (5 out of 147, see inset in Fig. 1B) investigated, the ESR peak was split by Δf = 231 ± 5 MHz, where f is the frequency of the applied radio frequency voltage. We attributed this splitting to the presence of a nuclear moment $I = {}^1/_2$ for ^{57}Fe (2.1% natural abundance). By contrast, the I = 0 isotopes, predominantly ^{56}Fe, showed no splitting. We could not distinguish different I = 0 isotopes and thus label them ^{56}Fe for simplicity. Because the hyperfine splitting Δf is much smaller than the thermal energy $k_BT/h \approx 25$ GHz, where k_B is the Boltzmann constant and h is the Planck constant, the two nuclear states $m_I = \pm{}^1/_2$ are occupied with equal probability. Thus, the ^{57}Fe peaks each had half the ESR peak height of the single peak observed for ^{56}Fe. Moreover, the simultaneous presence of two peaks indicates that the nuclear spin relaxation time was here much shorter than the time scale of the measurement (~1 ms) (*22*).

For the magnetic field used here, the electronic Zeeman energy is large compared to the hyperfine interaction. Because the Fe atom on MgO shows a large out-of-plane magnetic anisotropy, the electron spin is quantized along the out-of-plane direction z (*23*). Thus, the nuclear spin was quantized along z as well, leading to the spin Hamiltonian (*24*)

$$H = g_z\mu_B B_z S_z + A_z I_z S_z \qquad (1)$$

where μ_B is the Bohr magneton and S_z and I_z are the z-axis spin operators for the electron and nuclear spin, respectively. Parameters g_z and A_z are the z components of the electron g-factor and the hyperfine coupling constant, respectively. B_z is the z component of the magnetic field. The very small nuclear Zeeman energy is neglected here.

When determining the eigenstates of Eq. 1, the spins can either be aligned or anti-aligned with the magnetic field to give product states $|m_S, m_I\rangle$, where m_S and m_I are the electron and nuclear quantum numbers (Fig. 1C). Transitions occurred between the $m_S = \pm 2$ electronic states ($\Delta m_S = 4$) (*17*) and left the nuclear spin state unchanged ($\Delta m_I = 0$), leading to one resonance peak for ^{56}Fe. The spectra were centered at frequency f_0 given by the Zeeman energy $hf_0 = g_z\mu_B B_z \Delta m_S$. The two frequencies observed for ^{57}Fe in Fig. 1B differed because of the relative alignment of the electron and nuclear spin, to give a peak splitting $\Delta f = \Delta m_S \Delta m_I A_z = 4A_z$, where $\Delta m_I = 1$ (*24*).

In contrast to that of Fe, the hyperfine interaction of Ti atoms was more complex and reflected changes in the local electronic environment (*24*). The Ti atoms studied here were hydrogenated (*19*), resulting in an electronic spin of ${}^1/_2$ that lacked magnetic anisotropy. Thus, to good approximation, the electron spin followed the magnetic-field direction (*19*, *20*). The hyperfine spectra for Ti located on a bridge binding site (Ti_B) of MgO (*20*) (Fig. 2) showed $2I + 1$ peaks, one for each nuclear spin state, resulting in six peaks for ^{47}Ti ($I = {}^5/_2$, 7.4% natural abundance), eight peaks for ^{49}Ti ($I = {}^7/_2$, 5.4%), and a single peak for the nuclear spin–free isotopes, predominantly ^{48}Ti (I = 0, 73.7%). We found a splitting Δf between adjacent peaks of ~47 MHz for both ^{47}Ti and ^{49}Ti, implying that their nuclear gyromagnetic ratios were equal [see section 2 in (*22*)]. For all investigated atoms, we found no appreciable dependence of the hyperfine splitting on the magnetic field caused by the proximity of

[1]IBM Almaden Research Center, San Jose, CA 95120, USA. [2]Center for Quantum Nanoscience, Institute for Basic Science (IBS), Seoul 03760, Republic of Korea. [3]Department of Physics, Ewha Womans University, Seoul 03760, Republic of Korea. [4]QuantaLab, International Iberian Nanotechnology Laboratory (INL), 4715-310 Braga, Portugal. [5]Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland. [6]Instituto de Modelado e Innovación Tecnológica (CONICET-UNNE), Facultad de Ciencias Exactas, Naturales y Agrimensura, Universidad Nacional del Nordeste, W3404AAS Corrientes, Argentina. [7]Centre for Advanced Electron Spin Resonance, Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK.
*On leave from Departamento de Física Aplicada, Universidad de Alicante, San Vicente del Raspeig 03690, Spain.
†Corresponding author. Email: cplutz@us.ibm.com (C.P.L.); heinrich.andreas@qns.science (A.J.H.)

the tip (*19*) nor on the local electric field induced by the dc bias voltage.

In contrast to ensemble measurements, the binding site of individual atoms could be determined from STM images, and the atom could be moved among binding sites by using atom manipulation. We moved a ^{47}Ti atom to different binding sites (Fig. 3A). We observed a pronounced reduction of the splitting Δf to ~10 MHz (Fig. 3B) when the Ti atom was moved from a bridge to an oxygen site (Ti_O). The larger splitting was restored by moving the same atom back to a bridge binding site. In Fig. 3C, we show the exact position of the atoms obtained from Fig. 3A and show density functional theory (DFT) calculations (*22*) of the respective electronic ground state. The statistics for different atoms on the same type of binding site (Fig. 3D) revealed a small, but reproducible, difference in Δf for the two inequivalent bridge binding sites. This difference was caused by different values of $g_x \neq g_y$ and $A_x \neq A_y$ (resulting from the low symmetry of the bridge binding site) along with a magnetic field direction $B_x \neq B_y$ [Fig. 3C and section 1 in (*22*)].

Hyperfine spectra for both isotopes of Ti_O are shown in high resolution in Fig. 3E. In addition to the strong decrease in Δf compared with that for Ti_B, the interval between peaks as well as their amplitudes were nonuniform. To describe the spectra, we generalize the hyperfine Hamiltonian in Eq. 1 to

$$H = H_{EZ} + H_{HF} + H_{NQ} \quad (2)$$

Here, the electron Zeeman term $H_{EZ} = \mu_B \sum_{i=x,y,z} g_i B_i S_i$ and the hyperfine term $H_{HF} = \sum_{i=x,y,z} A_i I_i S_i$ were generalized to include the x and y spin components, needed for electronic spins lacking strong anisotropy (*19*), and the hyperfine interaction was allowed to have anisotropy. Moreover, $H_{NQ} = \sum_{i=x,y,z} P_i I_i^2$ is the nuclear electric quadrupole interaction, caused by the electric field gradient at the position of the nucleus (*24*), where P_i are the components of the nuclear electric quadrupole tensor (*22*). This quadrupole term is irrelevant for $I = 1/2$ systems, as in the case of Fe (*24*). Using the known symmetry of the system, only three fit parameters, A_z, $A_x = A_y$, and P_z were required to give good fits to the complex hyperfine spectrum of Ti_O (*22*).

Simulated ESR spectra of the experimental data obtained by using EasySpin (*25*) are superimposed on the data in Fig. 3E. The fits reflect that the hyperfine and the nuclear quadrupole interaction have comparable energies for Ti_O, resulting in peaks that are irregularly spaced because they are not well approximated by eigenstates of either operator alone. Furthermore, these measurements allowed us to determine the ratio of the nuclear quadrupole moments of $^{49}Q/^{47}Q \approx 0.79$. The excellent agreement with ensemble measurements [$^{49}Q/^{47}Q = 0.82$ (*26*)] shows the fidelity of the model and that good accuracy was obtained with this limited number of fit parameters.

Although the electric quadrupole interaction changed the peak intensities and spacings for Ti_O, the overall reduction in Δf compared to that for Ti_B was caused by the reduction in the hyperfine coupling constant $\boldsymbol{A}$. We consider here two contributions (*24*). The first is the isotropic Fermi contact term originating from a finite spin density of unpaired s electrons at the nucleus induced by interaction with the d electrons. The second is the anisotropic magnetic dipolar interaction of the nuclear spin with the surrounding d electrons, which depends on the orbital symmetry (*27*). The observed changes in Δf for the different binding sites of Ti were caused by the occupation of different orbitals in the magnetic ground state of the atom (Fig. 3C). We found a larger Fermi contact contribution for Ti_B (+50 MHz) compared with that for Ti_O (+19 MHz), presumably caused by a difference in covalency (*24*) [section 2 in (*22*)].

More importantly, the three spatial components of the dipolar contribution changed in magnitude and sign for the different binding sites. Because the magnetic field was applied

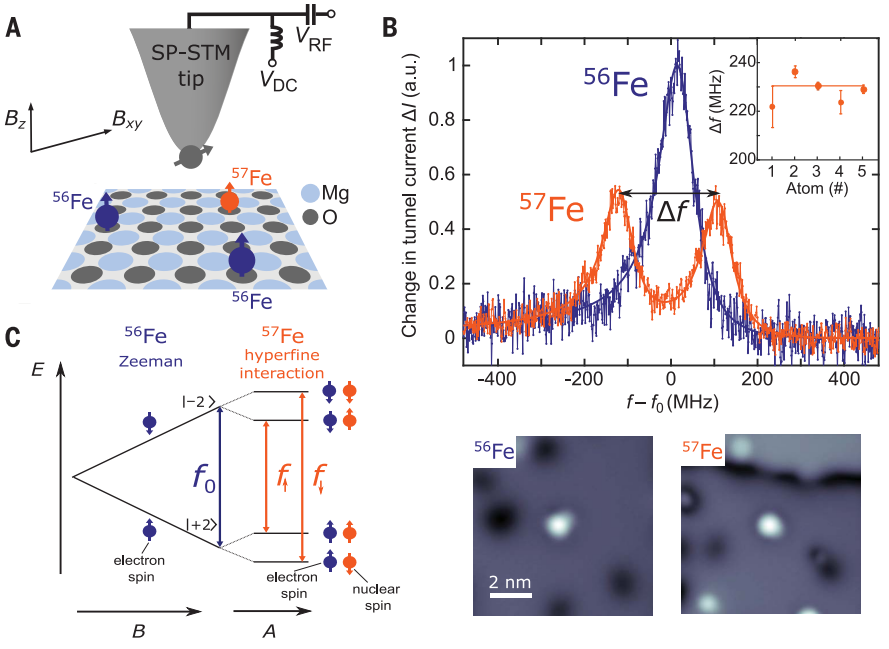


Fig. 1. Hyperfine interaction studied by ESR in a scanning tunneling microscope. (**A**) Experimental setup showing different isotopes of single Fe atoms on a bilayer MgO film on Ag(001) in an STM (B = 0.9 T, B_z = 0.1 T, and T = 1.2 K). ESR was performed by applying a radio frequency voltage (V_{RF}) to the tunneling junction (*17*, *18*). SP, spin-polarized; V_{DC}, direct current voltage. (**B**) ESR spectra of the change in tunnel current for different Fe isotopes (top). The blue line shows Fe that has zero nuclear spin (likely ^{56}Fe; see main text). The orange line shows ^{57}Fe with nuclear spin $I = 1/2$. Spectra were normalized to unity for ^{56}Fe and given in arbitrary units (a.u.). Electron Zeeman energy gives the center frequency f_0, here 19.89 GHz for ^{56}Fe and 19.87 GHz for ^{57}Fe (tunnel conditions: I_{set} = 12 pA, V_{DC} = 60 mV, and V_{RF} = 60 mV). The inset shows the peak splitting Δf for five different ^{57}Fe atoms. The solid line indicates the error-weighted mean. Shown at the bottom are topographic images of both atoms. (**C**) Schematic of the energy levels E of the ^{56}Fe atom, compared to those of ^{57}Fe, including hyperfine interaction as a function of the external magnetic field B and the hyperfine constant A (the small nuclear Zeeman energy was neglected).

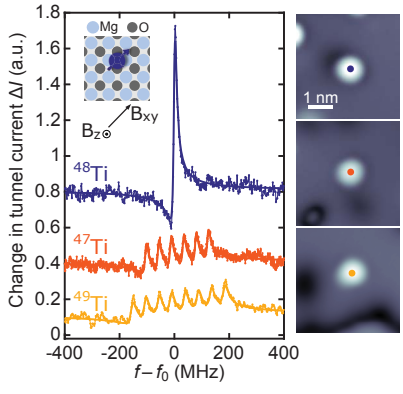


Fig. 2. Hyperfine interaction for titanium on a bridge binding site. On the left are ESR spectra for ^{48}Ti (blue, nuclear spin I = 0), ^{47}Ti (orange, $I = 5/2$) and ^{49}Ti (yellow, $I = 7/2$). The inset shows the bridge binding site. On the right are STM images for each atom. Experimental conditions for ^{48}Ti: f_0 = 23.03 GHz, I_{set} = 8 pA, V_{DC} = 40 mV, and V_{RF} = 30 mV; for ^{47}Ti: f_0 = 22.99 GHz, I_{set} = 8 pA, V_{DC} = 40 mV, and V_{RF} = 30 mV; and for ^{49}Ti: f_0 = 22.69 GHz, I_{set} = 20 pA, V_{DC} = 60 mV, and V_{RF} = 40 mV.

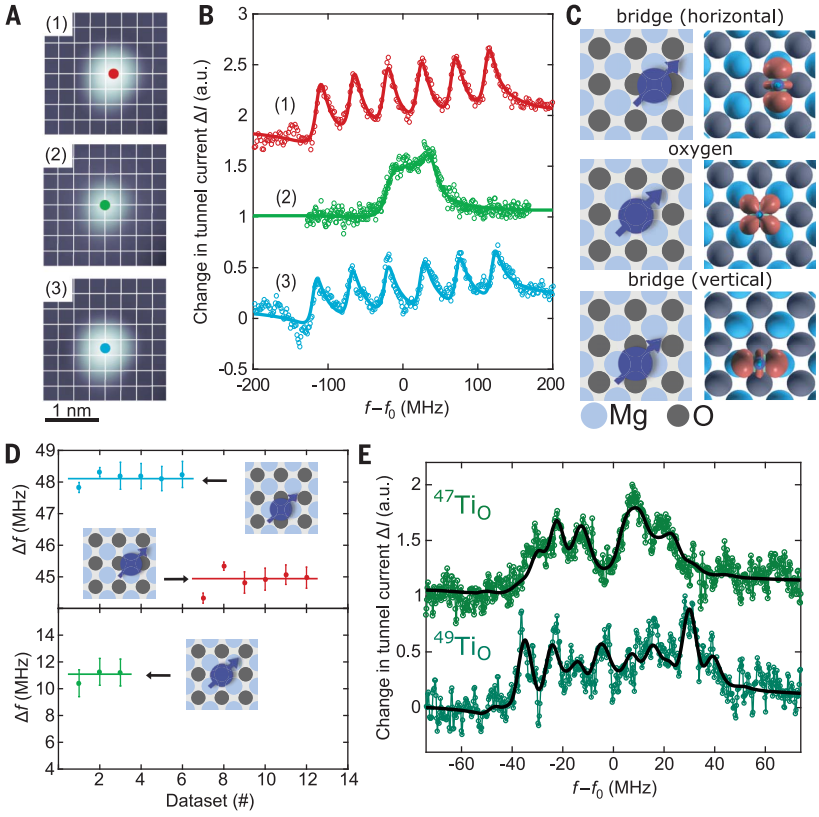


Fig. 3. Binding site dependence of the hyperfine spectrum of titanium. (**A**) By using atom manipulation, a ^{47}Ti atom on a bridge site (site 1, red, Ti_B) was moved to an oxygen binding site (site 2, green, Ti_O) and subsequently to a different bridge binding site (site 3, blue). White lines indicate the MgO lattice, with the intercepts corresponding to the positions of oxygen atoms. (**B**) ESR spectra for the binding sites in (A) (site 1: I_{set} = 10 pA, V_{DC} = 40 mV, V_{RF} = 40 mV, and f_0 = 22.89 GHz; site 2: I_{set} = 10 pA, V_{DC} = 60 mV, V_{RF} = 7 mV, and f_0 = 22.91 GHz; and site 3: I_{set} = 10 pA, V_{DC} = 40 mV, V_{RF} = 40 mV, and f_0 = 22.69 GHz). (**C**) Sketches (left) show the binding site. Calculated spin density (right) of the binding configurations obtained by DFT. (**D**) Statistics of the hyperfine splitting Δf for the different binding sites (horizontal bridge site in red, Δf = 44.9 ± 0.5 MHz; vertical bridge site in blue, Δf = 48.1 ± 0.3 MHz; and oxygen site in green, Δf = 10.8 ± 0.7 MHz. For the latter, Δf was obtained by neglecting electric quadrupole interaction which leads to unequal spacing of the peaks). (**E**) High-resolution ESR spectra for ^{47}Ti (top, I_{set} = 1.5 pA, V_{DC} = 60 mV, V_{RF} = 25 mV, T = 0.6 K, and f_0 = 22.49 GHz) and ^{49}Ti (bottom, I_{set} = 2.5 pA, V_{DC} = 60 mV, V_{RF} = 25 mV, T = 0.6 K, and f_0 = 22.49 GHz) on an oxygen binding site. Black lines are fits to the data, including the anisotropic hyperfine and nuclear electric quadrupole interaction (see main text).

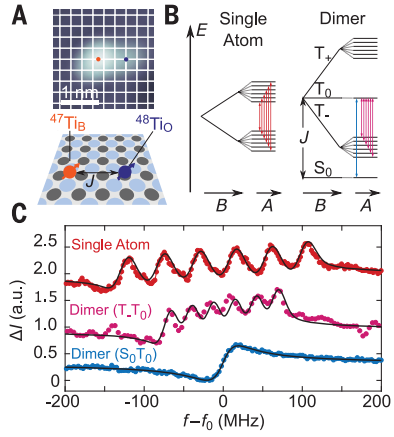


Fig. 4. Tuning the hyperfine splitting of a coupled two-atom nanostructure. (**A**) Topography of a dimer of Ti atoms created by atom manipulation (separation 7.2 Å). J denotes the exchange coupling strength between them. The bridge-site atom has a nuclear magnetic moment ($^{47}Ti_B$, $I = {}^5/_2$); the oxygen-site atom is nuclear spin–free ($^{48}Ti_O$, I = 0). (**B**) Energy levels of the single $^{47}Ti_B$ atom (left) and the dimer (right). For the dimer, the electronic spins form a singlet-triplet system because of exchange coupling (coupling strength J/h = 29.1 ± 0.2 GHz). (**C**) ESR spectrum of the transitions between T_0 and T_- taken on $^{47}Ti_B$ in the dimer reveals a decrease in hyperfine splitting Δf (middle, I_{set} = 10 pA, V_{DC} = 40 mV, V_{RF} = 20 mV, and f_0 = 22.89 GHz) compared with that of the isolated case (top, same data as in Fig. 2A). In the case of the singlet-triplet transition, S_0T_0, no hyperfine splitting was observed (bottom, I_{set} = 5 pA, V_{DC} = 40 mV, V_{RF} = 20 mV, and f_0 = 29.72 GHz).

nearly in-plane (Fig. 1A), the in-plane components A_x and A_y contributed the most to Δf. In the case of Ti_B, the dipolar contribution added to the Fermi contact part for one of the two components ($A_x \approx +61$ MHz; $A_y \approx +29$ MHz). By contrast, for Ti_O, both in-plane directions opposed the Fermi contact interaction, resulting in $A_x = A_y \approx +10$ MHz. As a result, Δf is smaller for Ti_O than for Ti_B. The strength of the hyperfine interaction and the importance of quadrupole interaction revealed the profound change in the chemical environment of the Ti atom upon moving it from a bridge site to an oxygen site: Modeling the hyperfine interaction and taking the DFT ground state into account, we find that the bonding with oxygen reduces the electron spin density at the Ti nucleus for Ti_O [~−1 instead of ~−2.8 atomic units for Ti_B (*22*)]. Moreover, the radial spread of the spin-polarized orbital $\langle r^{-3} \rangle$ can be deduced from the hyperfine splitting and the electric quadrupole contribution, which lies at ~(0.5 Å)$^{-3}$ for both binding sites (*22*).

Hyperfine spectra can also reveal changes in the magnetic environment, as we demonstrate here using assembled structures of Ti atoms. Figure 4A shows a $^{47}Ti_B$ atom with a nuclear spin $I = {}^5/_2$ that was moved to a position ~7 Å from a $^{48}Ti_O$ atom (I = 0). Because of the exchange interaction between the electronic spins, an electronic singlet-triplet system was formed [right panel of Fig. 4B and section 3 in (*22*)] (*19*, *20*). The Zeeman term split the triplet states (T_-, T_0, and T_+), whereas the energies of both the singlet S_0 and the triplet state T_0 remained independent of magnetic field.

Analogously, the hyperfine interaction did not split either S_0 or T_0, and for the T_+ and T_- states, the splitting remained the same as for the single atom (Fig. 4B). Thus, the hyperfine splitting for the T_-T_0 transition, which involved one of the magnetic field–independent states (T_0), decreased to only 27.2 ± 0.4 MHz, roughly half the value of the isolated Ti case (Fig. 4C). Accordingly, the hyperfine splitting of the singlet-triplet transition (S_0T_0) in Fig. 4C was essentially zero (less than our ~10 MHz linewidth). These transitions allowed us to probe the polarization of the coupled-atom states and to quantify the degree of state mixing to 87% in the singlet state [section 3 in (*22*)]. Here $\boldsymbol{A}$ remained constant, and instead, the electronic spin magnetization was changed (*20*). This Ti dimer structure showed that the collective properties that emerged in a correlated multispin structure yielded characteristics sharply different than those of the constituents.

REFERENCES AND NOTES

1. B. E. Kane, *Nature* **393**, 133–137 (1998).
2. B. B. Blinov, D. Leibfried, C. Monroe, D. J. Wineland, *Quantum Inform. Process.* **3**, 45–59 (2004).

3. S. Thiele *et al*., *Science* **344**, 1135–1138 (2014).
4. J. Köhler, A. C. J. Brouwer, E. J. J. Groenen, J. Schmidt, *Science* **268**, 1457–1460 (1995).
5. R. Vincent, S. Klyatskaya, M. Ruben, W. Wernsdorfer, F. Balestro, *Nature* **488**, 357–360 (2012).
6. P. Neumann *et al*., *Science* **329**, 542–544 (2010).
7. J. J. Pla *et al*., *Nature* **496**, 334–338 (2013).
8. Y. Sugimoto *et al*., *Nature* **446**, 64–67 (2007).
9. L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, *Science* **325**, 1110–1114 (2009).
10. A. J. Heinrich, J. A. Gupta, C. P. Lutz, D. M. Eigler, *Science* **306**, 466–469 (2004).
11. S. Loth, M. Etzkorn, C. P. Lutz, D. M. Eigler, A. J. Heinrich, *Science* **329**, 1628–1630 (2010).
12. A. A. Khajetoorians *et al*., *Science* **339**, 55–59 (2013).
13. B. C. Stipe, M. A. Rezaei, W. Ho, *Science* **280**, 1732–1735 (1998).
14. F. D. Natterer, F. Patthey, H. Brune, *Phys. Rev. Lett.* **111**, 175303 (2013).
15. F. Delgado, J. Fernández-Rossier, *Phys. Rev. Lett.* **107**, 076804 (2011).
16. C. R. Ast *et al*., *Nat. Commun.* **7**, 13009 (2016).
17. S. Baumann *et al*., *Science* **350**, 417–420 (2015).
18. W. Paul, S. Baumann, C. P. Lutz, A. J. Heinrich, *Rev. Sci. Instrum.* **87**, 074703 (2016).
19. K. Yang *et al*., *Phys. Rev. Lett.* **119**, 227206 (2017).
20. Y. Bae *et al*., arXiv:1807.04322 [cond-mat.mes-hall] (11 July 2018).
21. P. Willke *et al*., *Sci. Adv.* **4**, q1543 (2018).
22. Materials and methods are available as supplementary materials.
23. W. Paul *et al*., *Nat. Phys.* **13**, 403–407 (2017).
24. A. Abragam, B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Oxford Univ. Press, 2012).
25. S. Stoll, A. Schweiger, *J. Magn. Reson.* **178**, 42–55 (2006).
26. K. H. Channappa, J. M. Pendlebury, *Proc. Phys. Soc.* **86**, 1145–1146 (1965).
27. F. E. Mabbs, D. Collison, *Electron Paramagnetic Resonance of d Transition Metal Compounds* (Elsevier, 2013), vol. 16.

ACKNOWLEDGMENTS

We thank B. Melior for expert technical assistance and W. Paul, F. Natterer, and S. Fatayer for further discussions. **Funding:** We gratefully acknowledge financial support from the Office of Naval Research. P.W., Y.B., T.C., and A.J.H. acknowledge support from the Institute for Basic Science under grant IBS-R027-D1. A.F. acknowledges funding from CONICET (PIP11220150100327) and FONCyT (PICT-2012- 2866). J.L.L. is grateful for financial support from the ETH Fellowship program. J.F.-R. thanks the Fundação para a Ciência e a Tecnologia, under project no. PTDC/FIS-NAN/4662/2014 (016656). A.A. thanks the Engineering and Physical Sciences Research Council, under project no. EP/L011972/1. P.W. acknowledges support from the Alexander von Humboldt Foundation. **Author contributions:** A.J.H. and C.P.L. supervised the project. P.W., Y.B., K.Y., and T.C. carried out the STM measurements. P.W. performed the analysis and wrote the manuscript with help from all authors. A.F., J.L.L., and J.F.-R. performed the DFT calculations. All authors discussed the results and edited the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials. Additional data related to this paper may be requested from the authors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/336/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Table S1
References (*28*–*32*)

27 March 2018; accepted 15 August 2018
10.1126/science.aat7047

NEURODEVELOPMENT

Supracellular contraction at the rear of neural crest cell groups drives collective chemotaxis

Adam Shellard[1], András Szabó[1], Xavier Trepat[2,3,4], Roberto Mayor[1]*

Collective cell chemotaxis, the directed migration of cell groups along gradients of soluble chemical cues, underlies various developmental and pathological processes. We use neural crest cells, a migratory embryonic stem cell population whose behavior has been likened to malignant invasion, to study collective chemotaxis in vivo. Studying *Xenopus* and zebrafish, we have shown that the neural crest exhibits a tensile actomyosin ring at the edge of the migratory cell group that contracts in a supracellular fashion. This contractility is polarized during collective cell chemotaxis: It is inhibited at the front but persists at the rear of the cell cluster. The differential contractility drives directed collective cell migration ex vivo and in vivo through the intercalation of rear cells. Thus, in neural crest cells, collective chemotaxis works by rear-wheel drive.

Directed migration orchestrates events in development, homeostasis, and disease (*1*–*4*). Most long-range directed migration in vivo occurs by chemotaxis (*2*, *4*–*9*), in which cells follow gradients of soluble chemical cues. This has been best understood in individually migrating cells, for which several mechanisms have been proposed (*10*–*13*), but collective migration has been less well studied.

In collective migration, leader cells have dynamic actin-based protrusions (Fig. 1A, darker red) (*1*, *6*), form contacts with follower cells and with the extracellular matrix, and are responsive to chemotactic signals (*3*, *14*, *15*). We asked whether cells at the group's rear (Fig. 1A, dotted rectangle) contribute to collective cell chemotaxis. To investigate the mechanism of collective chemotaxis ex vivo and in vivo, we studied *Xenopus* and zebrafish cranial neural crest cells, an embryonic cell population that undergoes collective cell migration (*6*, *16*) in a manner similar to cancer cells (*17*), unlike neural crest cells of other species or those in the trunk, where less is known about the collectiveness (*18*). Although contact inhibition of locomotion and cluster confinement (*19*, *20*) are needed for cephalic neural crest directional movement in *Xenopus* and zebrafish, they are not sufficient, as collective chemotaxis toward SDF1 is essential for long-range directed movement (*6*).

Imaging of fluorescently tagged actin and myosin in neural crest explants revealed the presence of a multicellular actomyosin ring localized at the periphery of the cell group, in both the absence and presence of an SDF1 gradient (Fig. 1B and fig. S1, A and B). Enrichment of N-cadherin near the actomyosin cable at the cell junction (Fig. 1, C to F, and fig. S1, C to E) suggests that this cable is supracellular. Premigratory neural crest and neural crest overexpressing E-cadherin, but not N-cadherin, have internalized myosin localization, rather than myosin at the cluster periphery (fig. S1, F to J), suggesting that the switch of cadherin expression during the epithelial-to-mesenchymal transition may be required for the formation of the actomyosin cable.

To determine whether the actomyosin cable is contractile, we performed laser photoablation of the structure, resulting in recoil of both the actomyosin cable and cell-cell junctions (fig. S2, A and B), followed by the cable's reformation (fig. S2, C and D). To assess contractility, we measured actomyosin length, and we found frequent shortening (Fig. 1, G and H) independent of SDF1. These contractions were multicellular, as adjacent cells contracted synchronously (Fig. 1I and fig. S2E). A second ablation in a nearby cell after an initial ablation resulted in reduced actomyosin recoil (fig. S2, F and G), indicating that the tension of the cable is transmitted between cells.

[1]Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK. [2]Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute for Science and Technology (BIST), Barcelona 08028, Spain. [3]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Barcelona 08028, Spain. [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain.
***Corresponding author. Email: r.mayor@ucl.ac.uk**

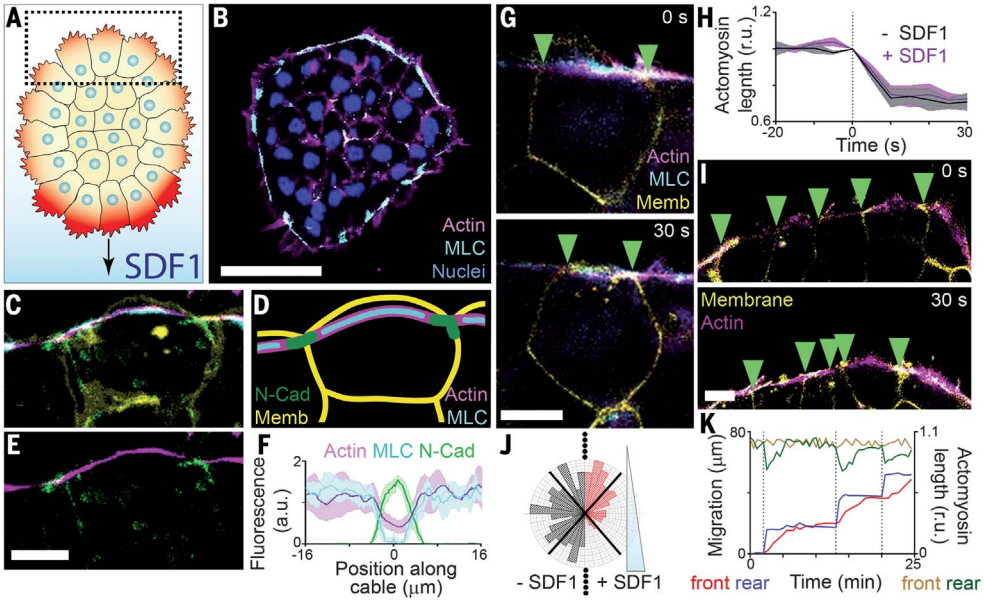


Fig. 1. *Xenopus* neural crest clusters exhibit a contractile actomyosin ring. (**A**) A neural crest cluster with protrusions (red) at the edge undergoes chemotaxis to SDF1. SDF1 stabilizes the protrusions at the front (darker red) (*7*). Dotted rectangle, rear cells. (**B**) Immunofluorescence of a neural crest explant in the absence of SDF1. MLC, myosin light chain. Scale bar, 50 μm. (**C** to **E**) Immunofluorescence of a cell at the edge of a neural crest explant (C and E) and a corresponding diagram (D). N-Cad, N-cadherin; Memb, membrane. Scale bar, 10 μm. (**F**) Protein fluorescence levels (means ± SEM) along the actin cable. Position 0 μm represents the cell contact. a.u., arbitrary units. n = 8 cells. (**G**) Spontaneous contraction of the actomyosin cable. Green arrowheads, cell-cell contacts. Scale bar, 10 μm. (**H**) Actomyosin length (means ± SEM) measured over time. Contractions start at 0 s. n = 20 cells. (**I**) Multicellular contraction of the actomyosin cable. Scale bar, 10 μm. (**J**) Distribution of actomyosin contractility at different angles without (−SDF1) or with (+SDF1) an SDF1 gradient. n = 150 contractions. (**K**) Relative actomyosin length at the front (brown line) and rear (green line) of a cluster and the positions of the front (red line) and rear (blue line) of the cluster. r.u., relative units.

In epithelial cells, the presence of an actomyosin cable seems to inhibit protrusion formation (*21*), but this inhibition does not occur in mesenchymal neural crest cells (fig. S2, H and I).

Although exposure to SDF1 gradients did not affect the magnitude of actomyosin contractions (Fig. 1H), contractions occurred less frequently in front cells during collective chemotaxis without affecting cells at the rear (Fig. 1J and fig. S3A). A similar inhibition of front contractions was observed with the chemoattractant PDGF-A (*22*) (fig. S3B). Mechanistically, this contractility gradient is likely set up by SDF1 activation of Rac1 in front cells, which inhibits RhoA and myosin phosphorylation (fig. S4). Uniform SDF1, unlike the SDF1 gradient, did not inhibit contractility (fig. S5A), suggesting that the cluster responds to the chemotactic gradient instead of to absolute SDF1 levels. This was further supported by the observation that rear contractility (fig. S5B) and cluster speed (fig. S5C) were unchanged when clusters were closer to the chemoattractant source, where higher SDF1 levels should be present.

To explore the connection between the asymmetric actomyosin contraction and collective chemotaxis, we simultaneously measured the positions of front and rear cells of explants during migration, as well as the length of the actomyosin cable at the front and rear. Pulsatile contraction of the cable at the rear (Fig. 1K, green lines, and fig. S6A) coincided with the forward movement of the rear (Fig. 1K, blue lines, and fig. S6A). Both events immediately preceded the movement of the front of the cluster (Fig. 1K, red lines, and fig. S6A). A similar local contraction precedes a short forward movement in the absence of SDF1 (fig. S6, B and C), but with no long-range directed movement. Together, these results suggest that supracellular actomyosin contractility at the rear may drive collective cell chemotaxis.

We tested the role of rear contractility of the actomyosin ring in collective chemotaxis by performing laser ablation. Chemotaxis was impaired by ablation of the actomyosin ring in rear cells but not by equivalent ablations in front cells or by other control ablations (Fig. 2, A to C, and fig. S7), suggesting the necessity of a rear supracellular actomyosin cable for chemotaxis. To test the requirement of rear contractility, we used an optogenetic system (*23*, *24*) to either increase (via the protein construct optoGEF-contract) or decrease (via the construct optoGEF-relax) contractility and myosin phosphorylation in the actomyosin cable upon illumination with low doses of blue light (fig. S8). No effect was observed on cell protrusions (fig. S9), focal adhesions (fig. S10), cell dispersion (fig. S11), or the phosphorylation of myosin located basally outside the cable (fig. S8K) upon illumination under the conditions of our assay. We first tested whether high contractility at the rear is necessary for collective chemotaxis by photoactivating optoGEF-relax at the rear of migrating clusters exposed to SDF1 (Fig. 2D). Inhibition of contractility in rear cells (Fig. 2D) impaired chemotaxis (Fig. 2, E and F). By contrast, inhibition of contractility in front cells failed to affect collective chemotaxis (fig. S12). To determine whether rear contractility is sufficient to drive collective cell migration, we activated contractility in rear cells in the absence of SDF1 (Fig. 2G). Whereas control neural crest cells did not exhibit directional migration, activated neural crest cells moved forward, away from the region of photoactivation (Fig. 2, H and I).

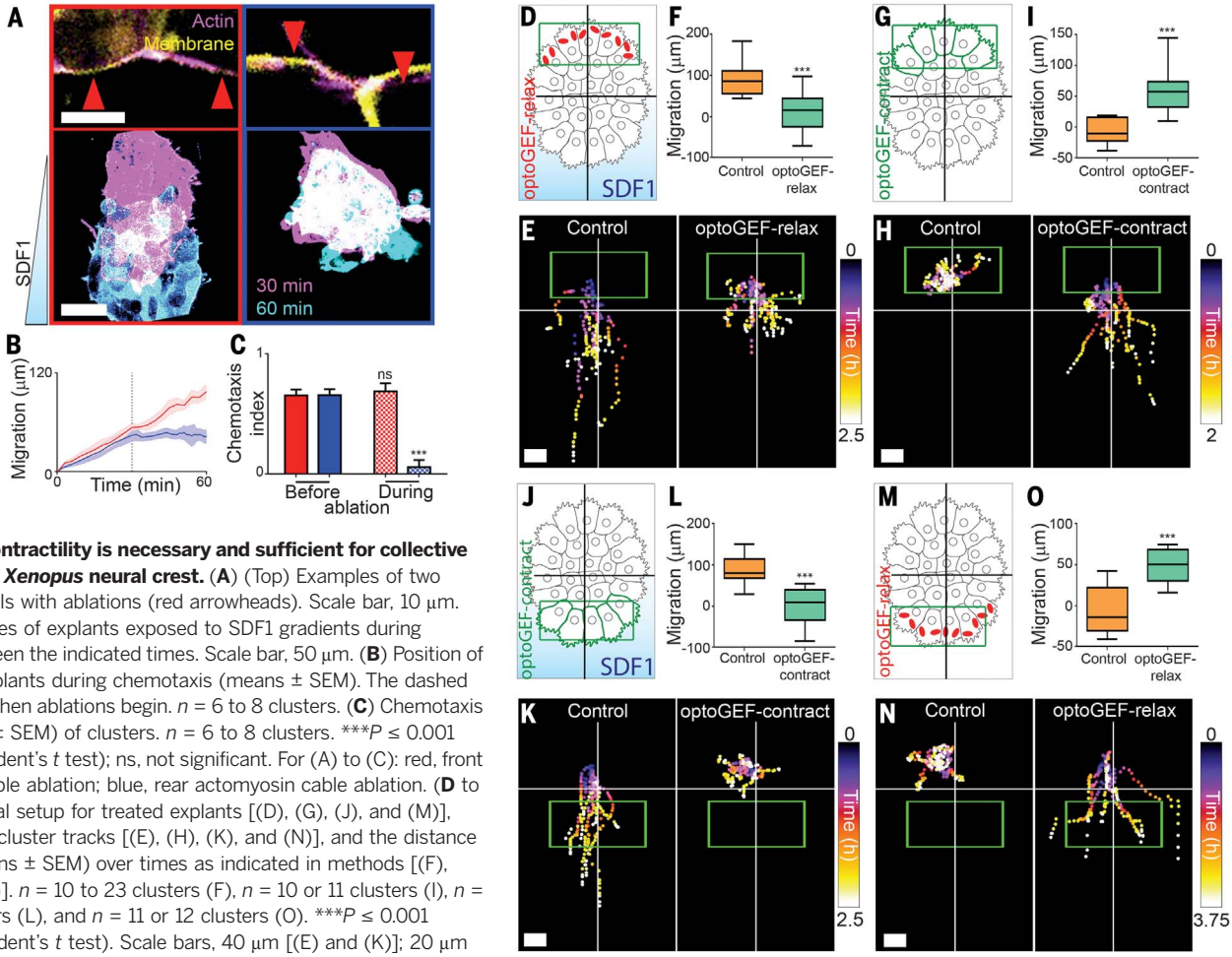


Fig. 2. Rear contractility is necessary and sufficient for collective chemotaxis of *Xenopus* neural crest. (**A**) (Top) Examples of two neighboring cells with ablations (red arrowheads). Scale bar, 10 μm. (Bottom) Images of explants exposed to SDF1 gradients during ablations between the indicated times. Scale bar, 50 μm. (**B**) Position of the front of explants during chemotaxis (means ± SEM). The dashed line indicates when ablations begin. n = 6 to 8 clusters. (**C**) Chemotaxis index (means ± SEM) of clusters. n = 6 to 8 clusters. ***$P \leq 0.001$ (two-tailed Student's t test); ns, not significant. For (A) to (C): red, front actomyosin cable ablation; blue, rear actomyosin cable ablation. (**D** to **O**) Experimental setup for treated explants [(D), (G), (J), and (M)], representative cluster tracks [(E), (H), (K), and (N)], and the distance migrated (means ± SEM) over times as indicated in methods [(F), (I), (L), and (O)]. n = 10 to 23 clusters (F), n = 10 or 11 clusters (I), n = 14 to 18 clusters (L), and n = 11 or 12 clusters (O). ***$P \leq 0.001$ (two-tailed Student's t test). Scale bars, 40 μm [(E) and (K)]; 20 μm [(H) and (N)]. Green box, initial illumination area; cross, initial cluster position. The top of all pictures is the rear.

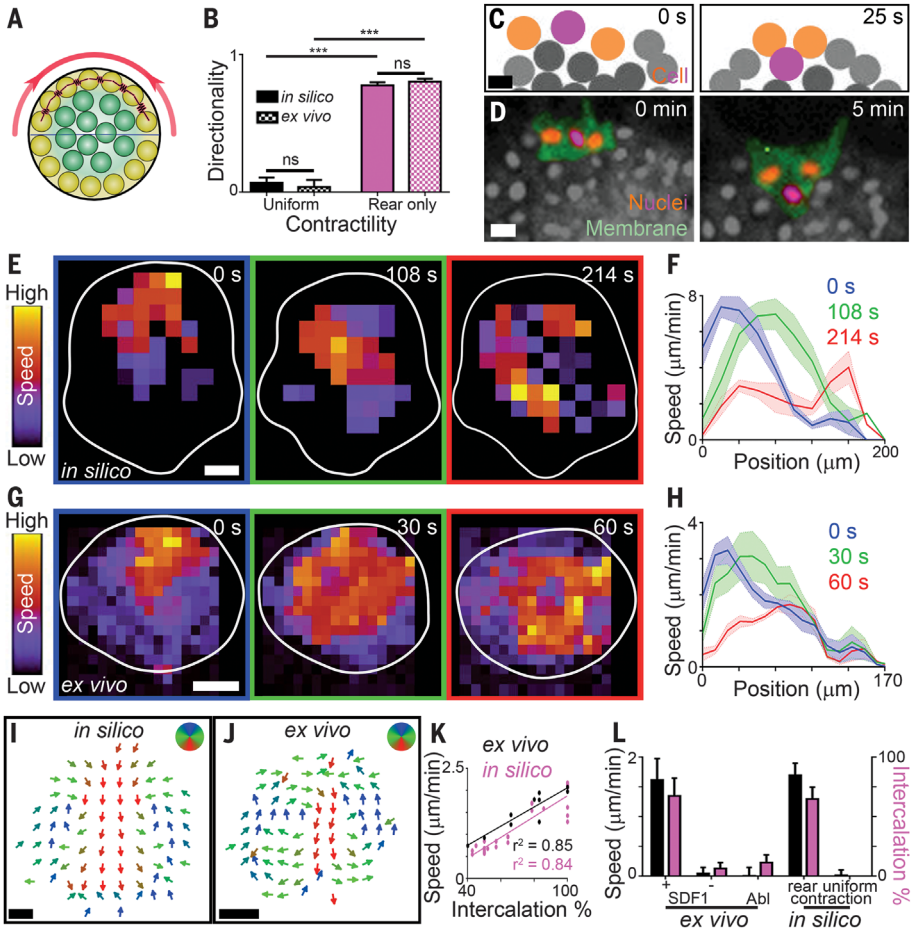


Fig. 3. Modeling contractility-driven collective migration. (**A**) Illustration of the computational model cluster. Yellow, edge cells; green, internal cells; red, contraction; horizontal line, distinction between front and rear, with rear outer cells contracting (red spring). (**B**) Directionality (means ± SEM) of clusters. n = 10 clusters. ***$P \leq 0.001$ (two-tailed Student's t test); ns, not significant. (**C** and **D**) Intercalation of a rear cell (purple) between two adjacent cells (orange) in silico (C) and ex vivo (D) during directional migration. Scale bars, 20 μm. (**E** to **H**) Wave of contraction. Heat maps indicate speed during migration in silico (E) and ex vivo (G). Scale bars, 40 μm. Graphs show speed profiles (means ± SEM) from clusters in silico (F) and ex vivo (H) at different times during directional migration. Position 0 μm represents the rear of the cluster; positions 200 and 170 μm [(F) and (H), respectively] represent the front of the cluster. n = 5 clusters. (**I** and **J**) Direction of intracluster cell movements shown from time-averaged cell tracks in silico (I) and particle image velocimetry ex vivo (J) after subtraction of cluster movement. n = 5 clusters. Scale bars, 40 μm. (**K**) Cluster speed and rear cell intercalation during migration. (**L**) Cluster speed (means ± SEM) and rear cell intercalation (means ± SEM) of clusters. Abl, laser ablation of the actomyosin ring in rear cells. n = 6 to 21 clusters. The top of all pictures is the rear.

To test whether SDF1-dependent inhibition of contractility in front cells is required for collective chemotaxis, we activated contractility in front cells of migrating clusters exposed to SDF1 (Fig. 2J); this repressed chemotaxis (Fig. 2, K and L), suggesting that low front contractility is essential for collective chemotaxis. Lastly, we asked whether front inhibition of contractility by SDF1 was sufficient to generate directed migration. We inhibited front contractility in the absence of SDF1 (Fig. 2M), which resulted in directional migration (Fig. 2, N and O). These optogenetic treatments affected contractility (fig. S13) (*23*) and not cell motility (fig. S14). Together, these results suggest that collective migration requires greater contractility at the rear than at the front of the cell cluster.

To understand how rear cell contractility might drive directed collective cell migration, we implemented a cell-centered computational model of a cell group with contractile edge cells (methods; Fig. 3A; and fig. S15, A to C). Cells interact through a soft-core repulsion and mid-range attraction; to model contractions, cells at the edge (either around the cluster or at the rear) periodically attract one another with additional force (Fig. 3A, red springs). Similar to ex vivo clusters, only simulations with rear but not uniform contractility were able to migrate forward (Fig. 3B, fig. S15D, and movie S1). Other migration parameters were comparable between in silico and ex vivo clusters (fig. S15, E and F). Unexpectedly, analysis of cell movements in silico revealed that rear cells in contractile regions were intercalated forward, into the cell group (Fig. 3C). As predicted by the model, we found an equivalent intercalation at the rear of neural crest clusters (Fig. 3D and fig. S15G). Furthermore, our simulations predicted that the effect of this local cell rearrangement is spread through the whole cell group such that when the cluster's rear contracts, the rear cells are intercalated, triggering a wave of cell movement that propagates from the rear toward the front of the cluster (Fig. 3, E and F). A similar wave was observed ex vivo (Fig. 3, G and H), as predicted by the model. This suggests that rear cell intercalation after rear contractions pushes cells forward progressively over time. Averaging cell movement over time and subtracting cluster movement reveals an intracluster flow of cells in silico, whereby rear cell intercalation causes a drift forward through the middle of the group and cells at the front and sides move backward, replacing rear cells (Fig. 3I). This was then confirmed to occur ex vivo as well (Fig. 3J). We found a positive correlation between the speed of ex vivo and in silico clusters during collective migration and the amount of rear cell intercalation (Fig. 3, K and L), consistent with this mechanism's driving cluster movement. Nonmigratory ex vivo and in silico clusters had low intercalation, and migratory clusters had comparable cluster speeds (Fig. 3L). We observed that contractions were normally accompanied by relaxation events (fig. S16A, green and red bars); however, we showed that ex vivo and in silico clusters were able to migrate directionally, independently of the level of rear relaxation (fig. S16, A and B, and movie S2). Altogether, these results suggest that rear contractility drives collective cell migration by inducing cell intercalation, which pushes the group forward.

Next, we analyzed whether this model of collective cell chemotaxis explains the in vivo migration of neural crest cells. As in ex vivo clusters, an actomyosin cable is present at the edge of the neural crest in both *Xenopus* (Fig. 4, A and B, and fig. S17, A and B) and zebrafish (fig. S18, A and B). Live imaging of the actomyosin cable shows that it is a contractile structure in vivo in both *Xenopus* (Fig. 4C and fig. S17C) and zebrafish (fig. S18, C and D) and contracts more often at the rear of the neural crest stream than at the front (fig. S17D). Rear contractility precedes forward movement of the cluster in vivo (Fig. 4D), as it does ex vivo. Less phosphomyosin was present at the front than at the rear at the beginning of migration (figs. S17, E to H, and S19). To identify whether individual neural crest cells flowed through clusters, as predicted from in silico and ex vivo results, we tracked live cells during migration. In both *Xenopus* and zebrafish, cells that were initially at the rear of the group were intercalated forward during migration (Fig. 4E and

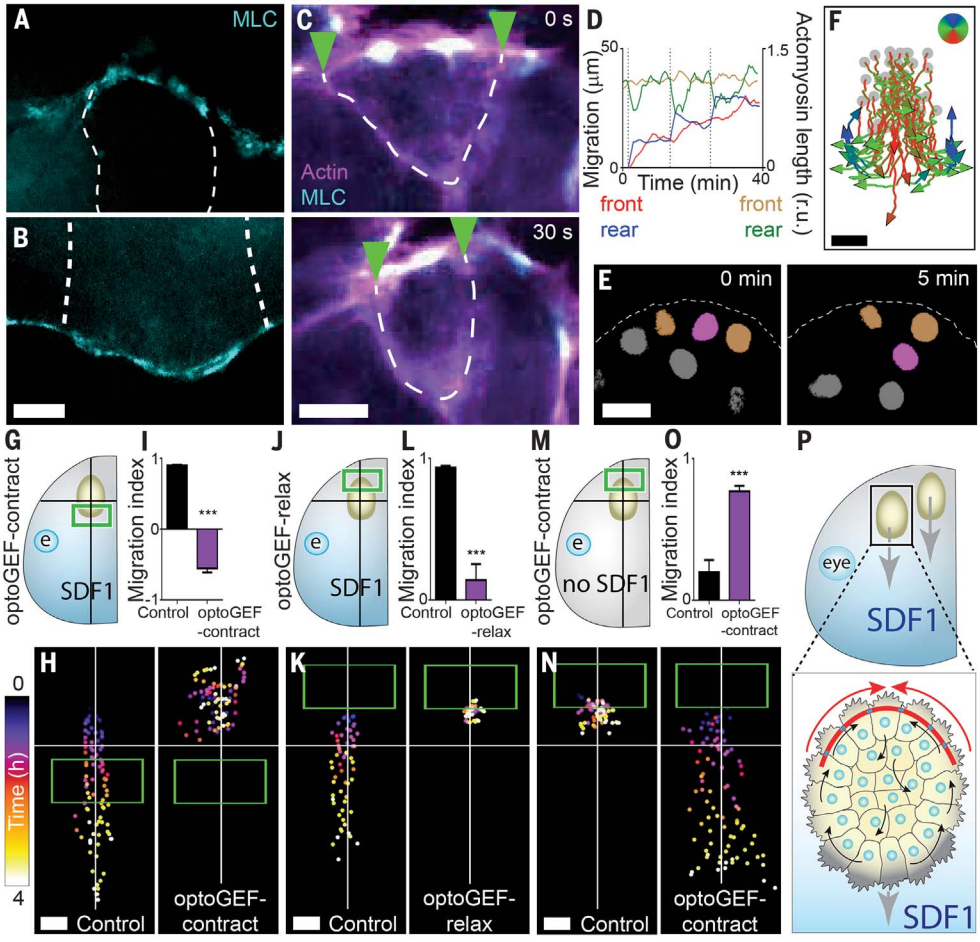


Fig. 4. Actomyosin drives collective chemotaxis in vivo in *Xenopus*. (**A** and **B**) Immunofluorescence of the rear (A) and front (B) of the *Xenopus* neural crest stream. Dashed lines, cell-cell contacts between neural crest cells. Scale bar, 10 μm. (**C**) Contraction of the actomyosin cable of *Xenopus* neural crest in vivo. Green arrowheads, cell-cell contacts; dashed lines, cell edges. Scale bar, 10 μm. (**D**) Actomyosin length at the front (brown line) and rear (green line) of a *Xenopus* cluster in vivo and the positions of the front (red line) and rear (blue line) of the cluster. (**E**) Intercalation of a rear cell (purple) between two adjacent cells (orange) in vivo. Scale bar, 20 μm. (**F**) Tracks of rear neural crest cells in vivo after subtraction of the cluster movement. Gray dots, initial cell positions. Scale bar, 30 μm. (**G** to **O**) Experimental design of treated *Xenopus* embryos [(G), (J), and (M)], representative tracks of neural crest clusters [(H), (K), and (N)], and migration indexes (means ± SEM) [(I), (L), and (O)]. Green boxes, initial illumination area; crosses, starting position of the explant. $n = 10$ clusters. $^{***}P \leq 0.001$ (two-tailed Student's t test). Scale bars, 50 μm. (**P**) The model: Collective cell chemotaxis is driven by actomyosin contractility at the rear (red arrows). The top of all pictures is the rear.

fig. S20, A and B). As with the ex vivo and in silico data, subtracting cluster movement to in vivo cell tracks revealed an intracluster flow (Fig. 4F and fig. S20C). This suggests that rear contractility is driving neural crest migration in vivo.

To test whether rear contractility is required for neural crest migration in vivo, we grafted neural crest expressing optoGEF-contract or optoGEF-relax into wild-type *Xenopus* embryos. Activation of contractility at the front of the stream (Fig. 4, G to I, and movie S3) or inhibition at the rear (Fig. 4, J to L, and movie S4) impaired neural crest migration, indicating that greater contractility at the rear than at the front was necessary for migration in the embryo. Neural crest grafted into host embryos lacking SDF1 failed to migrate, but activation of contractility at the rear of such grafts rescued migration (Fig. 4, M to O, and movie S5), demonstrating that high actomyosin contractility at the rear can drive directed collective migration in vivo. We conclude that rear contractility, as produced by a supracellular actomyosin cable, can drive collective cell chemotaxis in vivo (Fig. 4P).

The theory of active gels shows how anisotropies in viscoelastic materials can generate rotating flows similar to the cellular flows described here (*25*, *26*). In addition, physicists have proposed that cells can move by using tangential retrograde movement of their surfaces (*27*) and that this movement is more energetically efficient than other modes of swimming (*28*). However, only recently has such surface retrograde propulsion been described for the migration of single cells (*29*). Our work identifies an equivalent surface retrograde propulsion for collective cell migration, suggesting that the whole cluster behaves as a "supracell."

It is likely that for in vivo collective chemotaxis, rear actomyosin contractility works together with protrusions at the front to drive migration. Notably, peripheral actomyosin has been similarly observed in the collective migration of other cell types, including cancer cells (*30*, *31*), suggesting that other cell types may migrate under similar principles.

REFERENCES AND NOTES

1. P. Friedl, D. Gilmour, *Nat. Rev. Mol. Cell Biol.* **10**, 445–457 (2009).
2. P. Roca-Cusachs, R. Sunyer, X. Trepat, *Curr. Opin. Cell Biol.* **25**, 543–549 (2013).
3. R. Mayor, S. Etienne-Manneville, *Nat. Rev. Mol. Cell Biol.* **17**, 97–109 (2016).
4. A. Haeger, K. Wolf, M. M. Zegers, P. Friedl, *Trends Cell Biol.* **25**, 556–566 (2015).
5. R. Majumdar, M. Sixt, C. A. Parent, *Curr. Opin. Cell Biol.* **30**, 33–40 (2014).
6. E. Theveneau *et al.*, *Dev. Cell* **19**, 39–53 (2010).
7. D. Gilmour, M. Rembold, M. Leptin, *Nature* **541**, 311–320 (2017).
8. P. Haas, D. Gilmour, *Dev. Cell* **10**, 673–680 (2006).
9. E. Donà *et al.*, *Nature* **503**, 285–289 (2013).
10. M. D. Welch, *Cell* **160**, 581–582 (2015).
11. E. K. Paluch, E. Raz, *Curr. Opin. Cell Biol.* **25**, 582–590 (2013).
12. R. J. Petrie, K. M. Yamada, *Curr. Opin. Cell Biol.* **42**, 7–12 (2016).
13. R. H. Insall, *Nat. Rev. Mol. Cell Biol.* **11**, 453–458 (2010).
14. D. Cai *et al.*, *Cell* **157**, 1146–1159 (2014).
15. X. Wang, L. He, Y. I. Wu, K. M. Hahn, D. J. Montell, *Nat. Cell Biol.* **12**, 591–597 (2010).
16. M. Kaucka *et al.*, *Sci. Adv.* **2**, e1600060 (2016).
17. L. Kerosuo, M. Bronner-Fraser, *Semin. Cell Dev. Biol.* **23**, 320–332 (2012).
18. E. Theveneau, R. Mayor, *Genesis* **49**, 164–176 (2011).
19. C. Carmona-Fontaine *et al.*, *Nature* **456**, 957–961 (2008).
20. A. Szabó *et al.*, *J. Cell Biol.* **213**, 543–555 (2016).
21. M. Reffay *et al.*, *Nat. Cell Biol.* **16**, 217–223 (2014).
22. I. Bahm *et al.*, *Development* **144**, 2456–2468 (2017).
23. L. Valon, A. Marín-Llauradó, T. Wyatt, G. Charras, X. Trepat, *Nat. Commun.* **8**, 14396 (2017).
24. M. Uroz *et al.*, *Nat. Cell Biol.* **20**, 646–654 (2018).
25. K. Kruse, J. F. Joanny, F. Jülicher, J. Prost, K. Sekimoto, *Phys. Rev. Lett.* **92**, 078101 (2004).
26. R. Voituriez, J. F. Joanny, J. Prost, *Phys. Rev. Lett.* **96**, 028102 (2006).
27. E. M. Purcell, *Am. J. Phys.* **45**, 3–11 (1977).
28. A. M. Leshansky, O. Kenneth, O. Gat, J. E. Avron, *New J. Phys.* **9**, 145 (2007).
29. P. R. O'Neill *et al.*, *Dev. Cell* **46**, 9–22.e4 (2018).
30. C. Hidalgo-Carcedo *et al.*, *Nat. Cell Biol.* **13**, 49–58 (2011).
31. O. Zajac *et al.*, *Nat. Cell Biol.* **20**, 296–306 (2018).

ACKNOWLEDGMENTS

We thank G. Charras, L. Cramer, C. Stern, B. Stramer, and D. Wilkinson for critical reading of the manuscript and members of the Mayor laboratory for discussions. We thank M. Tada, M. Meyer, and E. Sahai for providing us with vectors and antibodies, and E. Scarpa for preliminary data. **Funding:** This study was supported by

grants from the Medical Research Council (M010465 and J000655 to R.M.), the Biotechnology and Biological Sciences Research Council (M008517 to R.M.), the Wellcome Trust (102489/Z/13/Z Wellcome Trust Ph.D. fellowship to A.Sh.), the Spanish Ministry of Economy and Competitiveness/FEDER (BFU2015-65074-P to X.T.), the Generalitat de Catalunya and CERCA program (2014-SGR-927 to X.T.), the European Research Council (CoG-616480 to X.T.), and the European Commission (project H2020-FETPROACT-01-2016-731957 to X.T.) and by a Marie Curie fellowship (329968 to A.Sz.).

Author contributions: R.M. conceptualized the study; A.Sh., A.Sz., X.T., and R.M. performed the methods; A.Sz. provided software; R.M. provided resources; A.Sh. and R.M. wrote the original draft of the manuscript; A.Sh., A.Sz., X.T., and R.M. reviewed and edited the manuscript; and R.M. supervised the study, provided project administration, and acquired funding. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/339/suppl/DC1
Materials and Methods
Figs. S1 to S20
References (*32*–*47*)
Movies S1 to S5

30 May 2018; accepted 10 September 2018
10.1126/science.aau3301

PLANT SCIENCE

The opium poppy genome and morphinan production

Li Guo[1,2,3]*, Thilo Winzer[4]*, Xiaofei Yang[1,5]*, Yi Li[4]*, Zemin Ning[6]*, Zhesi He[4], Roxana Teodor[4], Ying Lu[7], Tim A. Bowser[8], Ian A. Graham[4]†, Kai Ye[1,2,3,9]†

Morphinan-based painkillers are derived from opium poppy (*Papaver somniferum* L.). We report a draft of the opium poppy genome, with 2.72 gigabases assembled into 11 chromosomes with contig N50 and scaffold N50 of 1.77 and 204 megabases, respectively. Synteny analysis suggests a whole-genome duplication at ~7.8 million years ago and ancient segmental or whole-genome duplication(s) that occurred before the Papaveraceae-Ranunculaceae divergence 110 million years ago. Syntenic blocks representative of phthalideisoquinoline and morphinan components of a benzylisoquinoline alkaloid cluster of 15 genes provide insight into how this cluster evolved. Paralog analysis identified P450 and oxidoreductase genes that combined to form the *STORR* gene fusion essential for morphinan biosynthesis in opium poppy. Thus, gene duplication, rearrangement, and fusion events have led to evolution of specialized metabolic products in opium poppy.

Throughout history, opium poppy (*Papaver somniferum* L.) has been both a friend and foe of human civilization. In use since Neolithic times (*1*), the sap, known as opium, contains various alkaloids, including morphine and codeine, with effects ranging from pain relief and cough suppression to euphoria, sleepiness, and addiction. Opioid-based analgesics remain among the most effective and cheap treatments for the relief of severe pain and palliative care, but, owing to their addictive properties, careful medical prescription is essential to avoid misuse. Access to morphine equivalents to alleviate serious health-related suffering is unequal: In the United States and Canada, more than 3000% of the estimated need is met and in Western Europe, 870% is met, whereas in China, Russia, India, and Nigeria, 16%, 8%, 4%, and 0.2%, respectively, is met (*2*). Addressing the lack of access to pain relief or palliative care, especially among poor people in low to middle income countries, has been recognized as a global health and equity imperative (*2*).

Chemical synthesis or synthetic biology approaches are not yet commercially viable for any of the morphinan subclass of benzylisoquinoline alkaloids (BIAs) (*3*–*5*), and opium poppy remains the only commercial source. Genome rearrangements have been important in the evolution of BIA metabolism in opium poppy. For example, a cluster of 10 genes encodes enzymes for production of the antitussive and anticancer compound noscapine, which belongs to the phthalideisoquinoline subclass of BIAs (*6*), and a P450 oxidoreductase gene fusion (*4*, *7*, *8*) is responsible for the key gateway reaction directing metabolites toward the morphinan branch and away from the noscapine branch. Here we present the sequence assembly of the opium poppy genome to aid investigation into the evolution of BIA metabolism and provide a foundation for further yield improvement of this medicinal plant.

Large, complex plant genomes with an abundance of repeated sequences still pose challenges for genome analysis. In this study, we combined sequencing technologies (fig. S1), including Illumina paired-end and mate-pair (214X), 10X Genomics linked reads (40X), and Pacific Biosciences long-read (66.8X) sequencing and, for quality checking, Oxford Nanopore and Illumina sequencing of bacterial artificial chromosomes (Table 1 and tables S1 and S2). The final genome assembly of 2.72 Gb covers 94.8% of the estimated genome size (figs. S2 to S4 and table S3), and 81.6% of sequences have been assigned into individual chromosomes (fig. S5 and table S4) using a linkage map generated by sequence-based genotyping of 84 F2 plants (tables S5 and S6). We annotated the genome by using the MAKER pipeline (*9*) to incorporate protein homolog and transcriptome data from seven tissues (table S7). This annotation predicted 51,213 protein-coding genes (Table 1 and fig. S6) and 9494 noncoding RNAs (table S8). Benchmarking Universal Single-Copy Orthologs (BUSCO) (*10*) analysis based on plant gene models estimates 95.3% completeness (fig. S7). All predicted protein-coding genes are supported by RNA sequencing (RNA-seq) data or homologs, whereas 68.8% have significant hits in the InterPro database (Table 1). Repetitive elements make up 70.9% of the genome. Of the repetitive elements, 45.8% are long terminal repeat retrotransposons (fig. S8).

Synteny analysis revealed a relatively recent whole-genome duplication (WGD) event and traces of what we consider to be ancient segmental

Table 1. Assembly and annotation statistics of opium poppy genome.

Parameter	Value
Contig assembly	
Total number of contigs	65,578
Assembly size	2.71 Gb
N50	1.77 Mb
N90	590 kb
Largest contig	13.8 Mb
Scaffold assembly	
Total number of scaffolds	34,388
Assembly size	2.72 Gb
N50	204.5 Mb
N90	9.9 Mb
Largest scaffold	270.4 Mb
Annotation	
GC content	30.5%
Repeat density	70.9%
Number of protein-coding genes	51,213
Average length of protein-coding genes	3454 bp
Supported by RNA-seq or homologs	100%
Supported by protein families	68.8%

[1]MOE Key Lab for Intelligent Networks and Networks Security, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, 710049 China. [2]The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, 710061 China. [3]The School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049 China. [4]Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK. [5]Computer Science Department, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, 710049 China. [6]The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [7]Shanghai Ocean University, Shanghai, 201306 China. [8]Sun Pharmaceutical Industries Australia, Princes Highway, Port Fairy, VIC 3284, Australia. [9]Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China.
*These authors contributed equally to this work.
†Corresponding author. Email: ian.graham@york.ac.uk (I.A.G.); kaiye@xjtu.edu.cn (K.Y.)

duplications, although an ancient WGD cannot be ruled out (Fig. 1C and figs. S9 to S14). Distribution of synonymous substitutions per synonymous site (*Ks*) for *P. somniferum* paralogous genes and syntenic blocks confirmed a major WGD peak (Fig. 1C; figs. S13, A and D, and S15; and table S9) and a minor segmental duplication peak (Fig. 1C and fig. S13D). Intergenomic colinearity analysis indicated that *P. somniferum* did not experience γ, the hexaploidization event shared in core eudicots, as demonstrated by a 3:2 syntenic relationship between grape (*Vitis vinifera*) (*11*) and *P. somniferum* (fig. S9, C and F). Syntenic blocks (>132 kb) account for 86% total coverage across the whole genome (Fig. 1A and tables S10 to S13). Of the 25,744 genes arising from the WGD, 89.3% are present as two copies and 10.7% are present in more than two copies (fig. S10). Gene ontology analysis suggests that gene duplicates from the WGD are enriched with terms such as "cell redox homeostasis" and "positive regulation of transcription" (fig. S16). Comparison of *P. somniferum* with an ancestral eudicot karyotype (AEK) genome (*12*) (Fig. 1B) and with grape also supported the *P. somniferum* WGD (figs. S9, S11, and S12). We used OrthoFinder (*13*) to identify 48 single-copy orthologs shared across 11 angiosperm species, and phylogenetic analysis of these using BEAST (*14*) indicated that *P. somniferum* diverged from *Aquilegia coerulea* (Ranunculaceae) and *Nelumbo nucifera* (Nelumbonaceae) around 110 million years (Ma) ago and 125 Ma ago, respectively (Fig. 1D). Using divergence time and mean *Ks* values of syntenic blocks between *P. somniferum* and *A. coerulea*, we estimated the synonymous substitutions per site per year as 6.98×10^{-9} for Ranunculales, which led to the estimated time of the WGD at around 7.8 Ma ago (Fig. 1C, figs. S13 and S17, and table S9). By applying a phylogenomic approach that traces the history of paralog pairs using phylogenetic trees (*15*), we constructed 95 rooted maximum likelihood trees containing a pair of opium poppy paralogs from under the *Ks* peak around 1.5 as anchors. Additionally, we found that 65% of the trees support segmental duplications originating from multiple events occurring before the Papaveraceae-Ranunculaceae divergence at 110 Ma ago, whereas 35% support events occurring after the divergence (table S14 and data S1 and S2). After applying a more stringent approach with a bootstrap threshold cutoff at 50% for ancient gene pairs (*15*), we determined that 21% of trees support duplication events occurring before the Papaveraceae-Ranunculaceae divergence. The *Ks* distributions of *A. coerulea* paralogs and syntenic genes support a WGD event in this representative of the Ranunculaceae at 111.3 ± 35.6 Ma ago (Fig. 1C and table S9). A synteny dot plot of the opium poppy genome assembly with the *A. coerulea* genome assembly (*16*) revealed a 2:2 syntenic relationship (fig. S9D), which suggests that both the opium poppy and *A. coerulea* WGD events happened after the Papaveraceae-Ranunculaceae divergence, as shown in Fig. 1D.

The genome assembly allowed us to locate all of the functionally characterized genes of BIA metabolism in opium poppy, as well as their closely related homologs, to either chromosomes or unplaced scaffold positions (table S15). The noscapine gene cluster is located within a 584-kb region on chromosome 11, along with the *(S)- to (R)-reticuline* (*STORR*) gene fusion plus the remaining four genes in the biosynthetic pathway to production of the morphinan alkaloid, thebaine (Fig. 2, A and B). These genes are coexpressed in stems (Fig. 2C and fig. S18), and we refer to them as the BIA gene cluster. None of the other genes known to be associated with BIA metabolism—including *BERBERINE BRIDGE ENZYME* (*BBE*); *TETRAHYDROPROTOBERBERINE N-METHYLTRANSFERASE* (*TNMT*); and the bifurcated morphinan branch pathway genes *CODEINE 3-O-DEMETHYLASE* (*CODM*), *THEBAINE 6-O-DEMETHYLASE* (*T6ODM*), and *CODEINONE*

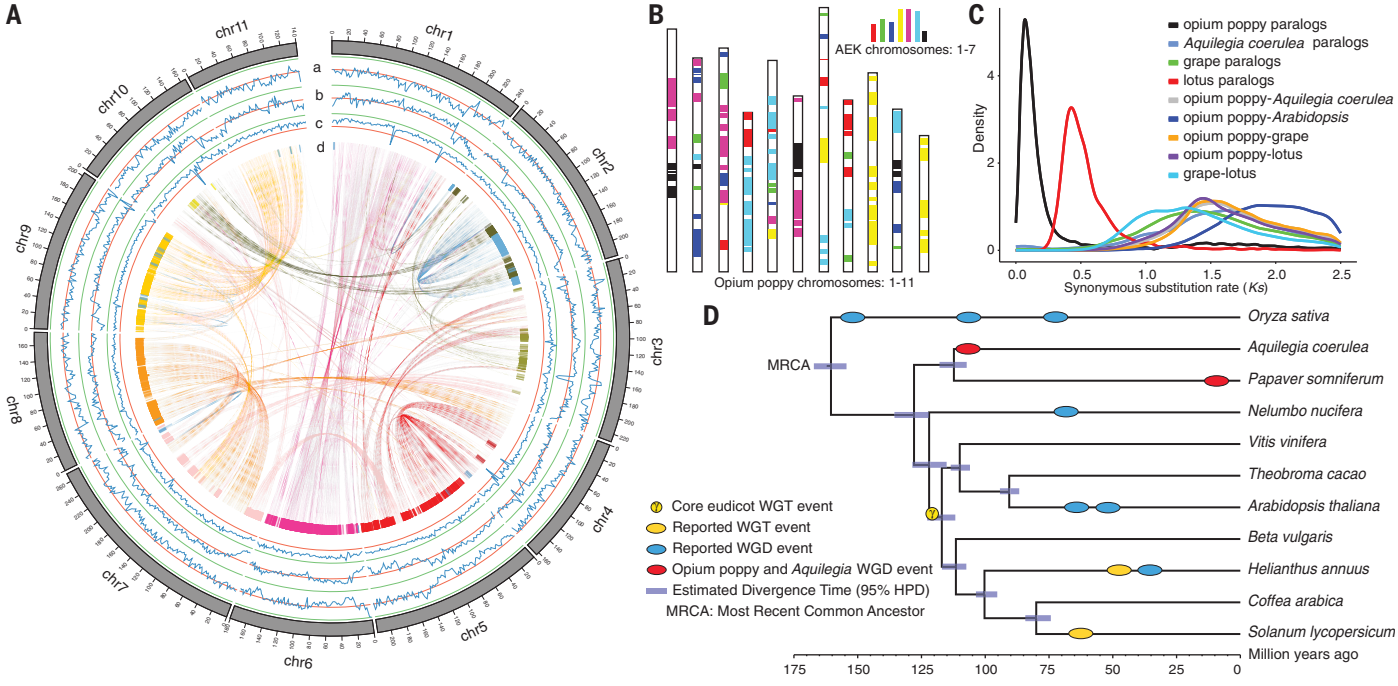


Fig. 1. Opium poppy genome features and whole-genome duplication. (**A**) Characteristics of the 11 chromosomes of *P. somniferum*. Tracks a to c represent the distribution of gene density, repeat density, and GC density, respectively, with densities calculated in 2-Mb windows. Track d shows syntenic blocks. Band width is proportional to syntenic block size. (**B**) Comparison with ancestral eudicot karyotype (AEK) chromosomes reveals synteny. The syntenic AEK blocks are painted onto *P. somniferum* chromosomes. (**C**) Synonymous substitution rate (*Ks*) distributions of syntenic blocks for *P. somniferum* paralogs and orthologs with other eudicots are represented with colored lines as indicated. (**D**) Inferred phylogenetic tree with 48 single-copy orthologs of 11 species identified by OrthoFinder (*13*). Posterior probabilities for all branches exceed 0.99. The timing of the *P. somniferum* and *A. coerulea* whole-genome duplications (WGDs) was estimated in this study, and other reported whole-genome triplication (WGT)/WGD timings are superimposed on the tree. Divergence timings are estimated using BEAST (*14*) and are indicated by light blue bars at the internodes with 95% highest posterior density (HPD).

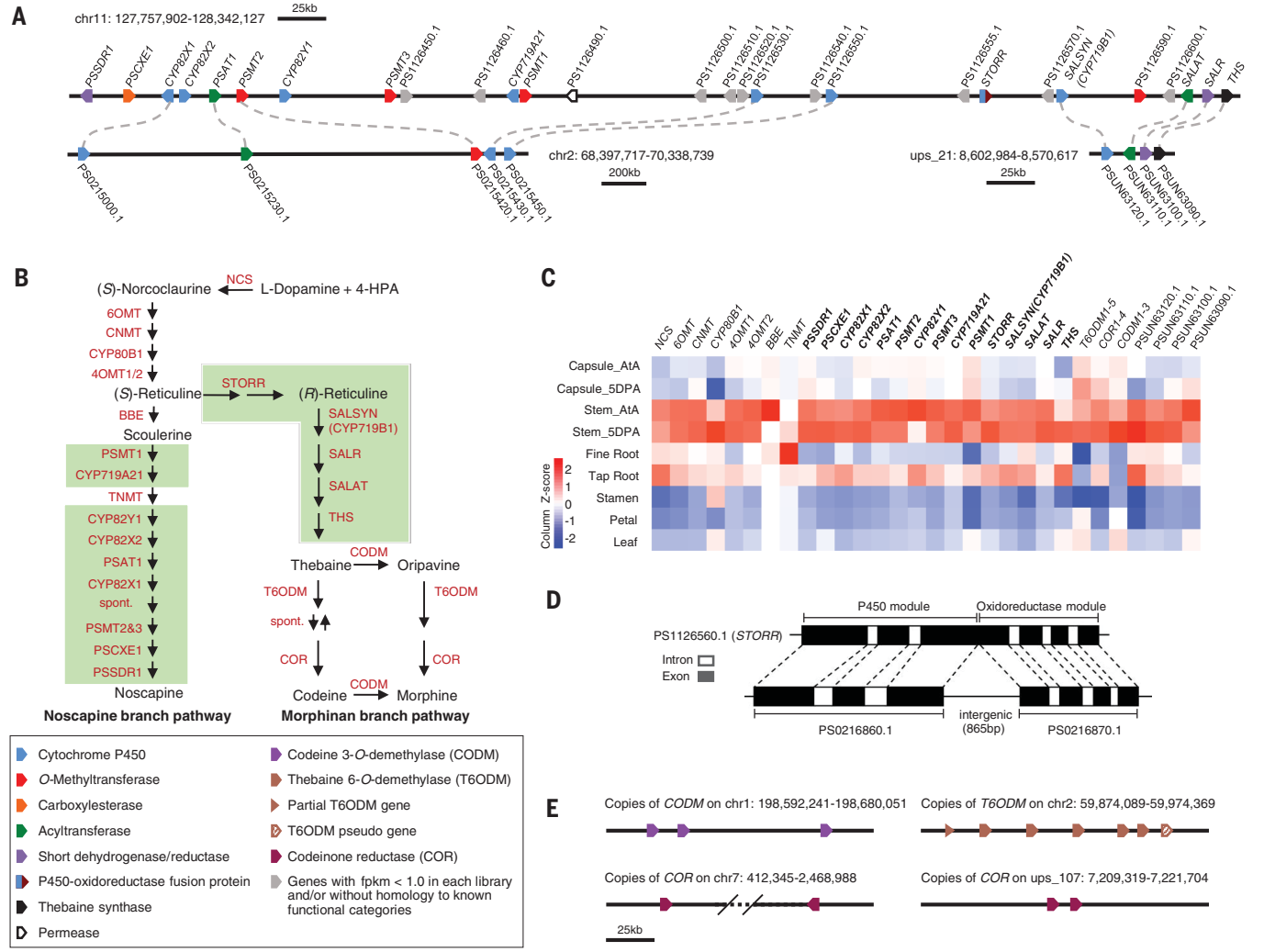


Fig. 2. Genomic arrangement of key genes of BIA metabolism. (**A**) Arrangement and chromosomal position as indicated of the 584-kb BIA gene cluster on chromosome 11 (chr11) encoding 15 coexpressed genes involved in noscapine and morphine biosynthesis (cluster 49, table S16). Below the BIA gene cluster is shown a syntenic block from chr2 (clusters 11 and 12, table S16) associated with the noscapine component of the cluster and a syntenic block from unplaced scaffold 21 associated with the morphinan component (cluster 70, table S16). Syntenic gene pairs are indicated by dashed lines. (**B**) Schematic representation of noscapine and morphinan branch pathways with the reactions associated with BIA gene cluster highlighted in green boxes. "spont." indicates spontaneous reactions. 4-HPA, 4-hydroxyphenylacetaldehyde; fpkm, fragments per kilobase of transcript per million mapped reads. Gene abbreviations are defined in table S15. (**C**) Tissue-specific expression of noscapine and morphinan biosynthesis genes across different tissues. AtA, at anthesis; 5DPA, 5 days postanthesis. BIA pathway genes and their expression values (converted to *z* scores) across different tissues are displayed as a heat map. Genes located in the BIA gene cluster are shown in bold. (**D**) Schematic structure of *STORR* on chr11 and the genomic region on chr2 containing its closest paralogs corresponding to the P450 and reductase modules. Dashed lines denote exon-intron boundaries. (**E**) Arrangement of locally duplicated copies of *CODEINE 3-O-DEMETHYLASE* (*CODM*), *THEBAINE 6-O-DEMETHYLASE* (*T6ODM*), and *CODEINONE REDUCTASE* (*COR*) genes (other annotated genes in the associated genomic regions are not shown).

REDUCTASE (*COR*)—are in a biosynthetic gene cluster (table S15). We used the plantiSMASH genome mining algorithm (*17*) to search the 11 chromosomes and unplaced scaffolds that encode annotated genes for additional gene clusters predicted to be associated with plant specialized metabolism. This approach detected the BIA gene cluster and a number of functionally uncharacterized genes across the same region, among which a cytochrome P450 (PS1126530.1) and a methyltransferase (PS1126590.1) exhibited a similar expression pattern to that of the 15 genes of BIA metabolism (Fig. 2A and table S16). Expression of genes immediately outside the 584-kb BIA gene cluster boundaries was low in aerial tissue (table S16). These genes include *METHYLSTYLOPINE 14-HYDROXYLASE* (*MSH*), which is involved in the sanguinarine branch of BIA metabolism (*18*). *MSH* is expressed in root tissue together with other sanguinarine pathway genes, which are dispersed across the genome (table S15). The plantiSMASH algorithm also found 49 other possible gene clusters across the 11 chromosomes and 34 on unplaced scaffolds, several of which show tissue-specific expression patterns (table S16). Paralogs of the morphinan pathway genes *SALUTARIDINE SYNTHASE* (*SALSYN*), *SALUTARIDINE REDUCTASE* (*SALR*), and *SALUTARIDINOL-7-O-ACETYL TRANSFERASE* (*SALAT*) and the recently discovered *THEBAINE SYNTHASE* (*19*) were identified on an unplaced scaffold in synteny with the BIA biosynthesis gene cluster on chromosome 11 (Fig. 2A and fig. S19 and S20). The expression pattern of these genes matches that of the genes in the BIA cluster (Fig. 2C and table S16).

To investigate the evolutionary history of the BIA gene cluster, we used MCScanX (*20*) to perform two rounds of synteny analysis with either the "all BLASTp" result as input for blocks with distant homology or the default "top 5 BLASTp" result for blocks with close homology. We found that the top-ranked syntenic block with the noscapine branch genes has distant homology and is on chromosome 2 [expect value (E-value) = 8.1×10^{-15}, all BLASTp], whereas the top-ranked syntenic block for morphinan pathway genes has close homology and is on unplaced scaffold 21 (E-value = 0, top 5 BLASTp) (Fig. 2A and tables S17 and S18). *Ks* and amino acid identity of syntenic gene pairs (fig. S21 and table S17) demonstrate that the syntenic block associated with the noscapine branch component of the BIA gene cluster is due to an ancient duplication with a median *Ks* value of 3.9, whereas the syntenic block associated with the morphinan branch component is due to a much more recent duplication occurring in the same time frame as the WGD event around 7.8 Ma ago.

The fusion event resulting in *STORR* was key for evolution of morphinan biosynthesis in the Papaveraceae (Fig. 2B) (*7*). Gene family analysis of P450 and reductase modules of *STORR* revealed their closest paralogs located 865 base pairs (bp) apart on chromosome 2 (Fig. 2D and fig. S22). These paralogs have the same gene orientation and exon-intron boundaries as the *STORR* modules, and on this basis we propose that the *STORR* gene fusion involved an 865-bp deletion after a duplication (Fig. 2D and table S19). *STORR* and its closest paralogs show amino acid sequence identity of 75 and 82% for the P450 and oxidoreductase modules, respectively, which suggests that the duplication leading to the *STORR* gene fusion occurred earlier than the WGD event (fig. S21).

From thebaine, the bifurcated morphinan branch giving rise to codeine and morphine requires three enzymatic reactions, two catalyzed by the 2-oxoglutarate/Fe(II)-dependent dioxygenases, CODM and T6ODM, and a third catalyzed by COR (Fig. 2B). The genome assembly reveals that *CODM* and *T6ODM* are encoded by colocalized gene copies on chromosomes 1 and 2, respectively (Fig. 2E and table S19). Phylogenetic analysis shows that copies of both CODM and T6ODM share more than 97% protein sequence identity, whereas the closest paralogs of CODM and T6ODM share 75.6 and 88.6%, respectively (figs. S21 and S22 and table S19). There are four copies of *COR*, two dispersed on chromosome 7 and two adjacent on unplaced scaffold 107 (Fig. 2E). Copies of COR share more than 95% protein sequence identity, and the closest paralogs share ~74% (figs. S21 and S22 and table S19). This closest-paralog analysis indicates that—as is the case with *STORR*—*T6ODM*, *CODM*, and *COR* emerged before the WGD (fig. S21). Because T6ODM, CODM, and COR use thebaine and downstream intermediates, we assume that the ability to produce thebaine had evolved before the WGD. The near sequence identity between the copies within each of the *T6ODM*, *CODM*, and *COR* gene families indicates that the increase in copy number of these genes occurred more recently than the WGD event. On the basis of the above timing of events, we speculate that the BIA gene cluster was assembled before the WGD event and, after duplication, underwent deletion of the noscapine component and STORR, leaving the morphinan component on unplaced scaffold 21.

The presence of genes exclusively associated with biosynthesis of both phthalideisoquinolines and morphinans in the BIA gene cluster implies a selection pressure that favors clustering of genes associated with these classes of alkaloids. *BBE* and *TNMT* functions are not exclusive for noscapine biosynthesis: Both are required for sanguinarine biosynthesis, which occurs predominantly in root rather than aerial tissues where noscapine and morphine biosynthesis occurs. Selective pressure on *BBE* and *TNMT* associated with their involvement in the biosynthesis of sanguinarine in root tissue may have kept them from being part of the BIA gene cluster even though they are also both expressed in stem tissue (Fig. 2C). Coordinate regulation of gene expression is considered to be part of the selective pressure resulting in gene cluster formation (*21*). In opium poppy, the exclusivity of gene function and complexity of the gene expression pattern could have determined which genes are clustered.

REFERENCES AND NOTES

1. S. Jacomet, *Veg. Hist. Archaeobot.* **18**, 47–59 (2009).
2. F. M. Knaul *et al.*, *Lancet* **391**, 1391–1454 (2018).
3. M. Gates, G. Tschudi, *J. Am. Chem. Soc.* **78**, 1380–1393 (1956).
4. S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante, C. D. Smolke, *Science* **349**, 1095–1100 (2015).
5. A. Nakagawa *et al.*, *Nat. Commun.* **7**, 10390 (2016).
6. T. Winzer *et al.*, *Science* **336**, 1704–1708 (2012).
7. T. Winzer *et al.*, *Science* **349**, 309–312 (2015).
8. S. C. Farrow, J. M. Hagel, G. A. Beaudoin, D. C. Burns, P. J. Facchini, *Nat. Chem. Biol.* **11**, 728–732 (2015).
9. M. S. Campbell, C. Holt, B. Moore, M. Yandell, *Curr. Protoc. Bioinformatics* **48**, 4.11.1–4.11.39 (2014).
10. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, *Bioinformatics* **31**, 3210–3212 (2015).
11. O. Jaillon *et al.*, *Nature* **449**, 463–467 (2007).
12. F. Murat, A. Armero, C. Pont, C. Klopp, J. Salse, *Nat. Genet.* **49**, 490–496 (2017).
13. D. M. Emms, S. Kelly, *Genome Biol.* **16**, 157 (2015).
14. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
15. Y. Jiao *et al.*, *Genome Biol.* **13**, R3 (2012).
16. The *Aquilegia coerulea* v3.1 genome data (released 8 August 2014; http://phytozome.jgi.doe.gov/) were produced by the U.S. Department of Energy Joint Genome Institute.
17. S. A. Kautsar, H. G. Suarez Duran, K. Blin, A. Osbourn, M. H. Medema, *Nucleic Acids Res.* **45**, W55–W63 (2017).
18. G. A. W. Beaudoin, P. J. Facchini, *Biochem. Biophys. Res. Commun.* **431**, 597–603 (2013).
19. X. Chen *et al.*, *Nat. Chem. Biol.* **14**, 738–743 (2018).
20. Y. Wang *et al.*, *Nucleic Acids Res.* **40**, e49 (2012).
21. A. Osbourn, *Plant Physiol.* **154**, 531–535 (2010).

ACKNOWLEDGMENTS

We thank Y. Jiao and L. Chen for helpful discussions on genome analysis and J. Hai, J. Mitchell, S. Donninger, and J. Hudson for administrative and technical support. This work involved the use of the York Bioscience Technology Facility. **Funding:** K.Y., L.G., and X.Y. are supported by the National Science Foundation of China (31671372, 31701739, and 61702406), the National Key R&D Program of China (2017YFC0907500 and 2018YFC0910400), "the Thousand Talents Plan," and the Key Construction Program of the National "985" Project. Z.N. is funded by the Wellcome Trust (WT098051), and I.A.G. received support from the UK Biotechnology and Biological Sciences Research Council (BB/K018809/1) and the Garfield Weston Foundation. **Author contributions:** K.Y., Z.N. and I.A.G. designed and supervised research. T.W., R.T., T.A.B., and Y.Li collected materials for sequencing and generated transcriptome data and linkage maps. Z.N. performed the genome assembly. X.Y., L.G., Y.Lu, T.W., and Y.Li performed genome annotation and synteny analysis. L.G., X.Y., Y.Li, and T.W. performed transcriptome and phylogenomic analysis. T.W., Y.Li, Z.H., X.Y., L.G., K.Y., and I.A.G. interpreted evolution of BIA metabolism. X.Y., L.G., Z.N., Y.Li, and T.W. prepared figures and tables. I.A.G., L.G., T.W., X.Y., Y.Li, Z.N., and K.Y. wrote the paper. **Competing interests:** None declared. **Data and materials availability:** All raw reads and assembled sequence data have been uploaded to NCBI under BioProjectID PRJNA435796. The Whole Genome Shotgun project has been deposited under accession number PUWZ00000000. The nine transcriptome datasets have accession numbers GSM3022361 to GSM3022369, and the 15 BAC sequences have accession numbers MG995579 to MG995593. HN1 is available from the University of York on behalf of Sun Pharmaceutical Industries Australia for experimental purposes, subject to a material transfer agreement.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/343/suppl/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S19
References (*22*–*76*)
Data S1 and S2

25 February 2018; resubmitted 26 May 2018
Accepted 21 August 2018
Published online 30 August 2018
10.1126/science.aat4096

SCIENCE AND SOCIETY

Realizing private and practical pharmacological collaboration

Brian Hie[1]*, Hyunghoon Cho[1]*, Bonnie Berger[1,2]†

Although combining data from multiple entities could power life-saving breakthroughs, open sharing of pharmacological data is generally not viable because of data privacy and intellectual property concerns. To this end, we leverage modern cryptographic tools to introduce a computational protocol for securely training a predictive model of drug–target interactions (DTIs) on a pooled dataset that overcomes barriers to data sharing by provably ensuring the confidentiality of all underlying drugs, targets, and observed interactions. Our protocol runs within days on a real dataset of more than 1 million interactions and is more accurate than state-of-the-art DTI prediction methods. Using our protocol, we discover previously unidentified DTIs that we experimentally validated via targeted assays. Our work lays a foundation for more effective and cooperative biomedical research.

Collaborative efforts to develop new, life-saving drug therapies have recently begun to take shape among pharmaceutical companies and academic labs, despite the highly competitive nature of the industry (*1*–*5*). Driving this transformation is the stalled or declining productivity of existing drug development pipelines amidst growing financial and regulatory pressures. Many in industry and academia are realizing that the difficult task of identifying novel drug candidates would be more successful if they could leverage pooled experimental datasets and knowledge that go beyond any single organization (*6*–*8*).

Until now, however, such forms of collaboration (*1*–*3*), including open-access data sharing partnerships like the Structural Genomics Consortium (www.thesgc.org), have been of limited scope because pharmacological data sharing is fundamentally restricted by concerns about intellectual property and other financial interests. Currently, entities must moderate the amount of data they share in order to maintain the confidentiality of drugs under development or the set of potential targets being tested, both of which may contain sensitive information about underlying research or business strategies.

Modern cryptography offers techniques to broaden pharmacological collaboration by greatly mitigating privacy concerns associated with data sharing. Secure multiparty computation (MPC) protocols (*9*) allow multiple entities to compute over their private datasets without revealing any information about the underlying raw data, except for the final computational output. Unfortunately, the promise of privacy-preserving collaboration has been severely hindered by the inability of existing secure computation frameworks to scale to complex computations over large datasets. In particular, analyzing a large amount of experimental data to predict new therapeutic interactions typically involves sophisticated algorithms that present a major computational challenge for MPC.

Here we introduce a proof-of-concept, end-to-end pipeline for collaborative drug–target interaction (DTI) prediction based on secure MPC. Conceptually, our protocol divides computation across collaborating entities while ensuring that none of the entities have any knowledge about the private data (Fig. 1). We achieve this result by using a cryptographic framework known as "secret sharing" (*10*) in which a private value (a "secret") is collectively represented by multiple entities. Each entity is given a random number (a "share") in a finite field (i.e., integers modulo some prime number p) such that the sum of all shares modulo p equals the secret. Any strict subset of entities cannot extract any information about the underlying secret using the subset's shares. Various protocols have been developed for performing elementary operations (e.g., addition or multiplication) over secret-shared inputs (*10*, *11*), which, taken together, form the building blocks for a general-purpose MPC framework.

Although secret sharing–based MPC typically requires overwhelming amounts of data communication between entities for complex and large-scale computations, very recent optimizations have leveraged techniques such as generalized Beaver multiplication triples and shared pseudorandom number generators to substantially reduce communication cost, thus enabling practical secure computation for challenging problems such as genome-wide association studies for 1 million individuals (*12*). Even with these advances, however, secure MPC is still infeasible for existing DTI prediction methods (*13*–*16*), primarily because their computations scale quadratically with the number of drugs (n) and the number of targets (m) in the dataset (e.g., n^2 or nm), which is prohibitive for realistic datasets with millions of compounds.

To achieve scalable computation while maintaining high accuracy, we draw from recent advances in deep learning (*17*) to train a neural network model for DTI prediction. Our neural network takes feature representations of a compound and a target as input and predicts the interactivity of the given pair. Although we use chemical structure fingerprints and protein domain annotations as input features in our computational experiments (materials and methods), our framework readily generalizes to alternative features. We circumvent the quadratic complexity of existing methods by training our neural network over a dataset consisting of only the observed DTIs and a comparable number of putatively noninteracting drug–target pairs; this dataset is typically much smaller than the full drug-by-target matrix. Furthermore, we greatly reduce the cryptographic overhead of secure neural network training by optimizing our architectural choices for efficient MPC, such as using the rectifier (*18*) as our activation function and hinge loss as our loss function, both of which require only a single data-oblivious comparison to evaluate (materials and methods). These operations can be more efficiently implemented in MPC than alternatives such as the sigmoid function, which requires many such comparisons to accurately approximate. Taken together, our efficient protocol allows our neural network to securely train over a wide area network (WAN) in less than 4 days on a dataset with more than 1 million training instances (table S1). In contrast, a recently proposed protocol for privacy-preserving neural network training (*19*) requires months of communication time over a WAN to train on an image dataset of a smaller scale (60,000 examples, 784 features).

To demonstrate the accuracy of our securely trained neural network for DTI prediction (named Secure DTI), we compared it to state-of-the-art DTI prediction techniques, including those based on matrix factorization with side information (CMF) (*14*), network diffusion (NetLapRLS and BLMNII) (*16*, *20*), and heterogeneous data integration (DTINet and HNM) (*15*, *21*) on a standard benchmark dataset (*22*) with 1923 observed interactions (materials and methods). In addition to newly ensuring the confidentiality of the pooled input data during the computation, Secure DTI surpasses the performance of all baseline plaintext methods in cross-validation accuracy (Fig. 2A and fig. S1), a surprising result in light of the optimizations we made to achieve practical scalability. Our improvement over the best-performing baseline method (DTINet) is statistically significant (one-sided Wilcoxon rank-sum P value = 0.006) and further pronounced in a more challenging but realistic evaluation setting where the entire interaction space is considered as the test data rather than a balanced number of positive and negative examples (fig. S1F). None of the baseline methods can be efficiently implemented in MPC for the purposes of secure collaboration, owing to their quadratic

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Department of Mathematics, MIT, Cambridge, MA 02139, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: bab@mit.edu

complexity in the number of drugs and targets. In contrast, matrix factorization without side information (MF) lends itself to efficient MPC (*23*) but at the cost of greatly diminished model performance (Fig. 2, A and B, and fig. S1).

We next set out to demonstrate the scalability and predictive performance of Secure DTI on a much larger dataset that more accurately represents the scale of cross-institutional collaboration. We obtained 969,817 interactions from the STITCH 5 human dataset (*24*), which is, to our knowledge, the largest publicly available DTI dataset, and evaluated the cross-validation performance of Secure DTI. Even on the challenging task of predicting DTIs of previously unseen compounds, Secure DTI achieved high accuracy [area under the precision-recall curve (AUPR) of 0.95], which substantially outperforms matrix factorization methods (AUPRs of 0.50 and 0.43) (Fig. 2B and fig. S2). Owing to their quadratic scalability, other baseline methods could not be reasonably applied to a dataset of this size (even in plaintext). In contrast, Secure DTI took less than 4 days to train on millions of interactions over a WAN (materials and methods) and efficiently scaled with a linear dependence on the number of interactions in the dataset (Fig. 2C and table S1). Even when training Secure DTI on 2 million interactions, we extrapolate the total runtime for one epoch (one linear pass over the full, shuffled training set) to be ~2.2 days. In practice, we expect our model to achieve high accuracy in only a few training epochs; we obtained all of our reported results after 1.5 epochs. Additionally, given that our protocol admits flexibility in the choice of predictive model, we also securely trained a

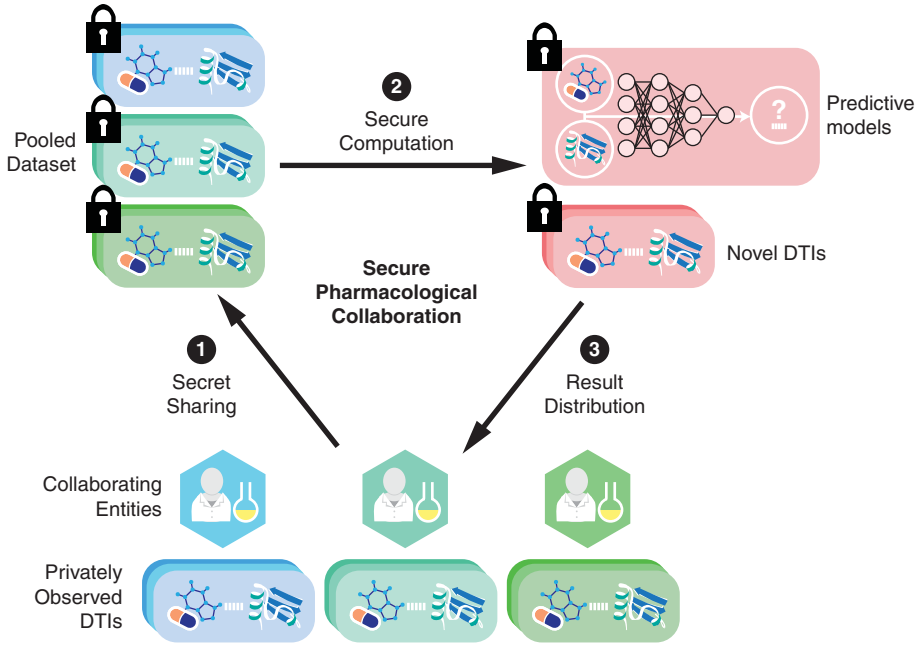


Fig. 1. Secure pipeline for pharmacological collaboration. Collaborating entities (e.g., pharmaceutical companies or research laboratories) have large private datasets of DTIs, as well as corresponding chemical structures and protein sequences. In our protocol, the entities first use secret sharing to pool their data in a way that reveals no information about the underlying drugs, targets, or interactions (step 1). The collaborating entities then jointly execute a cryptographic protocol that trains a predictive model (e.g., a neural network) on the pooled dataset (step 2). The final model can be made available to participating entities or may be used to distribute DTI predictions to participants in a way that encourages greater data sharing (step 3).

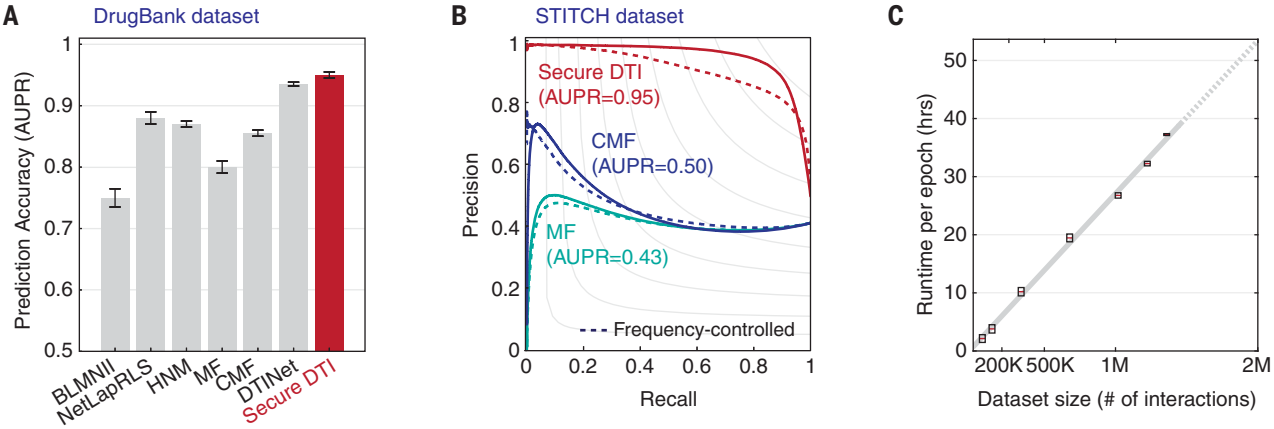


Fig. 2. Prediction of DTIs. (**A**) Predictions from the DrugBank 3.0 dataset. Bar height corresponds to mean AUPR (area under the precision-recall curve), and error bars indicate SD. We compared Secure DTI to the plaintext methods BLMNII (*20*), NetLapRLS (*16*), HNM (*21*), MF (*13*), CMF (*14*), and DTINet (*15*), as reported in Luo *et al*. (*15*), by means of 10-fold cross-validation on balanced training and test sets (materials and methods; see fig. S1 for other evaluation settings). (**B**) Predictions form the STITCH 5 dataset with more than 1 million drug–target pairs. Secure DTI is compared with matrix factorization with (CMF) and without (MF) side information (see fig. S2 for other evaluation settings). Solid lines, sampling negative examples randomly; dashed lines, sampling negative examples while matching the relative frequencies of drugs and targets to those in the positive examples, representing a more challenging test case. Reported AUPRs are for the solid curves. (**C**) Runtime of our training protocol, over a local area network (LAN), for different dataset sizes. Box height represents SD.

Table 1. Predicted out-of-dataset DTIs. We trained Secure DTI on all human DTIs from STITCH 5, which we used to score and rank all pairs of drugs and targets that are not in the STITCH database. We implemented two methods to control for model bias toward overrepresented drugs and targets, either (i) filtering out predictions involving a drug and target that are both highly represented in STITCH (Secure DTI-A) or (ii) retraining Secure DTI such that the negative training examples had an equal representation of drugs and targets compared with the positive training examples (Secure DTI-B) (materials and methods). Interactions labeled N/A involve commercially unavailable compounds and thus could not be tested. "Active" interactions have median inhibitory concentrations <100 μM, whereas "inconclusive" interactions demonstrate observable activity only at one or two high-concentration levels, a potential artifact of compound aggregation. We labeled the interaction between actinomycin D and PARP1 as "weakly active" because consistent activity was observed over a wide range of concentrations, including near-50% inhibition at 100 μM (our highest tested concentration). However, it should be noted that its dose-response curve does not follow a typical sigmoidal shape. References and additional information are provided in table S2.

Rank	Secure DTI-A			Secure DTI-B		
	Drug	Target	Experimental validation	Drug	Target	Experimental validation
1	Droloxifene	ERα	Active*	Droloxifene	ERβ	Active*
2	Droloxifene	ERβ	Active*	CHEMBL601690	p110α	Active
3	Imatinib	ErbB3	Inconclusive*	Droloxifene	ERα	Active*
4	Imatinib	ErbB4	Active*	Seocalcitol	VDR	Active
5	Nutlin-3	PARP1	Inactive*	AGN-PC-0A9TBG	PPARγ	Active
6	Droloxifene	PgR	Inactive*	CHEMBL589864	p110α	Active
7	Actinomycin D	PARP1	Weakly active*	T5958429	PARP1	Active
8	Hoechst 33258	PARP1	Inactive*	AGN-PC-0N7PYE	Factor Xa	Active
9	GW-501516	GR	Inactive*	AGN-PC-00DJ3O	PPARγ	Active
10	AGN-PC-0BFP0W	Lck	Active	AGN-PC-0NA8NJ	PTPRZ1	N/A
11	CHEMBL2332055	mGluR1	Inconclusive*	AGN-PC-0NA8NJ	PTPRG	N/A
12	CHEMBL2332055	mGluR5	Inconclusive*	AGN-PC-088DZ9	PROC	N/A

*Predicted interactions, including all testable interactions without existing literature support, that were experimentally validated in our study.

support vector machine (SVM) instead of a neural network; the SVM reduced predictive performance (fig. S2).

To go beyond cross-validation and demonstrate the potential for novel discoveries that can result from our collaborative pipeline, we trained Secure DTI on all STITCH 5 interactions and scored the remaining possible drug–target pairs for interactivity, which is closer to how our pipeline would be used in a real-world setting. We controlled for bias toward highly represented drugs and targets in the dataset by either (i) filtering out any prediction involving both a drug and target highly represented in the original dataset (Secure DTI-A) or (ii) sampling negative examples (i.e., noninteractions) during model training such that each drug or target was seen at the same relative frequency in the negative examples as in the positive examples (Secure DTI-B). In both cases, many of our top predictions (5 of 12 for Secure DTI-A and 9 of 12 for Secure DTI-B) were validated by our own targeted assay experiments or by published experimental studies that have not yet been deposited into the STITCH database (Table 1). Our validation experiments suggest a previously unknown interaction between imatinib and ErbB4, for which we could not find any existing experimental support. It will be interesting to find out whether this interaction is confirmed by other studies. The top prediction from both methods was an interaction between the estrogen receptor (ER) and droloxifene, which had reached phase 3 clinical trials as an ER modulator for advanced breast cancer (*25*). Similarly, the predicted interaction between the vitamin D receptor (VDR) and seocalcitol has been clinically well established (*26*). Furthermore, some predictions without direct activity have strong evidence for an indirect functional interaction; for example, nutlin-3 has been shown to inhibit poly(ADP-ribose) polymerase 1 (PARP1) protein levels through p53-dependent proteasomal degradation in mouse fibroblasts (*27*). All of our findings were obtained without revealing any information about the underlying drugs, targets, and interactions during the computation.

To provide enhanced interpretability of our reported results, we incorporated an additional step into our pipeline for securely evaluating the impact of individual input features on the prediction outcome (supplementary text). When applied to our top predictions from the STITCH database, this capability linked drugs to specific ligand-binding or functional sites within the predicted target (table S3).

We envision a real-world scenario in which multiple participating entities contribute secret-shared data to train a machine learning model on a joint pharmacological dataset (Fig. 1). After training, the model could be made available to all participants or could remain private such that entities receive a number of predictions commensurate with the amount of data they contribute in order to incentivize participation. As more training data will most likely result in greatly improved performance (fig. S3 and supplementary text), entities will be incentivized to share information in a way that is mutually beneficial and has provable privacy guarantees.

Our pipeline is secure under the "honest-but-curious" security model in which the collaborating entities follow the correct protocol and do not collude to reconstruct the data. This is a substantial improvement over the current state of biomedical research, in which privacy concerns hinder any collaboration, but our framework can also be extended to achieve even stronger security guarantees. Because our security guarantee holds as long as at least one entity is honest during the main computation (materials and methods), we can relax the no-collusion requirement by introducing additional collaborating entities into our protocol, which does not substantially increase total computation time but does increase communication linearly in the number of entities. If we require security against malicious entities who deviate from the protocol during the online computation, we can include a message authentication code (MAC) with each message. At the end of the protocol, the MAC is verified to ensure that each step was performed according to the protocol specification, a technique introduced in the SPDZ framework (*28*). This approach roughly doubles computation and communication, offering a trade-off between security and performance that can be adjusted according to specific study requirements.

Although we did not add noise to the final computation output of our pipeline to limit

information leakage, a technique known as differential privacy (*29*, *30*), a method being developed for differentially private neural networks can be used in conjunction with our protocol (*31*). An alternative strategy for collaborative neural network–based prediction is to train local models in plaintext and use secure protocols only when periodically averaging over these models, thus minimizing the amount of cryptographic overhead (*32*, *33*). However, this approach is vulnerable to reverse engineering–based attacks in which a malicious collaborator jointly trains a local model (e.g., a generative adversarial network) that uncovers information about private data owned by honest collaborators, even when differential privacy techniques are applied (*34*). In contrast, securely training a single model over a decentralized network of computing parties, as in our pipeline, is not vulnerable to such attacks.

Our privacy-preserving protocols generalize to other large-scale data sharing problems beyond drug discovery, with the highest potential for impact in areas that suffer from a lack of collaboration due to privacy concerns, such as predictive analyses of electronic health records. Our practical demonstration of secure, large-scale machine learning with neural networks may also provide a useful blueprint for enhancing privacy in many other domains where neural networks have been shown to be successful.

REFERENCES AND NOTES

1. S. Reardon, *Nature* 10.1038/nature.2014.14672 (2014).
2. J. Levy, The age of collaboration: why pharma companies now have to work together. *Pharmafile* (2015); www.pharmafile.com/news/501725/age-collaboration-why-pharma-companies-now-have-work-together.
3. M. Wilhelm, Big Pharma Buys Into Crowdsourcing for Drug Discovery. *Wired* (2017); www.wired.com/story/big-pharma-buys-into-crowdsourcing-for-drug-discovery/.
4. J. Hunter, Collaboration for innovation is the new mantra for the pharmaceutical industry. *Drug Discovery World* (2014); www.ddw-online.com/business/p217613-collaboration-for-innovation-is-the-new-mantra-for-the-pharmaceutical-industry-spring-14.html.
5. "Johnson & Johnson Innovation Announces New Collaborations Advancing Ground-Breaking Biomedical Innovation Around the Globe," press release; www.jnjinnovation.com/sites/default/files/jji_bio_2017_press_release_06-15-17.pdf.
6. I. Khanna, *Drug Discov. Today* **17**, 1088–1102 (2012).
7. D. Cressey, *Nature* **471**, 17–18 (2011).
8. S. M. Paul *et al*., *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
9. R. Cramer, I. B. Damgård, J. B. Nielsen, *Secure Multiparty Computation and Secret Sharing* (Cambridge Univ. Press, ed. 1, 2015).
10. M. Ben-Or, S. Goldwasser, A. Wigderson, in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing* (ACM, 1988), pp. 1–10.
11. O. Catrina, A. Saxena, in *Financial Cryptography and Data Security*, R. Sion, Ed. (Lecture Notes in Computer Science Series, Springer, 2010), pp. 35–50.
12. H. Cho, D. Wu, B. Berger, *Nat. Biotechnol.* **36**, 547–551 (2018).
13. M. C. Cobanoglu, C. Liu, F. Hu, Z. N. Oltvai, I. Bahar, *J. Chem. Inf. Model.* **53**, 3399–3409 (2013).
14. X. Zheng, H. Ding, H. Mamitsuka, S. Zhu, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2013), pp. 1025–1033.
15. Y. Luo *et al*., *Nat. Commun.* **8**, 573 (2017).
16. Z. Xia, L.-Y. Wu, X. Zhou, S. T. C. Wong, *BMC Syst. Biol.* **4** (suppl. 2), S6 (2010).
17. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436–444 (2015).
18. X. Glorot, A. Bordes, Y. Bengio, in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, M. Dudík, Eds. (JMLR W&CP, 2011), vol. 15, pp. 315–323.
19. P. Mohassel, Y. Zhang, in *Proceedings of the 2017 IEEE Symposium on Security and Privacy* (IEEE, 2017), pp. 19–38.
20. J. P. Mei, C. K. Kwoh, P. Yang, X. L. Li, J. Zheng, *Bioinformatics* **29**, 238–245 (2013).
21. W. Wang, S. Yang, X. Zhang, J. Li, *Bioinformatics* **30**, 2923–2930 (2014).
22. C. Knox *et al*., *Nucleic Acids Res.* **39** (suppl. 1), D1035–D1041 (2011).
23. V. Nikolaenko *et al*., in *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security* (ACM, 2013), pp. 801–812.
24. D. Szklarczyk *et al*., *Nucleic Acids Res.* **44**, D380–D384 (2016).
25. A. Buzdar *et al*., *Breast Cancer Res. Treat.* **73**, 161–175 (2002).
26. G. Tocchini-Valentini, N. Rochel, J. M. Wurtz, D. Moras, *J. Med. Chem.* **47**, 1956–1961 (2004).
27. S. Matsushima *et al*., *Biochem. Biophys. Res. Commun.* **407**, 557–561 (2011).
28. I. Damgård, V. Pastro, N. Smart, S. Zakarias, in *Advances in Cryptology – CRYPTO 2012*, R. Safavi-Naini, R. Canetti, Eds. (Lecture Notes in Computer Series, Springer, 2012), vol. 7417, pp. 643–662.
29. C. Dwork, F. McSherry, K. Nissim, A. Smith, *Theory Cryptogr.* 265–284 (2006).
30. S. Simmons, C. Sahinalp, B. Berger, *Cell Syst.* **3**, 54–61 (2016).
31. M. Abadi *et al*., in *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security* (ACM, 2016), pp. 308–318.
32. K. Bonawitz, *et al.*, in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (ACM, 2017), pp. 1175–1191.
33. R. Shokri, V. Shmatikov, in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing* (IEEE, 2016), pp. 909–910.
34. B. Hitaj, G. Ateniese, F. Perez-Cruz, arXiv:1702.07464 [cs.CR] (24 February 2017).

ACKNOWLEDGMENTS

We thank E. Irvine and P. Macaluso for assistance with validation experiments. **Funding:** B.H. and H.C. are partially supported by NIH grant R01GM081871 (to B.B.). H.C. is also partially supported by the Kwanjeong Educational Foundation. **Author contributions:** B.H., H.C., and B.B. developed the computational methods. B.H. and H.C. implemented the protocol and performed the analyses. B.B. guided the research. All authors wrote the manuscript. **Competing interests:** A provisional patent application on this work (serial no. 62/611,678) was filed on 29 December 2017. **Data and materials availability:** Code and data are available at http://secure-dti.csail.mit.edu. Drug–target interaction datasets were obtained from DrugBank version 3.0 (www.drugbank.ca/) and STITCH version 5.0 (http://stitch.embl.de/). Protein domain information is available from the Pfam database (https://pfam.xfam.org/).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/347/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Tables S1 to S3
References (*35*–*58*)

2 March 2018; accepted 28 August 2018
10.1126/science.aat4807

IMMUNOLOGY

DNGR-1 in dendritic cells limits tissue damage by dampening neutrophil recruitment

Carlos del Fresno[1]*†, Paula Saz-Leal[1]*, Michel Enamorado[1]‡, Stefanie K. Wculek[1], Sarai Martínez-Cano[1], Noelia Blanco-Menéndez[1], Oliver Schulz[2], Mattia Gallizioli[3], Francesc Miró-Mur[3], Eva Cano[4], Anna Planas[3,5], David Sancho[1]†

Host injury triggers feedback mechanisms that limit tissue damage. Conventional type 1 dendritic cells (cDC1s) express dendritic cell natural killer lectin group receptor-1 (DNGR-1), encoded by the gene *Clec9a*, which senses tissue damage and favors cross-presentation of dead-cell material to CD8$^+$ T cells. Here we find that DNGR-1 additionally reduces host-damaging inflammatory responses induced by sterile and infectious tissue injury in mice. DNGR-1 deficiency leads to exacerbated caerulein-induced necrotizing pancreatitis and increased pathology during systemic *Candida albicans* infection without affecting fungal burden. This effect is B and T cell–independent and attributable to increased neutrophilia in DNGR-1–deficient settings. Mechanistically, DNGR-1 engagement activates SHP-1 and inhibits MIP-2 (encoded by *Cxcl2*) production by cDC1s during *Candida* infection. This consequently restrains neutrophil recruitment and promotes disease tolerance. Thus, DNGR-1–mediated sensing of injury by cDC1s serves as a rheostat for the control of tissue damage, innate immunity, and immunopathology.

After sterile or infectious insults, injured tissues expose or release alarm signals that are detected by specific innate immune receptors on myeloid cells (*1*). This triggers an inflammatory response, which promotes the recruitment of myeloid cells into the damaged organ. This innate immune response must be tightly regulated to avoid additional tissue damage (*2*).

Among myeloid cell sensors of tissue damage, dendritic cell natural killer lectin group receptor-1 (DNGR-1; *Clec9a* gene) is a C-type lectin receptor (CLR) that detects F-actin exposed by damaged cells (*3*, *4*). DNGR-1 is mainly expressed by mouse and human conventional type 1 dendritic cells (cDC1s), including CD103$^+$CD11b$^-$ DCs in peripheral tissues (*5*, *6*). DNGR-1 favors the cross-presentation of dead cell–associated antigens to CD8$^+$ T cells (*7*–*9*). However, whether DNGR-1 plays any role in innate immunity is unknown. To address this issue, we used a mouse model of caerulein-induced acute necrotizing pancreatitis (Fig. 1A), which results in massive acinar cell death, leading to the infiltration of myeloid cells. This, in turn, triggers further pathology and edema (*10*). Upon caerulein treatment, there was increased pancreatic infiltration by neutrophils, but not monocytes, in DNGR-1–deficient (*Clec9a*$^{gfp/gfp}$) mice compared with that observed in wild-type (WT) mice (Fig. 1B). Neutrophil numbers in the bone marrow (BM) and blood were similar in both genotypes (fig. S1), ruling out an effect of DNGR-1 deficiency on neutrophil ontogeny and suggesting a local recruitment effect.

As a nongenetic approach, we used a DNGR-1–blocking antibody (*7*, *11*). Receptor blockade phenocopied the exacerbated pancreatic infiltration of neutrophils (Fig. 1C) but not monocytes (fig. S2A). Enhanced neutrophilia upon DNGR-1 blockade was lost in *Batf3*$^{-/-}$ mice (Fig. 1D and fig. S2B), which lack functional cDC1s (*12*), indicating that cDC1s are the key mediators. Pancreatic CXCR2-mediated neutrophil infiltration is pathological in acute pancreatitis (*13*). Consistently, caerulein-treated *Clec9a*$^{gfp/gfp}$ mice displayed exacerbated pancreatitis with increased serum lipase concentrations (Fig. 1E) and extended pancreatic edema (Fig. 1F).

The rapid kinetics of neutrophil infiltration suggested the involvement of an innate immune response. To test this, *Rag1*$^{-/-}$ (lacking B and T cells) and *Rag1*$^{-/-}$*Clec9a*$^{gfp/gfp}$ mice were subjected to caerulein-induced acute pancreatitis. Notably, the absence of DNGR-1 resulted in enhanced neutrophil infiltration (Fig. 1G and fig. S2C) and increased circulating lipase concentrations (Fig. 1H) in B and T cell–deficient mice. Thus, after tissue damage, DNGR-1 expressed on cDC1s regulates the recruitment of neutrophils without the involvement of B and T cells.

A reduction of neutrophil-mediated immunopathology is associated with disease tolerance upon infection, which limits the impact of damage-generating infectious challenges on host fitness without affecting pathogen burden (*14*, *15*). To test whether DNGR-1 affects disease tolerance, we used systemic *Candida albicans* infection, which generates extensive renal tissue necrosis (*16*). DNGR-1–deficient mice showed increased morbidity and mortality upon systemic candidiasis (Fig. 2, A and B), despite having a similar fungal burden (Fig. 2C). Extended pathology in the absence of DNGR-1 correlated with increased neutrophil infiltration in the kidney (Fig. 2, D and E). Neutrophil numbers in BM or blood of WT and *Clec9a*$^{gfp/gfp}$ mice were similar (fig. S3). DNGR-1 blockade in infected mice phenocopied increased neutrophilia (Fig. 2F), which was prevented in BATF3-deficient mice (Fig. 2G), indicating that cDC1s mediate the effect. Of note, *Rag1*$^{-/-}$*Clec9a*$^{gfp/gfp}$ mice also showed increased renal neutrophil numbers (Fig. 2H) and reduced survival after infection (Fig. 2I). Monocyte recruitment into *Candida*-infected kidneys was not significantly increased in any of the DNGR-1–deficient conditions (right panel of Fig. 2D and fig. S4). Thus, DNGR-1 dampens the recruitment of neutrophils to damaged tissues in both sterile and infectious settings in a B and T cell–independent manner.

Neutrophil-mediated renal immunopathology causes acute kidney failure and mortality during systemic candidiasis (*17*, *18*). Consistently, *C. albicans*–infected *Clec9a*$^{gfp/gfp}$ mice showed exacerbated kidney damage, with increased terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling (TUNEL)–positive cells (Fig. 2J), increased concentrations of serum creatinine (Fig. 2K), and enhanced expression of kidney injury molecule-1 (KIM-1) (Fig. 2L). Kidney neutrophilia was increased in *Clec9a*$^{gfp/gfp}$ mice three days after infection (Fig. 2M), along with enhanced KIM-1 expression (Fig. 2N). Thus, exacerbated renal damage caused by neutrophils could underlie increased pathology in *Candida*-infected *Clec9a*$^{gfp/gfp}$ mice.

We tested whether DNGR-1–regulated neutrophilia drives tissue damage in sterile pancreatitis (Fig. 3A). Partial depletion of neutrophils with an antibody against Ly6G (anti-Ly6G, or 1A8 antibody) (fig. S5) reverted the enhanced edematous lesions found in isotype-treated *Clec9a*$^{gfp/gfp}$ mice upon caerulein treatment (Fig. 3B). Assessing the impact of neutrophils in *C. albicans* infection is more complex, because the depletion of neutrophils is lethal (*19*). To circumvent this, we first used fungizone to eliminate the fungus starting at day 3 postinfection (Fig. 3C and fig. S6A), after initial tissue damage by the infection (Fig. 2N). Removal of *C. albicans* did not affect the exacerbated neutrophil infiltration (Fig. 3D) or renal damage (Fig. 3E) observed in *Clec9a*$^{gfp/gfp}$ mice. Thus, after the initial damage, the presence of fungus was not essential for the DNGR-1–dependent effect. Neutrophil depletion with 1A8 antibody in the presence of fungizone (Fig. 3, C and F, and fig. S6B) prevented the enhanced renal damage found in isotype-treated *Clec9a*$^{gfp/gfp}$ mice (Fig. 3, G and H), even though fungal burden was equivalent between genotypes (Fig. 3I).

[1]Immunobiology Laboratory, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. [2]Immunobiology Laboratory, The Francis Crick Institute, London, UK. [3]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. [4]Chronic Disease Programme–CROSADIS, Instituto De Salud Carlos III, Madrid, Spain. [5]Department of Brain Ischemia and Neurodegeneration, Institut d'Investigacions Biomèdiques de Barcelona (IIBB-CSIC), Barcelona, Spain.
*These authors contributed equally to this work.
†Corresponding author. Email: dsancho@cnic.es (D.S.); cdelfresno@cnic.es (C.d.F.) ‡Present address: Mucosal Immunology Section, Laboratory of Parasitic Diseases, NIAID, NIH, Bethesda, MD, USA.

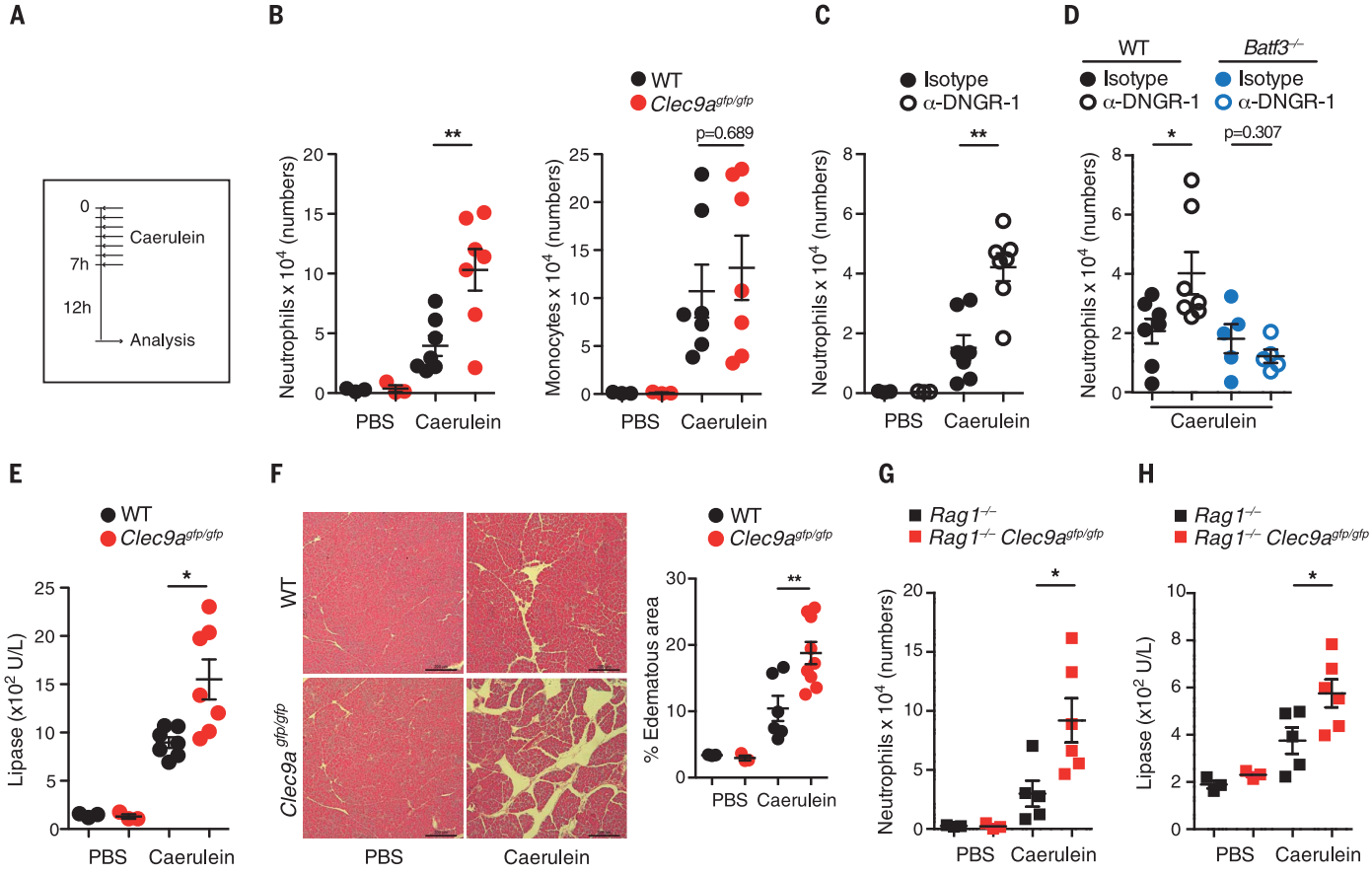


Fig. 1. DNGR-1 regulates neutrophil infiltration and tissue damage during acute pancreatitis. (**A**) Acute pancreatitis was induced by intraperitoneal injection of caerulein hourly for 6 hours in WT and DNGR-1–deficient (*Clec9a*$^{gfp/gfp}$) mice. Phosphate-buffered saline (PBS) injection was used as a control. Animals were analyzed 12 hours after the last injection. (**B**) Infiltrating neutrophils (left) and monocytes (right) in pancreas quantified by flow cytometry. (**C** and **D**) Anti–DNGR-1 (α-DNGR-1) or isotype control antibodies were intraperitoneally injected into WT (C) or WT and *Batf3*$^{-/-}$ (D) mice on the day before (day −1) and the day of (day 0) caerulein injection. Infiltrating neutrophils in pancreas were quantified by flow cytometry. (**E**) Concentrations of lipase were detected in serum from peripheral blood. U/L, units per liter. (**F**) Hematoxylin and eosin (H&E) staining in pancreatic sections (left). Representative images of $n = 5$ pancreata per experimental condition. Percentages of edematous area (right). (**G** and **H**) *Rag1*$^{-/-}$ and *Rag1*$^{-/-}$*Clec9a*$^{gfp/gfp}$ mice were subjected to pancreatitis as indicated in (A). Infiltrating neutrophils in pancreas quantified by flow cytometry (G) and serum lipase concentrations (H) are shown. In (B) to (H), each dot represents a single mouse. Data are means ± SEM of a representative experiment ($N \geq 2$ independent experiments). Significance was assessed by unpaired Student's *t* test between genotypes (B), (E), (F), (G), and (H) or treatments (C) and (D); $^*P < 0.05$, and $^{**}P < 0.01$.

Thus, neutrophil influx is the cellular mechanism driving the pathology in *Candida*-infected *Clec9a*$^{gfp/gfp}$ mice.

To decipher the mechanisms underlying the regulatory role of DNGR-1 on cDC1s in neutrophil infiltration, we used F-actin–myosin II complexes as DNGR-1 ligand (DNGR-1L) to robustly trigger the receptor (*20*). Plated DNGR-1L triggered signaling through the DNGR-1–SYK axis in B3Z-NFAT reporter cells (*7*) in a dose- (fig. S7A) and DNGR-1–dependent manner (fig. S7B). Then, we used a cDC1 cell line (MutuDC) (*21*) that expresses DNGR-1 as well as Dectin-1 (fig. S8A), a CLR critically involved in *C. albicans* recognition (*22*). Stimulation of MutuDCs with the Dectin-1 agonists whole β-glucan particles (WGP) or heat-killed *C. albicans* (*23*) induced the expression of proinflammatory factors such as *Tnf*, *Cxcl2*, and *Egr2*. This was reduced by concomitant exposure to DNGR-1L (Fig. 4A and fig. S8, B and C). Consistently, DNGR-1 triggering attenuated phospholipase Cγ2 (PLCγ2) phosphorylation and IκB degradation in response to WGP (Fig. 4B). DNGR-1 triggering had no impact on the response to toll-like receptor 9 (TLR9) ligand CpG (Fig. 4C and fig. S8, D and E), indicating specificity in the pathways modulated. Using a blocking antibody (fig. S8F), we confirmed that the effect elicited by DNGR-1L was DNGR-1 dependent (fig. S8G).

Regulatory phosphatases can couple to some immunoreceptor tyrosine-based activation motif (ITAM)–containing receptors (*24*, *25*). As DNGR-1 bears a hemi-ITAM (hemITAM) motif (*7*), we tested phosphatase activation upon DNGR-1L sensing. DNGR-1L induced SHP-1 phosphorylation (Fig. 4D) without affecting other CLR-related regulatory mechanisms (*26*, *27*) (fig. S8H). Treatment with the SHP inhibitor NSC-87877 (NSC) abolished the regulatory effect of DNGR-1L on responses elicited by WGP (Fig. 4E). Moreover, mice with SHP-1 depletion in the CD11c$^+$ compartment (CD11cΔSHP-1) (*28*), including cDC1s, phenocopied the exacerbated neutrophil infiltration observed in DNGR-1–deficient mice (Fig. 4F and fig. S9). These observations are consistent with an involvement of SHP-1 in the molecular mechanism that adjusts inflammatory responses in cDC1s after DNGR-1 engagement.

MIP-2 (encoded by *Cxcl2*) is a CXCR2 ligand fundamental for neutrophil mobilization from the BM (*29*) and local recruitment to *C. albicans*–infected tissues (*30*). We hypothesized that the MIP-2–CXCR2 axis could be mediating the boosted neutrophilia in the absence of DNGR-1. Administration of pepducin, a peptide that inhibits CXCR2 signaling (*13*), reverted the enhanced renal neutrophil recruitment observed in *Clec9a*$^{gfp/gfp}$ upon *Candida* infection (Fig. 4G). To dissect the contribution of DNGR-1 to the MIP-2–mediated process in vivo, we infected mice with *C. albicans*. After 60 hours, we measured *Cxcl2* expression in the renal immune infiltrate (fig. S10). Of all 10 immune populations tested, only neutrophil

frequencies were increased in *Clec9a*$^{gfp/gfp}$ mice (fig. S11). *Cxcl2* was expressed by neutrophils, macrophages, cDC2s, and cDC1s, but expression was enhanced only in cDC1s in *Clec9a*$^{gfp/gfp}$ mice (Fig. 4H). This suggests that DNGR-1 limits *Cxcl2* expression in cDC1s during *C. albicans* infection.

To investigate the relevance of this increased MIP-2 production by cDC1s on neutrophil recruitment under DNGR-1–deficient conditions, we generated mixed BM chimeric mice with specific MIP-2 deficiency in cDC1s (*Batf3*$^{-/-}$:*Cxcl2*$^{-/-}$) (see methods and fig. S12). After infection with *C. albicans*, DNGR-1 blockade generated an exacerbated renal neutrophil recruitment

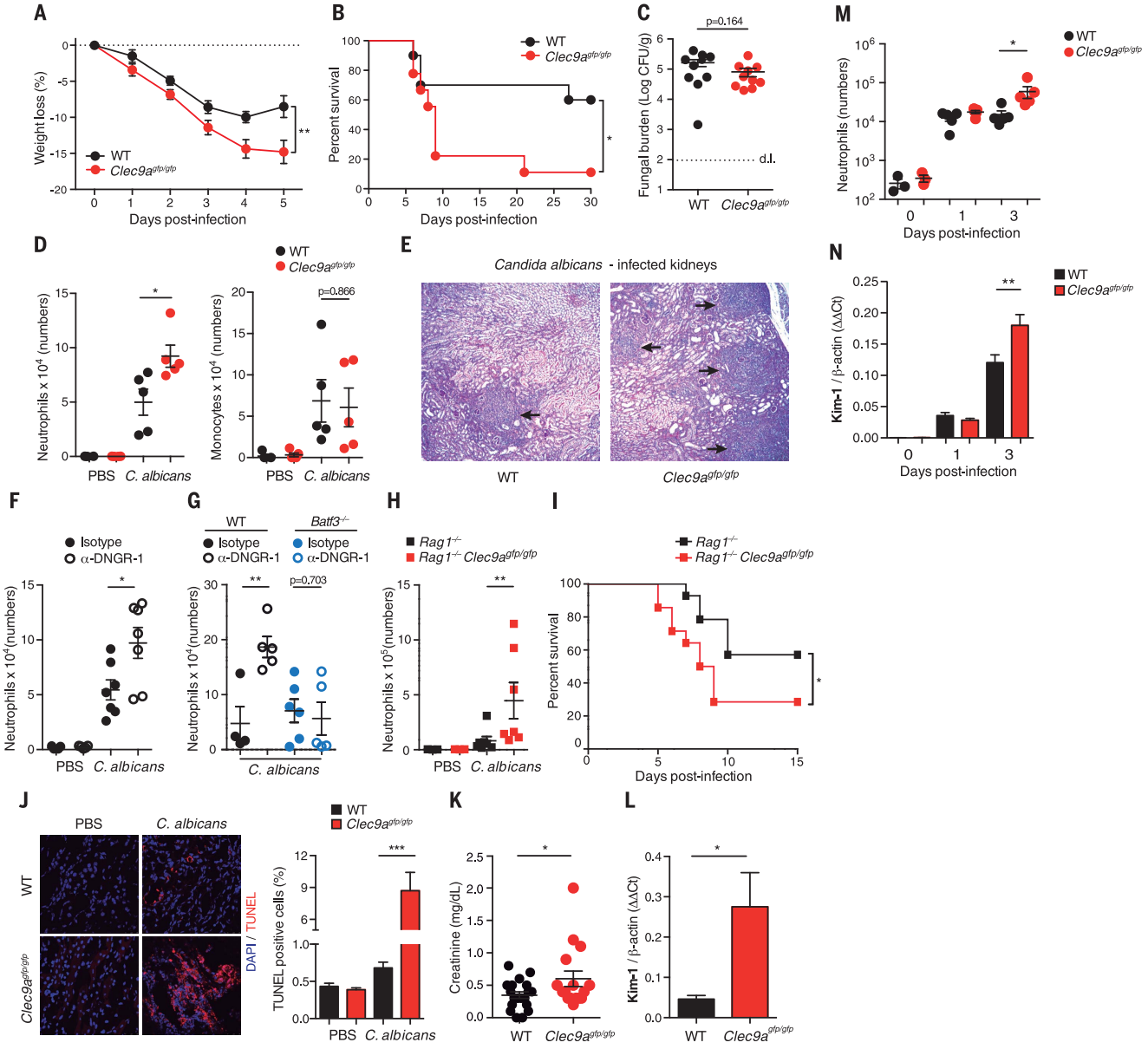


Fig. 2. DNGR-1 controls neutrophil recruitment and pathology associated with systemic candidiasis. *C. albicans* was intravenously injected into WT and *Clec9a*$^{gfp/gfp}$ mice. (**A** and **B**) Weight loss (A) and survival rate (B) were recorded. The dotted line in (A) shows the no-weight-loss reference. (**C** to **E**) After 6 days postinfection, renal fungal burden was detected (C), renal infiltrated neutrophils (left) and monocytes (right) were quantified by flow cytometry (D), and H&E staining was performed (E) (arrows indicate neutrophil accumulation). d.l., detection limit; CFU, colony forming unit. (**F** and **G**) Anti–DNGR-1 or isotype control antibodies were intraperitoneally injected in WT (F) or WT and *Batf3*$^{-/-}$ (G) mice on day −1 and daily after infection; infiltrating renal neutrophils were analyzed by flow cytometry. (**H** and **I**) *Rag1*$^{-/-}$ and *Rag1*$^{-/-}$*Clec9a*$^{gfp/gfp}$ mice were infected as indicated. Renal neutrophils (day 6 postinfection) were quantified by flow cytometry (H), and survival rate was determined (I). (**J**) 4′,6-diamidino-2-phenylindole (DAPI) and TUNEL staining in renal sections (left) and the percentage of TUNEL-positive cells (right). (**K** and **L**) Serum creatinine concentrations (K) and KIM-1 relative expression in total kidney (L) in WT and *Clec9a*$^{gfp/gfp}$ *Candida*-infected mice. β-actin expression was used for normalization. (**M** and **N**) Number of renal neutrophils (M) and KIM-1 expression in total kidney (N) at the indicated times postinfection. In (A), (C), (D), (F), (G), (H), and (M), data are means ± SEM of a representative experiment ($N \geq 2$ independent experiments), including at least five mice per condition. (B) and (I) show representative experiments ($N \geq 2$ independent experiments) with $n \geq 9$ mice per genotype. In (C), (D), (F), (G), (H), (K), and (M), each dot represents a single mouse. (E) and (J) show representative images of $n \geq 5$ kidneys per condition. In (K), (L), and (N), data are means ± SEM of ≥2 pooled experiments ($n \geq 5$ mice per experimental condition in each independent experiment). Significance was assessed by two-way analysis of variance (ANOVA) with Bonferroni post hoc test (A), log-rank (B) and (I), or unpaired Student's *t* test between genotypes (C), (D), (H), (J), (K), (L), (M), and (N) or treatments (F) and (G); $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

in *Batf3*$^{-/-}$:WT control chimeras (Fig. 4I). This boosted neutrophilia was lost in *Batf3*$^{-/-}$:*Cxcl2*$^{-/-}$ chimeras (Fig. 4I), thus relying on MIP-2 produced by cDC1s.

Furthermore, we crossed *Clec9a*$^{gfp/gfp}$ and *Cxcl2*$^{-/-}$ mice to further generate chimeric mice with cDC1s lacking both DNGR-1 and MIP-2 (*Batf3*$^{-/-}$:*Clec9a*$^{gfp/gfp}$ *Cxcl2*$^{-/-}$). Upon systemic candidiasis, *Batf3*$^{-/-}$:*Clec9a*$^{gfp/gfp}$ chimeras showed an exacerbated neutrophil infiltration into the kidney compared with *Batf3*$^{-/-}$:WT control chimeras (Fig. 4J). Notably, this boosted neutrophilia was lost in *Batf3*$^{-/-}$:*Clec9a*$^{gfp/gfp}$ *Cxcl2*$^{-/-}$ mice (Fig. 4J). Thus, in the absence of DNGR-1, MIP-2 produced by BATF3-dependent cDC1s is a key mediator for the enhanced neutrophil recruitment.

Infiltration of immune cells within injured tissues must balance pathogen control with increased damage caused by the inflammatory response. In particular, early infiltration by neutrophils to damaged tissues must be carefully regulated, because these cells can cause further tissue destruction (*13*, *17*, *18*). Disease tolerance to infections comprises mechanisms involved in the

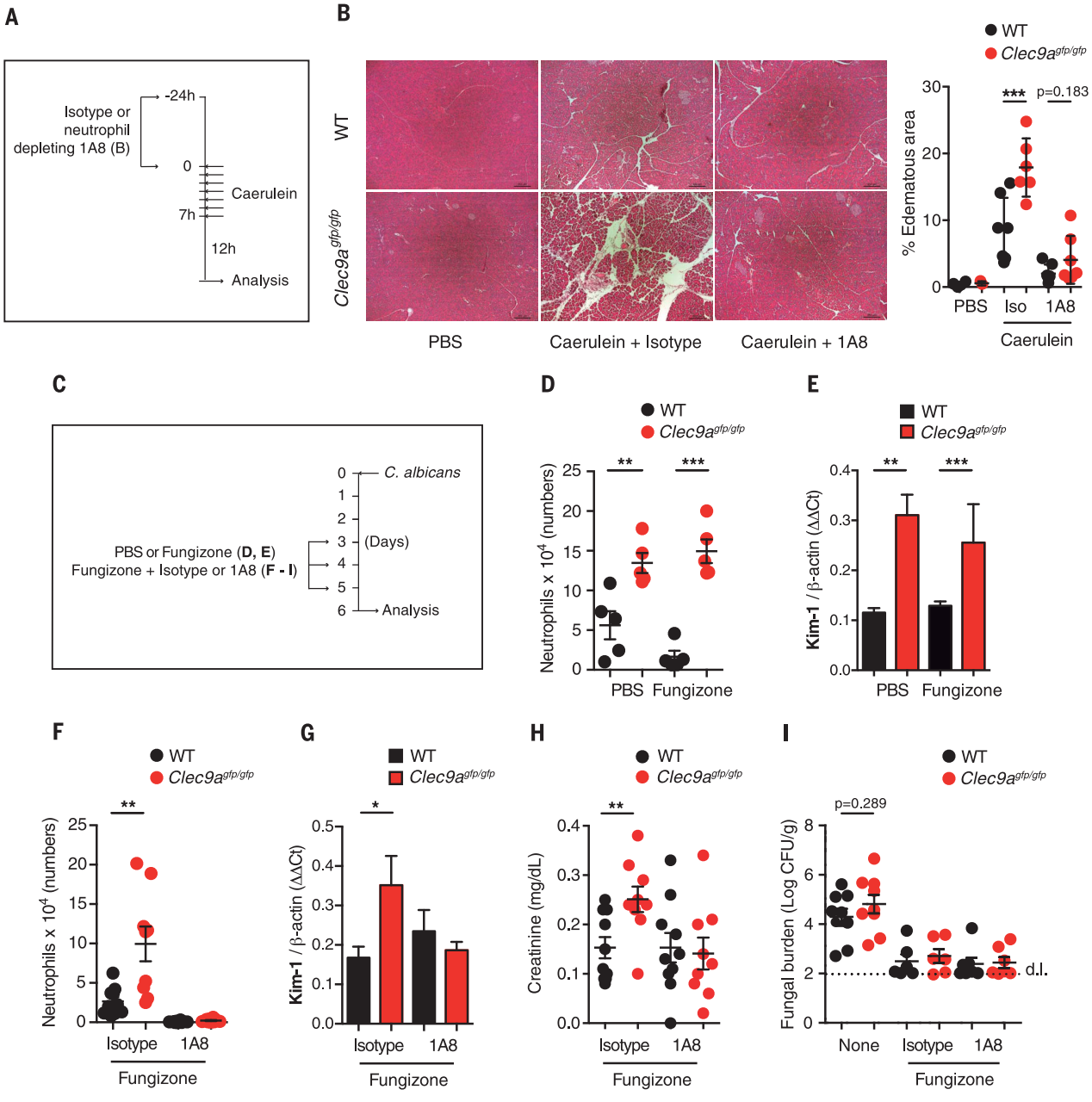


Fig. 3. DNGR-1 restrains tissue damage by dampening neutrophil-mediated immunopathology. (**A**) Representative scheme of pancreatitis induction. 1A8 neutrophil-depleting antibody or isotype (iso) control were administered intraperitoneally, as indicated. (**B**) H&E staining in pancreatic sections (left) and percentages of edematous area (right). Representative images of $n \geq 6$ pancreata per experimental condition. (**C**) Representative scheme of *C. albicans* infection; fungizone or PBS together or alone with 1A8 antibody or isotype control were intraperitoneally administered as indicated. (**D** to **I**) *C. albicans* was intravenously injected in WT and *Clec9a*$^{gfp/gfp}$ mice. After 6 days postinfection, renal-infiltrating neutrophils were quantified by flow cytometry (D) and (F), and KIM-1 expression in total kidney was measured (E) and (G). (H and I) Serum creatinine concentrations (H) and renal fungal burden (I). In (B) and (D) to (I), data are means ± SEM of a representative experiment ($N \geq 2$ independent experiments), including at least five mice per condition. Each dot represents a single mouse. Significance was assessed by unpaired Student's *t* test between genotypes; $^*P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

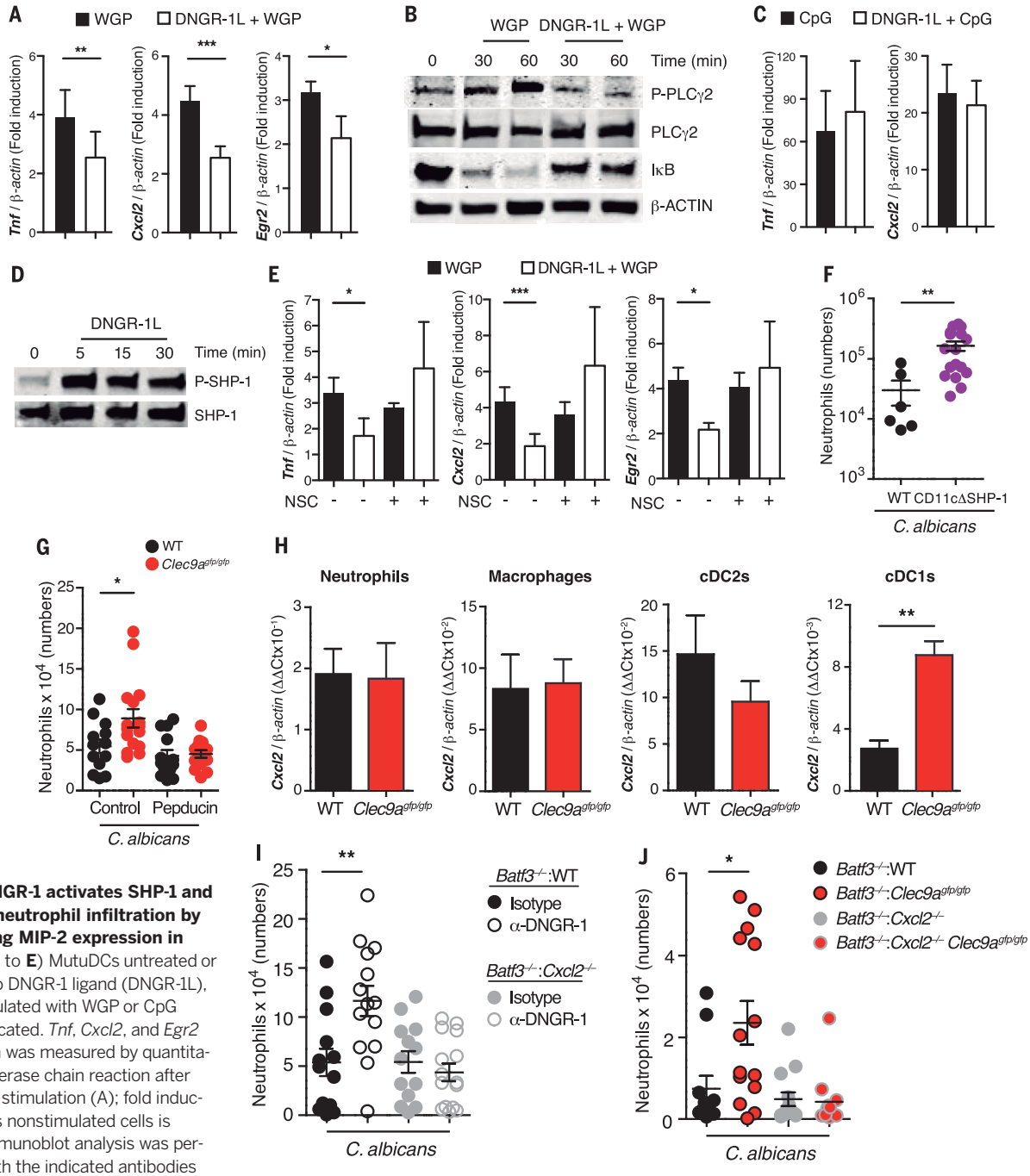


Fig. 4. DNGR-1 activates SHP-1 and controls neutrophil infiltration by dampening MIP-2 expression in cDC1s. (**A** to **E**) MutuDCs untreated or exposed to DNGR-1 ligand (DNGR-1L), were stimulated with WGP or CpG where indicated. *Tnf*, *Cxcl2*, and *Egr2* expression was measured by quantitative polymerase chain reaction after 4 hours of stimulation (A); fold induction versus nonstimulated cells is shown. Immunoblot analysis was performed with the indicated antibodies (B). P-PLCγ2, phosphorylated PLCγ2. In (C), *Tnf* and *Cxcl2* expression were measured as in (A). *Egr2* expression was not induced in response to CpG. MutuDCs were exposed to DNGR-1L and analyzed by immunoblot (D). P-SHP-1, phosphorylated SHP-1. In (E), *Tnf*, *Cxcl2*, and *Egr2* expression was measured in cells either preincubated with the SHP-inhibitor NSC (+) or not (−) and stimulated as in (A). (**F** to **J**) Mice were intravenously infected with *C. albicans*. Renal infiltrating neutrophils were quantified after 6 days in CD11cΔSHP-1 and WT littermates (F) and pepducin or control peptide-treated WT and *Clec9a*$^{gfp/gfp}$ mice (G). In (H), relative *Cxcl2* expression by immune cells in the kidney was measured 60 hours postinfection. In (I) and (J), renal infiltrating neutrophils were quantified at day 6 post infection. In (I), lethally irradiated B6/SJL CD45.1 recipient mice were reconstituted with a mixture of 50% BATF3-deficient BM cells (CD45.2, producing MIP-2 but lacking cDC1s) and 50% MIP-2–deficient BM cells (CD45.2, *Cxcl2*$^{-/-}$ mice). Thus, *Batf3*$^{-/-}$:*Cxcl2*$^{-/-}$ chimeric mice are defective for MIP-2 production only in cDC1s compared with control BM chimeras (*Batf3*$^{-/-}$:WT). Anti–DNGR-1 or isotype control antibodies were intraperitoneally injected on day −1 and daily after infection. In (J), neutrophil numbers are shown for the following mixed BM chimeric mice, generated as in (I): (i) *Batf3*$^{-/-}$:WT control chimeras; (ii) *Batf3*$^{-/-}$:*Clec9a*$^{gfp/gfp}$, which generate cDC1s lacking DNGR-1; (iii) *Batf3*$^{-/-}$:*Cxcl2*$^{-/-}$, which produce MIP-2–deficient cDC1s; and (iv) *Batf3*$^{-/-}$:*Clec9a*$^{gfp/gfp}$ *Cxcl2*$^{-/-}$, which generate cDC1s lacking both DNGR-1 and MIP-2. For (I) and (J), all cDC1s in the kidney were of donor origin (fig. S12, A and C), and the number of reconstituted cDC1s were equal in the different chimeric mice (fig. S12, B and D). In (A), (C), (E), and (H), data are means + SEM of pooled experiments, including $N \geq 3$ individual cultures or ≥4 mice per condition ($N \geq 4$ independent experiments). For (B) and (D), $N \geq 2$ representative immunoblots. In (F), (G), (I), and (J), data are means ± SEM of two pooled experiments. Each dot represents a single mouse. Significance was assessed by paired Student's *t* test between DNGR-1L–treated or untreated (A), (C), and (E) or between genotypes (H) or unpaired between genotypes (F), (G), (I), and (J); *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

control of tissue damage. This concept of "tissue damage control" is not restricted to infections and can also be applied to the regulation of damage from sterile inflammation (*14*, *31*). Notably, mediators involved in tissue damage control under both sterile and infectious conditions can be shared (*31*). Our data suggest that DNGR-1 acts as a shared checkpoint for sterile and infectious tissue damage control. Detection of tissue damage by cDC1s through DNGR-1 would act as a checkpoint for neutrophil infiltration and further immunopathology. Deficient sensing of tissue damage in the absence of DNGR-1 leads to higher production of MIP-2 by cDC1s. This increased MIP-2 production can ignite neutrophil infiltration that drives immunopathology within the damaged organ (fig. S13). Thus, DNGR-1 acts as a necrosis-sensing receptor that, depending on the inflammatory context, may promote a regulatory tissue damage–control mechanism by cDC1s or may contribute to cross-priming during adaptive immunity–related responses (*7*–*9*). This capacity to develop two different host protective functions and the regulation and implications of this dual role remain to be investigated.

REFERENCES AND NOTES

1. P. Matzinger, *Science* **296**, 301–305 (2002).
2. K. L. Rock, H. Kono, *Annu. Rev. Pathol.* **3**, 99–126 (2008).
3. S. Ahrens *et al.*, *Immunity* **36**, 635–645 (2012).
4. J.-G. Zhang *et al.*, *Immunity* **36**, 646–657 (2012).
5. B. U. Schraml *et al.*, *Cell* **154**, 843–858 (2013).
6. L. F. Poulin *et al.*, *Blood* **119**, 6052–6062 (2012).
7. D. Sancho *et al.*, *Nature* **458**, 899–903 (2009).
8. S. Iborra *et al.*, *J. Clin. Invest.* **122**, 1628–1643 (2012).
9. S. Zelenay *et al.*, *J. Clin. Invest.* **122**, 1615–1627 (2012).
10. M. M. Lerch, F. S. Gorelick, *Gastroenterology* **144**, 1180–1193 (2013).
11. S. Iborra *et al.*, *Immunity* **45**, 847–860 (2016).
12. K. Hildner *et al.*, *Science* **322**, 1097–1100 (2008).
13. C. W. Steele *et al.*, *J. Pathol.* **237**, 85–97 (2015).
14. R. Medzhitov, D. S. Schneider, M. P. Soares, *Science* **335**, 936–941 (2012).
15. A. M. Jamieson *et al.*, *Science* **340**, 1230–1234 (2013).
16. J. F. Fisher, K. Kavanagh, J. D. Sobel, C. A. Kauffman, C. A. Newman, *Clin. Infect. Dis.* **52** (suppl. 6), S437–S451 (2011).
17. O. Majer *et al.*, *PLOS Pathog.* **8**, e1002811 (2012).
18. M. S. Lionakis *et al.*, *PLOS Pathog.* **8**, e1002865 (2012).
19. T. Dejima *et al.*, *Infect. Immun.* **79**, 4503–4510 (2011).
20. O. Schulz *et al.*, *Cell Reports* **24**, 419–428 (2018).
21. S. A. Fuertes Marraco *et al.*, *Front. Immunol.* **3**, 331 (2012).
22. R. A. Drummond, G. D. Brown, *Curr. Opin. Microbiol.* **14**, 392–399 (2011).
23. H. S. Goodridge *et al.*, *Nature* **472**, 471–475 (2011).
24. S. Ben Mkaddem *et al.*, *J. Clin. Invest.* **124**, 3945–3959 (2014).
25. S. Iborra *et al.*, *Immunity* **45**, 788–801 (2016).
26. N. Blanco-Menéndez *et al.*, *J. Immunol.* **195**, 4466–4478 (2015).
27. T. S. Freedman *et al.*, *eLife* **4**, e09183 (2015).
28. C. L. Abram, G. L. Roberge, L. I. Pao, B. G. Neel, C. A. Lowell, *Immunity* **38**, 489–501 (2013).
29. K. J. Eash, A. M. Greenbaum, P. K. Gopalan, D. C. Link, *J. Clin. Invest.* **120**, 2423–2431 (2010).
30. E. Balish *et al.*, *J. Leukoc. Biol.* **66**, 144–150 (1999).
31. S. R. Wu, P. Reddy, *Trends Immunol.* **38**, 231–235 (2017).

ACKNOWLEDGMENTS

We are grateful to C. Reis e Sousa and C. A. Lowell for sharing essential reagents. We thank C. Reis e Sousa and members of the D.S. laboratory for discussions and critical reading of the manuscript. We appreciate R. Mota for his advice on the pancreatitis model. We thank the staff at the CNIC facilities for technical support. **Funding:** C.d.F. is supported by the AECC Foundation as the recipient of an "Ayuda Fundación Científica AECC a personal investigador en cancer." P.S.-L. is funded by grant BES-2015-072699 from the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU). M.E. is the recipient of a CNIC International Ph.D. Programme fellowship "la Caixa"-Severo Ochoa OSLC-CNIC-2013-04. S.K.W. is supported by a European Molecular Biology Organization (EMBO) Long-Term Fellowship (grant ALTF 438-2016) and a CNIC-International Postdoctoral Programme Fellowship (grant 17230-2016). Work in the D.S. laboratory is funded by the CNIC and grant SAF2016-79040-R from MCIU, Agencia Estatal de Investigación, and Fondos Europeos de Desarrollo Regional (FEDER); B2017/BMD-3733 Immunothercan-CM from Comunidad de Madrid; RD16/0015/0018-REEM from FIS-Instituto de Salud Carlos III, MICINN, and FEDER; Acteria Foundation; Constantes y Vitales prize (Atresmedia); La Marató de TV3 Foundation (201723); the European Commission (635122-PROCROP H2020); and the European Research Council (ERC-2016-Consolidator Grant 725091). The CNIC is supported by the MCIU and the Pro-CNIC Foundation and is a Severo Ochoa Center of Excellence (SEV-2015-0505). **Author contributions:** Conceptualization: C.d.F., P.S.-L., and D.S.; methodology, investigation, analysis, and validation: C.d.F., P.S.-L., M.E., S.K.W., S.M.-C., N.B.-M., O.S., M.G., and F.M.-M.; resources: O.S., E.C., and A.P.; writing of original draft: C.d.F., P.S.-L., and D.S.; editing of draft: all authors; supervision, project administration, and funding acquisition: D.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to understand and assess the conclusions of this research are available in the main text and supplementary materials. *Batf3*$^{-/-}$ mice were obtained from K. M. Murphy under a material transfer agreement with Washington University and the Howard Hugues Medical Institute. The MuTu1940 DC cell line was obtained from H. Acha-Orbea under a material transfer agreement with the University of Laussanne.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/351/suppl/DC1
Materials and Methods
Figs. S1 to S13
References

29 May 2017; resubmitted 13 June 2018
Accepted 3 September 2018
10.1126/science.aan8423

GERM CELL DEVELOPMENT

Generation of human oogonia from induced pluripotent stem cells in vitro

Chika Yamashiro[1,2], Kotaro Sasaki[1,2], Yukihiro Yabuta[1,2], Yoji Kojima[1,2,3,4], Tomonori Nakamura[1,2], Ikuhiro Okamoto[1,2], Shihori Yokobayashi[1,2,4], Yusuke Murase[1,2], Yukiko Ishikura[1,2], Kenjiro Shirane[5,6], Hiroyuki Sasaki[5,6], Takuya Yamamoto[3,4,7], Mitinori Saitou[1,2,3,4]*

Human in vitro gametogenesis may transform reproductive medicine. Human pluripotent stem cells (hPSCs) have been induced into primordial germ cell–like cells (hPGCLCs); however, further differentiation to a mature germ cell has not been achieved. Here, we show that hPGCLCs differentiate progressively into oogonia-like cells during a long-term in vitro culture (approximately 4 months) in xenogeneic reconstituted ovaries with mouse embryonic ovarian somatic cells. The hPGCLC-derived oogonia display hallmarks of epigenetic reprogramming—genome-wide DNA demethylation, imprint erasure, and extinguishment of aberrant DNA methylation in hPSCs—and acquire an immediate precursory state for meiotic recombination. Furthermore, the inactive X chromosome shows a progressive demethylation and reactivation, albeit partially. These findings establish the germline competence of hPSCs and provide a critical step toward human in vitro gametogenesis.

The germ-cell lineage arises from primordial germ cells (PGCs) that go through a multistep process to generate spermatozoa or oocytes. Methods for in vitro gametogenesis from pluripotent stem cells (PSCs) would provide a powerful tool with which to explore the mechanism of germ-cell development and its anomalies (*1*). Mouse PSCs have been induced into PGC-like cells (mPGCLCs), which contribute to spermatogenesis upon transplantation into testes (*2*, *3*) and to oogenesis upon aggregation with embryonic ovarian somatic cells [reconstituted ovaries (rOvaries)], followed by transplantation under ovarian bursa (*4*) or an appropriate culture (*5*). Resultant gametes from these procedures generate fertile offspring (*2*–*5*). Furthermore, human PSCs (hPSCs) have been induced into hPGCLCs that bear a gene-expression property of hPGCs just after their specification, opening the possibility for human in vitro gametogenesis (*6*, *7*). However, further differentiation of hPGCLCs has not been successful, and whether hPGCLCs can develop as mature germ cells remains unknown.

[1]Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [2]Japan Science and Technology Agency (JST), Exploratory Research for Advanced Technology (ERATO), Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [3]Institute for Integrated Cell-Material Sciences, Kyoto University, Yoshida-Ushinomiya-cho, Sakyo-ku, Kyoto 606-8501, Japan. [4]Center for iPS Cell Research and Application, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. [5]Division of Epigenomics, Medical Institute of Bioregulation, and Epigenome Network Research Center, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan. [6]Graduate School of Medical Sciences, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan. [7]Agency for Medical Research and Development–Core Research for Evolutional Science and Technology (AMED-CREST), AMED, 1-7-1 Otemachi, Chiyoda-ku, Tokyo, 100-0004, Japan.
*Corresponding author. Email: saitou@anat2.med.kyoto-u.ac.jp

At week 2 of development, hPGCs express key transcription factors (TFs) such as SOX17, TFAP2C, and BLIMP1 (also known as PRDM1). Then, at week 5, they migrate to and colonize the embryonic gonads to initiate differentiation into oogonia or gonocytes in embryonic ovaries or testes (*8*, *9*) that express RNA regulators such as DAZL and DDX4 [also known as human VASA homolog (hVH)] (*10*–*12*). Oogonia and gonocytes are very similar in morphology, gene expression, and epigenetic properties until they start sexual differentiation at week 10 into oocytes through meiotic prophase or fetal spermatogonia (*8*–*12*). A characteristic event in germ-cell development is epigenetic reprogramming, which occurs by week 10 and leads to genome-wide DNA demethylation, imprint erasure, and X reactivation (*10*–*12*).

We explored whether hPGCLCs can undergo further development in vitro in xenogeneic reconstituted ovaries (xrOvaries) with mouse embryonic ovarian somatic cells (fig. S1A). First, we induced male human-induced PSCs (hiPSCs) with the *BLIMP1-tdTomato*;*TFAP2C-EGFP* alleles [585B1 BTAG (XY)] into incipient mesoderm-like cells (iMeLCs) and then into hPGCLCs (*7*, *13*). BLIMP1 and TFAP2C are expressed in oogonia and gonocytes at least until week 10 (*10*–*12*). We isolated BTAG-positive (BT^+AG^+) hPGCLCs at day 6 of induction by means of fluorescence-activated cell sorting (FACS) and generated xrOvaries. In agreement with a previous report (*5*), mPGCLCs differentiated efficiently into primary oocytes and formed secondary follicles after a 21-day culture in rOvaries (fig. S1B). At culture day 7, the xrOvaries exhibited a round and flattened shape, and the BT^+AG^+ cells were distributed uniformly within the xrOvaries (fig. S1C). Subsequently, the xrOvaries expanded laterally with the formation of cyst-like structures (Fig. 1A). From culture days 21 to 77, the xrOvaries exhibited autofluorescence under fluorescence microscopy (fig. S1D). There were ~2000 BT^+AG^+ cells per xrOvary at culture day 21 and ~500 at culture day 77 (fig. S1E), indicating that only a fraction of the initial hPGCLCs (~5000) survived in xrOvaries.

At culture day 77, the AG^+ cells in xrOvaries existed as clusters, were positive for a human mitochondrial antigen, bore faint 4′,6-diamidino-2-phenylindole (DAPI) staining, and were delineated by $FOXL2^+$ mouse granulosa cells and their basement membrane (fig. S2, A to C). The intensity of the DAPI staining in AG^+ cells appeared to decline progressively during the culture (fig. S2D). The AG^+ cells expressed key TFs for early germ cells (TFAP2C, SOX17, and POU5F1), and some were mitotically active ($Ki67^+$) or in apoptosis (cleaved $CASPASE3^+$) (fig. S2A). At culture day 77, many AG^+ cells up-regulated DAZL and DDX4 (Fig. 1B and fig. S2E), suggesting that in xrOvaries, hPGCLCs not only survive as germ cells but also differentiate into oogonia and gonocytes. Accordingly, at culture day 77, electron microscopy revealed the presence of large cells, highly similar to oogonia and gonocytes, with clear cytoplasms with sparsely located mitochondria and round nuclei with loosely packed chromatin and prominent nucleoli (fig. S3) (*8*, *9*).

We next generated female hiPSCs that bear the AG;*DDX4/hVH-tdTomato* alleles [1390G3 AGVT (XX)] (fig. S4) (*13*), created xrOvaries, and cultured them up to culture day 120. The female AG^+ hPGCLCs in xrOvaries developed similarly to the male hPGCLCs (fig. S5) and up-regulated VT at culture day 77, and the AG^+VT^+ cells appeared to differentiate into AG-negative (AG^-) VT^+ cells at culture day 120 (Fig. 1C and fig. S5A). We examined the expression of key genes by means of quantitative polymerase chain reaction. Both male and female hPGCLC-derived cells expressed early germ-cell genes (*BLIMP1*, *TFAP2C*, *SOX17*, and *NANOS3*), core or naïve pluripotency genes (*POU5F1*, *NANOG*, *TCL1B*, and *TFCP2L1*), and up-regulated genes for oogonia and gonocytes (*DAZL* and *DDX4*) from around culture days 35 to 49 onward, and also genes for meiosis (*SYCP3* and *REC8*) (fig. S6). The AG^-VT^+ cells at culture day 120 down-regulated early germ-cell and core or naïve pluripotency genes and up-regulated *STRA8*, a gene essential for meiosis initiation (fig. S6), suggesting their developmentally advanced character. Consistently, the $DDX4^+$ cells at culture day 120 expressed SYCP3 (SCP3) but not γH2AX, DMC1, and SYCP1 (SCP1) proteins, indicating that they have not yet initiated meiotic recombination (fig. S7).

We analyzed the transcriptomes of these cell types (fig. S8A and table S1). Unsupervised hierarchical clustering (UHC) showed that hPGCLC-derived cells are distinct from hiPSCs and iMeLCs and can be subclassified according to their culture period (fig. S8B). Principle component analysis (PCA) gave a concordant result (fig. S8C). We identified the genes with significantly positive or negative PC1/2 loadings (453 genes) (fig. S8D and table S2). UHC classified them into five major clusters (Fig. 2A, fig. S8D, and table S2):

Cluster 1 represents the genes up-regulated upon hPGCLC specification or early in xrOvaries and expressed essentially continuously thereafter. Clusters 2 and 5 signify the genes transiently expressed in early hPGCLC-derived cells or down-regulated early in xrOvaries, and cluster 4 represents the genes down-regulated upon hPGCLC specification (Fig. 2A and table S2). Cluster 3, which shows a progressive and coordinated up-regulation from culture day 35 onwards, signifies genes for oogonia and gonocytes and is enriched in genes with gene ontology (GO) functional terms for male meiosis, fertilization, and Piwi-interacting RNA metabolic process (Fig. 2A and table S2). Using the 453 genes, we compared the gene-expression properties of hPGCLC-derived cells with those of oogonia and gonocytes (*12*). hPGCLC-derived cells from culture day 35 onward, particularly the culture day 77 BT^+AG^+ and culture day 120 AG^+ cells, exhibited a strong similarity to week 7 and 9 oogonia and gonocytes (Fig. 2B and fig. S10).

The culture day 120 AG^-VT^+ cells exhibited a developmentally advanced character (figs. S6, S7, and S8, B and C). To clarify this point further, we examined the expression of genes that distinguish relevant human fetal germ cells (FGCs) [mitotic (oogonia and gonocytes), retinoic acid (RA)–responsive (female), meiotic (female), and oogenesis (female)] (*14*) in culture day 120 AG^-VT^+ cells, which intriguingly revealed their similarity to RA-responsive FGCs. They down-regulate genes for early germ cells; further up-regulate *DAZL*, *DDX4*, *MAEL*, and *KRBOX1*; up-regulate RA- or bone morphogenetic protein (BMP)–responsive genes (*STRA8*, *REC8*, or *ID1/2/3/4*, *MSX1/2*); yet do not sufficiently up-regulate key meiosis genes (*SYCP1*, *DMC1*, *SPO11*, or *PRDM9*) (Fig. 2C and fig. S11). Thus, hPGCLC development in xrOvaries reconstitutes human germ-cell development, albeit with protracted kinetics, leading to the generation of oogonia and RA-responsive FGCs, a state responding to the signal for the meiotic entry and in preparation for the meiotic recombination (*14*).

We determined the genome-wide DNA methylation [5-methylcytosine (5mC)] profiles of hPGCLC-derived cells in xrOvaries by means of whole-genome bisulfite sequencing (WGBS) (fig. S12A and table S3). The genome-wide 5mC levels of hiPSCs and iMeLCs were both ~80% and decreased progressively in hPGCLC-derived cells, reaching ~20% in culture day 77 cells and ~13% in culture day 120 cells (fig. S12B), the levels comparable with that in oogonia and gonocytes at weeks 7 to 10 (*11*, *12*). The demethylation occurred throughout the genome (Fig. 3 and figs. S12C and S13, A and B), and the 5mC distribution profiles of the culture days 77 to 120 cells were very similar to those of oogonia and gonocytes at weeks 7 to 10 (Fig. 3), respectively, but not to those of the blastocysts (*15*) and naïve human embryonic stem cells (hESCs) (fig. S12D) (*16*, *17*) [and hPGCLCs reported by others (*18*)]. Thus, hPGCLC-derived cells demethylate their 5mCs in a fashion similar to that of oogonia and gonocytes but not early embryonic cells and their putative in vitro counterparts.

We examined whether hPGCLC-derived cells can erase the parental imprints. Blastocysts (*15*) and somatic cells (*11*) exhibited ~50% CpG methylation in the differentially methylated regions (DMRs) of paternally and maternally imprinted genes (Fig. 4A and table S4). By contrast, hiPSCs and primed hESCs (*12*, *16*, *17*) exhibited hypermethylation in some DMRs, whereas naïve hESCs (*16*, *17*) showed hypomethylation in nearly all the DMRs (Fig. 4A), indicating that hPSCs misregulate the imprint states. Similar to oogonia and gonocytes, hPGCLC-derived cells progressively erased the parental imprints, including the hypermethylated DMRs of hiPSCs (Fig. 4A). We determined the CpG sequences that bear a hypermethylation in hiPSCs compared with three independent hESC lines (table S5). The hPGCLC-derived cells erased such hiPSC-specific methylation in a nearly complete fashion (fig. S13C). Repeat elements are also demethylated in hPGCLC-derived cells, and the demethylation-resistant repeats in vivo, such as ERVK and SVA (*11*, *12*), showed similar resistance in hPGCLC-derived cells (fig. S13D). We defined the "escapees" that retained relatively high 5mC levels in culture day 77 to 120 cells and the week 7 oogonia and gonocytes (table S6). Essentially, all the escapees in the oogonia and gonocytes (*12*) were included in those in the hPGCLC-derived cells, which were enriched around SVA, ERVK, and ERV1 (fig. S13E).

We explored whether hPGCLC-derived cells can reactivate the inactive X chromosome (Xi). RNA fluorescence in situ hybridization (RNA FISH) revealed that the 1390G3 AGVT hiPSCs (passage 29) lack the expression of *XIST* from both alleles and show monoallelic expression of the X-linked genes (fig. S14, A to D), indicating that they bear one active X chromosome (Xa) and one Xi without *XIST* (Xi^{XIST-}) (*19*). Accordingly, the promoters of the X-linked genes in 1390G3 AGVT hiPSCs exhibited an intermediate

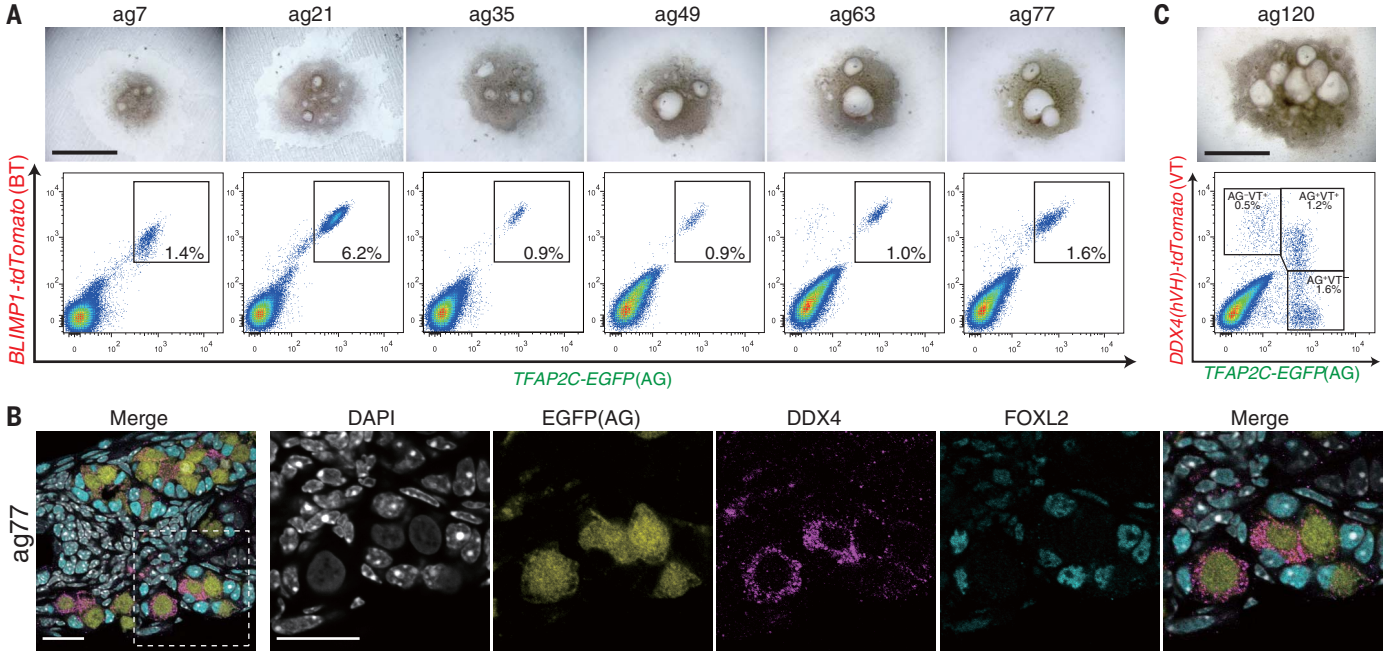


Fig. 1. hPGCLC differentiation in xrOvaries. (**A**) Bright-field (BF) images and FACS by BTAG of xrOvaries with 585B1 BTAG (XY) hPGCLC-derived cells from culture days (ag) 7 to 77. Scale bars, (A) and (C), 500 μm. (**B**) Expression of DDX4 (magenta) in AG^+ cells (yellow) in xrOvaries at culture day 77, with FOXL2 (cyan) and DAPI (white) staining. The boxed area at left (merged image) is magnified at right. Scale bars, 20 μm. (**C**) A BF image and FACS by AGVT of xrOvaries at culture day 120 with 1390G3 AGVT (XX) hPGCLC-derived cells.

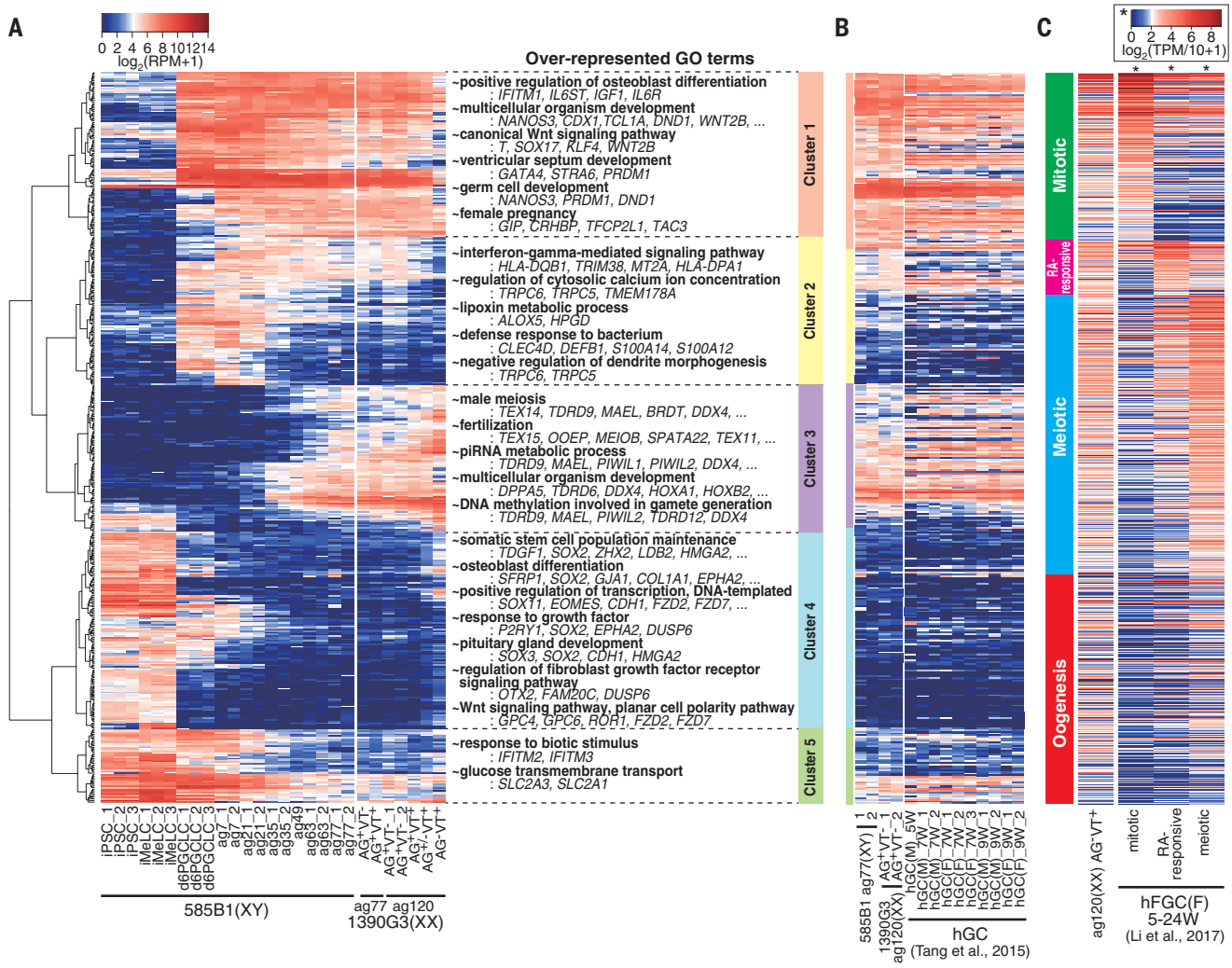


Fig. 2. Transcriptome dynamics of hPGCLC-derived cells in xrOvaries. (**A**) (Left) UHC and heatmap of the expression in the indicated cells of the 453 signature genes for the transitions of hPGCLC-derived cell properties in xrOvaries (fig. S8, C and D, and table S2). (Right) Representative genes in each cluster and their GO enrichments (table S2). (**B** and **C**) Comparison of the expression of (B) the 453 or (C) indicated genes (*14*) between (B) culture day 77 BT^+AG^+ and culture day 120 AG^+VT^+ cells or (C) culture day 120 AG^-VT^+ cells and the indicated human germ cells (*12*, *14*).

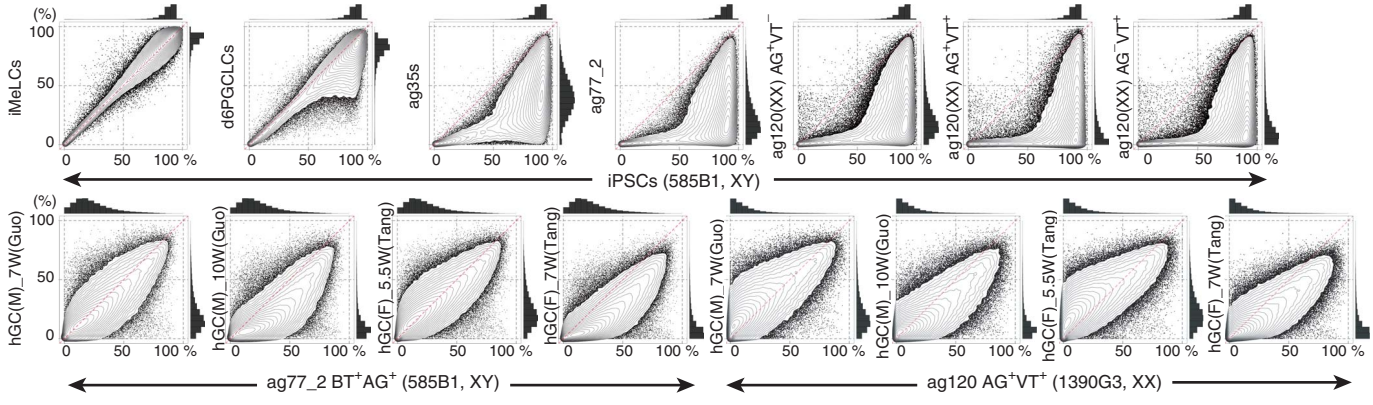


Fig. 3. Genome-wide DNA demethylation in hPGCLC-derived cells in xrOvaries. Comparisons of the 5mC levels (genome-wide 2-kb windows) by contour representations of scatter plots, combined with histogram representations (top and right of scatter plots), between hiPSCs (top) or ag77 BT^+AG^+ cells (bottom left) and ag120 AG^+VT^+ cells (bottom right) and the indicated cell types.

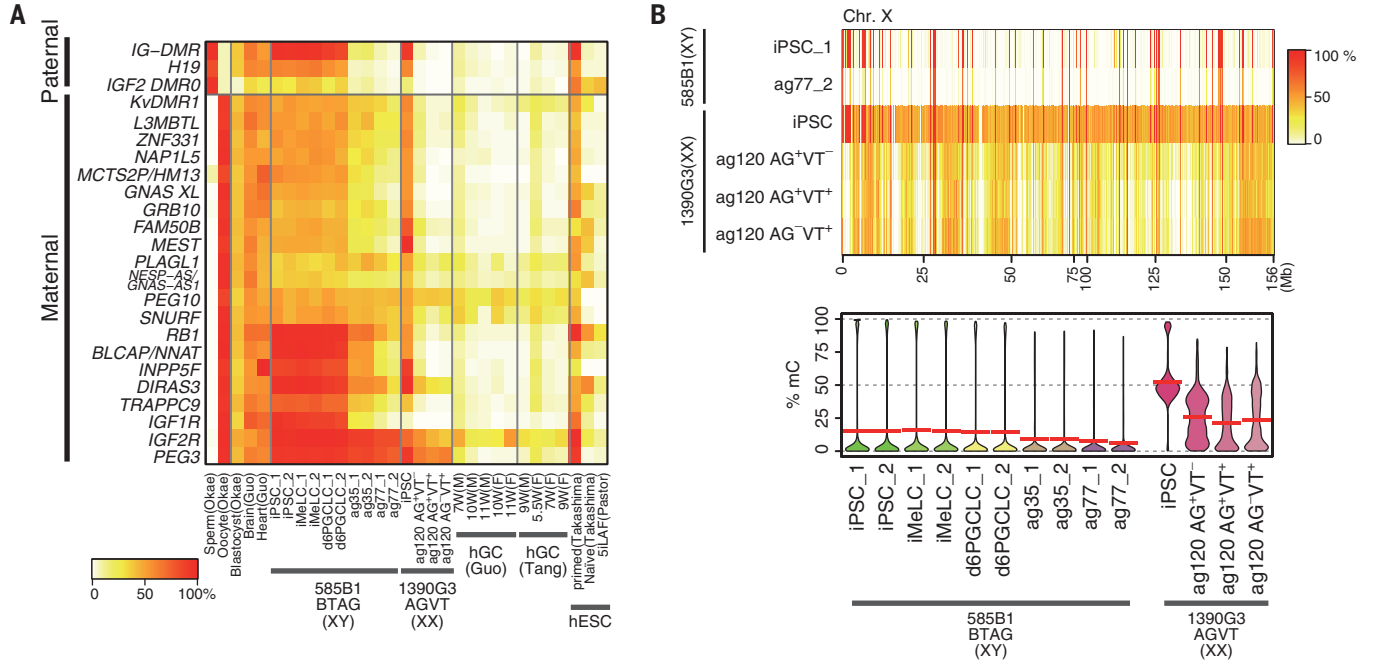


Fig. 4. Imprint erasure and X chromosome demethylation in hPGCLC-derived cells in xrOvaries. (**A**) Heatmaps showing the 5mC levels in the DMRs of the indicated imprinted genes in the indicated cells. (**B**) Heatmaps (top) and violin plots (bottom, average; red bars) showing the 5mC levels in the CpG islands on the X chromosomes in the indicated cells.

(~50%) 5mC level (Fig. 4B and figs. S14E and S15), suggesting that they are unmethylated on the Xa and nearly fully methylated on the Xi (*19*). The culture day 120 cells, but not day 6 hPGCLCs, exhibited a partial (~20%) reactivation—biallelic expression—of several X-linked genes but did not reactivate *XIST* (fig. S14, A to D). Consistently, the promoters of the X-linked genes in culture day 120 cells were moderately demethylated (~20%) (Fig. 4B and figs. S14E and S15). These findings demonstrate that hPGCLC-derived cells in xrOvaries undergo proper epigenetic reprogramming—genome-wide DNA demethylation, imprint erasure, and extinguishment of aberrantly acquired/persisting methylation of hiPSCs—reaching minimum 5mC levels (~13%) comparable with those reported in human germ cells. Nonetheless, the Xi^{XIST-} state in vitro is more resistant to reprogramming, indicating a distinctive epigenetic mechanism for the Xi in hPSCs, which warrants further investigation.

We have provided evidence that primed hiPSCs (*20*) are competent to generate germ cells with their hallmark of epigenetic reprogramming. In mouse-mouse rOvaries, ovarian somatic cells, likely granulosa cells, provide timely signals and environments for mPGCLCs to mature into primary oocytes and form secondary follicles (*5*). By contrast, in xrOvaries, mouse granulosa cells create a permissive environment for hPGCLCs to gradually mature into oogonia. Although the underlying mechanism remains unclear, because mPGCLCs undergo epigenetic reprogramming upon expansion and differentiate into oocytes in response to BMP and RA (*21*, *22*), hPGCLC-derived cell proliferation and signals from mouse granulosa cells (fig. S16) would allow hPGCLC-derived cells to mature into oogonia or RA-responsive FGCs (fig. S17). Because both male and female mPGCLCs enter into meiotic prophase under the same condition (*22*), male hPGCLCs would also enter into meiotic prophase in xrOvaries. Future studies will include exploring a strategy and the mechanism for the differentiation of hiPSC-induced oogonia into oocytes with meiotic recombination or for the differentiation of hPGCLCs into fetal spermatogonia, which promote human in vitro gametogenesis.

REFERENCES AND NOTES

1. M. Saitou, H. Miyauchi, *Cell Stem Cell* **18**, 721–735 (2016).
2. K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, M. Saitou, *Cell* **146**, 519–532 (2011).
3. Y. Ishikura *et al.*, *Cell Reports* **17**, 2789–2804 (2016).
4. K. Hayashi *et al.*, *Science* **338**, 971–975 (2012).
5. O. Hikabe *et al.*, *Nature* **539**, 299–303 (2016).
6. N. Irie *et al.*, *Cell* **160**, 253–268 (2015).
7. K. Sasaki *et al.*, *Cell Stem Cell* **17**, 178–194 (2015).
8. T. G. Baker, *Proc. R. Soc. London B Biol. Sci.* **158**, 417–433 (1963).
9. T. Fukuda, C. Hedinger, P. Groscurth, *Cell Tissue Res.* **161**, 55–70 (1975).
10. S. Gkountela *et al.*, *Cell* **161**, 1425–1436 (2015).
11. F. Guo *et al.*, *Cell* **161**, 1437–1452 (2015).
12. W. W. Tang *et al.*, *Cell* **161**, 1453–1467 (2015).
13. S. Yokobayashi *et al.*, *Biol. Reprod.* **96**, 1154–1166 (2017).
14. L. Li *et al.*, *Cell Stem Cell* **20**, 858–873.e4 (2017).
15. H. Okae *et al.*, *PLOS Genet.* **10**, e1004868 (2014).
16. Y. Takashima *et al.*, *Cell* **158**, 1254–1269 (2014).
17. W. A. Pastor *et al.*, *Cell Stem Cell* **18**, 323–329 (2016).
18. F. von Meyenn *et al.*, *Dev. Cell* **39**, 104–115 (2016).
19. S. Patel *et al.*, *Cell Reports* **18**, 54–67 (2017).
20. T. Nakamura *et al.*, *Nature* **537**, 57–62 (2016).
21. H. Ohta *et al.*, *EMBO J.* **36**, 1888–1907 (2017).
22. H. Miyauchi *et al.*, *EMBO J.* **36**, 3100–3119 (2017).

ACKNOWLEDGMENTS

We thank the members of our laboratory for their helpful input on this study: K. Hayashi for his advice on rOvaries; Y. Nagai, Y. Sakaguchi, and M. Kawasaki of the Saitou Laboratory, J. Oishi of the Sasaki Laboratory, and T. Sato and M. Kabata of the Yamamoto Laboratory for their technical assistance; and the Center for Anatomical, Pathological, and Forensic Medical Research at Kyoto University for the preparation of electron microscopy. We also thank T. Mori for his encouragement and support. **Author contributions:** C.Y. performed hPGCLC induction and xrOvary formation and analysis. K.S., Y.K., and S.Y. assisted in hPGCLC induction, and Y.I. assisted in xrOvary formation. T.N. and T.Y. contributed to the RNA sequencing (RNA-seq) and WGBS; Y.M., K.S., and H.S. contributed to the WGBS; and Y.Y. contributed to the analyses of RNA-seq and WGBS data. I.O. performed the RNA FISH. Y.K. and M.S. designed the experiments and wrote the manuscript. **Funding:** This work was supported by a Grant-in-Aid for Specially Promoted Research from the Japan Society for the Promotion of Science (JSPS) (17H06098) to M.S., by a JST-ERATO grant (JPMJER1104) to M.S., by funds from the HEALIOS to K.K. and the Pythias Fund to M.S., and by a Grant-in-Aid for Scientific Research on Innovative Areas and Specially Promoted Research from JSPS to H.S. (25112010 and 18H05214). **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The accession numbers for the RNA-seq and WGBS data generated in this study are GSE117101 (GEO) and DRA006618/DRA007077 (DDBJ), respectively.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6412/356/suppl/DC1
Materials and Methods
Figs. S1 to S17
References (*23*–*34*)
Tables S1 to S6

1 February 2018; accepted 10 September 2018
Published online 20 September 2018
10.1126/science.aat1674

sartorius

Take Your Microwell Plate Pipetting Back to the Comfort Zone

Looking for ways to speed up Microwell plate pipetting without errors?

Experience the Picus® with the unique Plate Tracker feature that keeps track of where you need to pipette next. In combination with the multi-dispensing mode, plate work has never been easier and faster!

To find out more, visit:
www.sartorius.com/picus

PCR

applied biosystems invitrogen

First-class PCR results served with every run

Applied Biosystems™ thermal cyclers

- Trusted for reliability and precision for over 30 years
- Verified for superior temperature accuracy and uniformity
- Comprehensive portfolio of instruments delivering first-class PCR results for every application

Press the call button at
thermofisher.com/thermalcyclerdemo

ThermoFisher SCIENTIFIC

Find out more about Invitrogen™ PCR reagents at
thermofisher.com/amplifly

CHALLENGE POSSIBILITIES

NTNU is Norway's leading science and technology university. Here, equality, stability and long-term government research funding create freedom to innovate, collaborate and explore new ideas. Together, we challenge — challenge established truths, naysayers and ourselves. We create knowledge for a better world.

www.ntnu.edu

NTNU

Knowledge for a better world

Ultra-High Mass Range Mass Spectrometer

The Q Exactive UHMR Hybrid Quadrupole–Orbitrap Mass Spectrometer from Thermo Fisher Scientific combines substantially increased sensitivity and mass resolution at high *m/z* with MS^2 and pseudo-MS^3 capabilities—creating a unique and powerful analytical solution for investigating the structure and heterogeneity of protein complexes. With its superior ability to perform native intact mass and top-down analyses, Q Exactive UHMR (ultra-high mass range) MS provides unprecedented structural detail not visible with other methods, enabling deeper understanding of protein function, disease mechanisms, potential drug targets, and biotherapeutic compounds. It can analyze intact megadalton assemblies and resolve small differences in masses that reveal key ligands, modifications, and interactions for better understanding of biological processes. With orders-of-magnitude enhanced sensitivity at high *m/z*, it can make high-confidence measurements with very small amounts of sample.

Thermo Fisher Scientific
For info: 800-955-6288
www.thermofisher.com/qexactiveuhmr

Fluorescent Secondary Antibodies

StarBright Blue 520 Fluorescent Secondary Antibodies from Bio-Rad are fluorescent dye–labeled secondary antibodies for use in multiplex Western blotting. They enable highly sensitive fluorescent detection, short exposure times, and easy multiplexing. These plug-and-play antibodies allow simultaneous detection of up to three proteins (two targets of interest and one housekeeping protein) on the same blot when used with hFAB Rhodamine Housekeeping Protein Fluorescent Primary Antibodies. The antibodies work seamlessly on nitrocellulose or low-fluorescence polyvinylidene fluoride (PVDF) membranes, and all other aspects of the standard Western blotting workflow remain unchanged. StarBright Blue 520 antibodies can detect low-abundance protein targets in exposure times that are two to four times shorter than those of traditional fluorescent antibodies.

Bio-Rad
For info: 800-424-6723
www.bio-rad.com/starbright520

Protein Tagging System

Promega Corporation's bioluminescent HiBiT Protein Tagging System provides a small, sensitive, easily quantifiable peptide tag that allows researchers to make better cell models. High Binary Technology (HiBiT) makes CRISPR-mediated tag knockins accessible by eliminating the need for molecular cloning prior to insertion and simplifying detection of tagged proteins. The small tag size makes CRISPR insertion efficient, and bioluminescence enables antibody-free detection with the sensitivity to detect endogenous expression. The HiBiT detection method also eliminates the need for antibody-based techniques for receptor internalization studies. With HiBiT, multiple antibody-binding steps and associated washes are eliminated from detection protocols—simply add the detection reagent and measure a luminescent signal. Classic epitope tagging methods are limited in throughput and sensitivity, require high-quality antibodies, and may only yield semiquantitative results. HiBiT tagging brings the simplicity and sensitivity of bioluminescence to protein abundance studies, quantifying proteins down to endogenous levels, even those maintained at low expression levels.

Promega
For info: 800-356-9526
www.promega.com

Complex Charged Molecule Separator

EAF2000 is a simultaneous electrical and asymmetrical flow field-flow fractionation (EAF4) system designed to enhance molecular separation and characterization. In this system, electrical and crossflow fields are applied simultaneously across the FFF channel, enabling separations by particle size and particle charge based on electrophoretic mobility. Combining these two powerful separation techniques in a single platform opens the door to characterizing complex proteins, antibodies, and viruses as well as environmental and charged nanoparticles or polymers that have proven resistant to other techniques.

Postnova Analytics
For info: +44-(0)-1885-475007
www.postnova.com

Aerosol Inlet System for PTR-TOFMS

The CHemical Analysis of aeRosol ON-line (CHARON) particle inlet, coupled to IONICON proton-transfer-reaction time-of-flight mass spectrometry (PTR-TOFMS) instruments, covers volatile organic compounds (VOCs) and now allows quantitative, molecular-level characterization of sub-µm particulate organic matter as well as particulate ammonium and nitrate at single-digit ng/m^3 mass concentration levels in real time. Future IONICON customers can use a single, real-time analytical instrument to monitor VOCs, intermediate-volatility organic compounds, semivolatile organic compounds, and low-volatility organic compounds, with the all-in-one CHARON PTR-TOFMS.

IONICON
For info: +43-(0)-512-214-800
www.ionicon.com

LC/MS/MS Sample Prep Kit for Monoclonal Antibodies

The nSMOL Antibody Bio-Analysis Kit from Shimadzu Scientific Instruments is a ready-to-use reagent kit for collection of monoclonal antibodies from blood or other biological samples using immunoglobulin collection resin, followed by selective proteolysis of the antigen-binding (Fab) region of these antibodies via trypsin-immobilized nanoparticles. Variable region–derived peptides produced by limited proteolysis can then be quantified via multiple reaction-monitoring measurements utilizing Shimadzu's high-performance LCMS-8050/8060 triple-quadrupole liquid chromatograph mass spectrometer. The simplified workflow eliminates the denaturing, reduction, and alkylation steps normally associated with protein digestion, enabling more efficient sample preparation and analysis. With the nSMOL (nano-surface and molecular orientation limited proteolysis) kit, the antibody region subject to analysis (Fab region) is selectively fragmented, resulting in a sample preparation protocol that dramatically suppresses background noise and ion suppression, and leads to improved response and quantitative repeatability. In addition, the selective collection of Fab peptides limits contamination from excessive peptides or trypsin.

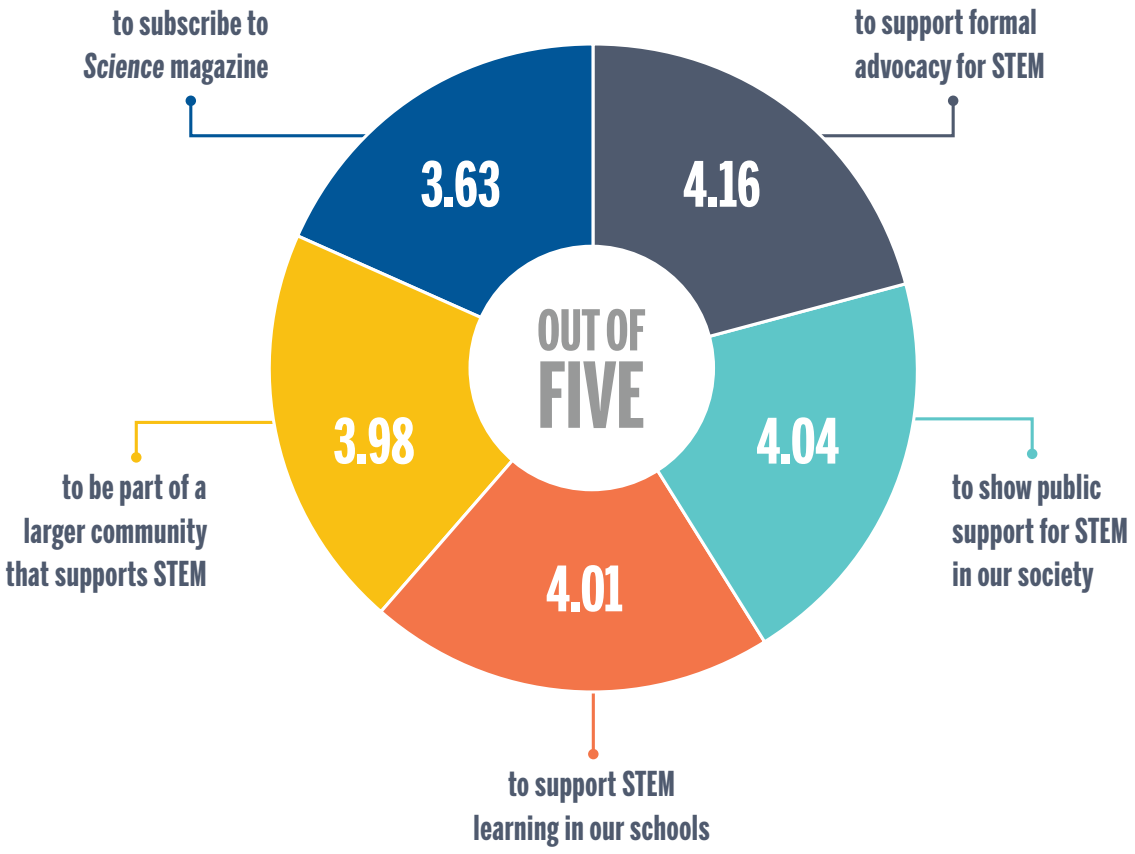
Shimadzu Scientific Instruments
For info: 800-477-1227
www.shimadzu.com/an/lcms/msmol.html

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.

AAAS IS THE FORCE FOR SCIENCE

According to the 2017 Member Survey, you joined AAAS ...



TELL US WHAT'S IMPORTANT TO YOU!

The 2018 Member Survey is launching in September. Look in your inbox for a link.

Your responses help us to better serve science, scientists, and the global community. Don't miss your chance to tell us what's most important to you!

BIOLOGY ■ EARTH ■ ENVIRONMENT ■ LIFE ■ INTERDISCIPLINARY ■ PHYSICAL ■ MATERIAL ■ SOCIAL SCIENCES

Science Advances | AAAS

OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

OPEN ACCESS

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

VISION AND VISIONARY

Tencent YouTu Computer Vision Summit (TCVS)

Tencent YouTu Lab forms strategic partnership with *Science* to promote global development of computer vision

On 6 September, 2018, Tencent YouTu Lab and *Science*—the official publication of the American Association for the Advancement of *Science* (AAAS)—jointly presented the Tencent YouTu Computer Vision Summit in Shanghai, China. Experts and professors from around the world gathered to discuss the development of computer vision technology and its applications.

Dowson Tong
Senior Executive Vice President, Tencent

"Tencent and *Science* will continue to focus on computer vision by gathering global resources and promoting the development of computer vision."

Bill Moran
Publisher, *Science* family of journals, *Science*/AAAS

"At *Science* and AAAS we believe in science without borders; it is all about collaboration and cooperation. Tencent shares this same vision with us."

▲ Tencent YouTu Lab forms a strategic partnership with *Science*.

Tencent YouTu Lab also confirmed its official partnership with *Science*, with the aim of bringing together their global resources in the field of artificial intelligence (AI) research through scholarships, industry-university-research exchanges, extensive cooperation in cutting-edge artificial intelligence research (including co-organizing a webinar on medical AI), and the joint establishment of global computer vision awards and other academic prizes.

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

How low can you go?

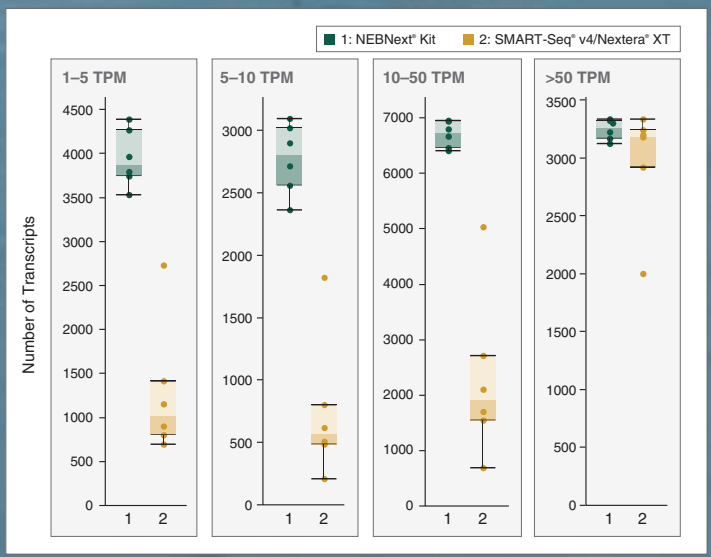
NEBNext® Single Cell/ Low Input Library Prep Kit

With this unique streamlined method, high-quality, full-length transcript sequencing libraries are made directly from single cells, or from as little as 2 pg – 200 ng of total RNA.

- Detect more transcripts, including low-abundance transcripts
- Obtain uniform, full-length transcript coverage, regardless of input amount or sample type
- Save time with a streamlined workflow, minimal handling steps and hands-on time

Visit **NEBNext.com** to request your sample today.

Superior transcript detection with the NEBNext Single Cell/Low Input RNA Library Prep Kit

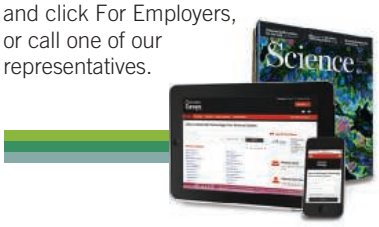


Sequencing libraries were generated from Jurkat single cells (6 replicates) using the NEBNext Single Cell/Low Input RNA Library Prep Kit, or the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing plus the Nextera® XT DNA Library Prep Kit. Libraries were sequenced on an Illumina® NextSeq® 500. Each dot represents the number of transcripts identified at the given Transcripts Per Kilobase Million (TPM) range, and each box represents the median, first and third quartiles per replicate and method. Salmon 0.6 was used for read mapping and quantification of all GENCODE v25 transcripts. Increased identification of low abundance transcripts is observed with the NEBNext libraries.

Science Careers

SCIENCE CAREERS ADVERTISING

For full advertising details, go to ScienceCcareers.org and click For Employers, or call one of our representatives.



AMERICAS

+1 202 326-6577
+1 202 326-6578
advertise@sciencecareers.org

EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527
advertise@sciencecareers.org

CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012
advertise@sciencecareers.org

JAPAN

+81 3-6459-4174
advertise@sciencecareers.org

CUSTOMER SERVICE

AMERICAS

+1 202 326-6577

REST OF WORLD

+44 (0) 1223 326528

advertise@sciencecareers.org

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

Faculty Hires in Marine Science

"Discovery Starts Here"

The University of Texas at Austin Marine Science Institute is experiencing exciting growth in its nationally and internationally recognized research programs and invites applications for a faculty position in the Department of Marine Science. We seek leading innovative and productive scientists at the forefront of their disciplines to join our faculty and expanding programs. The position, to be filled at the tenured faculty rank of Associate or Full Professor, will be in the area of ***Marine Chemical Ecology*** – eligible for the ***Mary Anderson Abell and Joseph Miles Abell, Jr., M.D. Endowed Chair in Marine Science***, as a Chair Fellow or Holder, respectively. Areas of interest include laboratory and field methods to study intra- and interspecific chemical cues and signals of marine flora and fauna; elucidate chemical structures of cues and signals and their roles in ecosystem structure and function; and investigate the identities and mechanisms of xenobiotic disruption of chemical cues and signals in the marine environment. The successful applicant will be expected to establish an internationally recognized research program at the Institute, mentor graduate students and postdoctoral scientists, and teach at the graduate and undergraduate levels. We are interested in individuals who will contribute to diversity and equal opportunity in higher education through their teaching, research, and service.

The position is located at the Marine Science Institute in Port Aransas, Texas, which offers close proximity to a variety of unique estuarine and coastal habitats as well as excellent shoreside facilities for experimental work. Our infrastructure includes a newly renovated Estuarine Research Center in partnership with NOAA that expands the research capacity of the Marine Science Institute. This building houses the Mission-Aransas National Estuarine Research Reserve, a Marine Library, and 12,000 square feet of space for faculty offices and state-of-the-art research laboratories. New additions under development include a 10,000 square foot building for the new Center for Coastal Ocean Science that will house chemical ecology and toxicology research. Successful applicants will have the freedom to follow independent and collaborative research, and will be provided with 9 months of state-funded salary support for research (75%) and teaching (25%).

The University of Texas Marine Science Institute is responsive to needs of dual career couples and is committed to equality of opportunity. We strongly encourage women and racial/ethnic minorities to apply. Applications should include a letter of interest, curriculum vitae, teaching statement, a research statement, and a list of three potential references sent electronically to Interfolio: **https://home.interfolio.com/15714**. Review of applications will begin **January 2019**.

The University of Texas at Austin is an Equal Opportunity Employer.
Background check conducted on applicant selected.

BNU Seeking Top Talents in Systems Science

Feel free to contact us:
Website: http://sss.bnu.edu.cn
Email: sss@bnu.edu.cn
Address: School of Systems Science, Beijing Normal University, No. 19, XinJieKouWai St., Haidian District, Beijing, P.R.China 100875

Beijing Normal University (BNU), as a venerable and dynamic institution, has a long-held tradition of partnering with leading academic institutions from around the world. Noted for its culturally rich and diverse environment, BNU endeavors to become a leader of higher education in China, Asia and even the world.

The School of Systems Science (SSS) at BNU advocates interdisciplinary research in systems science through collaboration with other disciplines in natural, social, technical and economic sciences inside and outside BNU. Having such a uniquely collaborative environment together with outstanding research facilities, SSS aspires to excel internationally in the field of Systems Science.

In 2018, SSS will establish an International Science Center for Complex Systems on the Zhuhai campus of BNU in southern China. This center aims at the frontiers of scientific research and technical innovation, specifically in the formation mechanism of human decision making behavior and its neural mechanism, together with data analysis of human behavior and artificial intelligence (including swarm intelligence). By mobilizing its top experts to lead key research projects in cooperation with talents around the world, this science center is expected to become an international platform for systems science development as well as an interface between academia, industry and governmental authorities.

In order to carry out cutting-edge research in systems science, SSS intends to expand its research team. Its current research portfolio includes, but is not limited to the following fields:

i. The fundamental theories of complex system.

ii. Social and economic system.

iii. The life ecosystem and the self-organizing behavior of brain and cognition.

iv. Multi-agent system and evolutionary algorithm.

v. Information technology of artificial intelligence systems.

vi. The science of science.

I Think the Next Century Will Be the Century of Complexity.

––Stephen Hawking, 2000.

SSS sincerely welcomes famous scholars, academic leaders, promising young scholars and postdocs in interdisciplinary areas of systems science. The applicant is expected to have a doctoral degree in relevant orientation, a solid research foundation and outstanding research findings in the above-mentioned fields, with great potential for being an excellent teacher and tutor, at the same time an excellent team player in interdisciplinary collaboration.

BNU will place each successful applicant in a proper position according to his/her academic background, research interests and teaching plan, and provide him/her with competitive salary, sufficient start-up fund, necessary laboratory and office space, and other reasonable supports. During the employment, BNU will evaluate each faculty member's capacity and potential and provide the best opportunities for vocational development in line with the international conventions and common practices in academic circles.

Over the years, SSS has made great contributions to the sustainable development of systems science in and outside China. Faced with the complex social and technological challenges, SSS will resume its effort to solve scientific problems in nature and society. By integrating many different yet relevant disciplines and building the international advanced research center of complex systems, SSS strives to cultivate high-level interdisciplinary talents so as to generate new knowledge, new minds, and new technologies that will promote the creation of a shared future for mankind. In time, SSS will become a crucial education and innovation incubator for big data, artificial intelligence, future brain and intellectual education.

SSS cordially welcomes job applicants and visiting scholars with expertise in systems science and related areas. The school also welcomes research and education collaboration from China and the rest of world.

TEXAS A&M UNIVERSITY
Biology

Two Faculty Positions in Microbiology

The Department of Biology at Texas A&M University (TAMU) invites applications for two tenure-track, assistant professor positions in microbiology.

We will consider candidates pursuing innovative research in all disciplines of microbiology. The criteria for selection will be uniqueness, creativity, and excellence in research and scholarship. We require all candidates to have a Ph.D. in Microbiology or a related field, and we strongly encourage applications from candidates who will increase the exposure of our students to a diverse culture.

The successful candidate will be expected to develop an externally funded research program and to teach undergraduate and graduate courses. We offer an interactive and collegial research environment, a modern infrastructure, and a competitive startup package. More information about our department can be found at www.bio.tamu.edu. For full consideration, applicants should email a letter of intent, *curriculum vitae*, statement of research and teaching interests, and three letters of recommendation by **November 15, 2018** to **microsearch@bio.tamu.edu**

If you have questions about this search, please direct e-mails to Dr. Bruce Riley, Chair of the Search Committee, at microsearch@bio.tamu.edu

The Texas A&M System is an Equal Opportunity/Affirmative Action/ Veterans/Disability Employer committed to diversity. Texas A&M University, the College of Science and the Department of Biology are dedicated to the goal of building an inclusive and culturally diverse faculty and staff who are committed to teaching and working in an environment of academic freedom and equality of opportunity.

TEXAS A&M UNIVERSITY
Biology

Faculty Position in Developmental Biology

The Department of Biology at Texas A&M University (TAMU) invites applications for a tenure-track, assistant professor position in developmental biology.

We will consider candidates pursuing innovative research in any area of developmental biology. The criteria for selection will be uniqueness, creativity, and excellence in research and scholarship. We require all candidates to have a Ph.D. in biology or a related field, and we strongly encourage applications from candidates who will increase the exposure of our students to a diverse culture.

The successful candidate will be expected to develop an externally funded research program and to teach undergraduate and graduate courses. We offer an interactive and collegial research environment, a modern infrastructure, and a competitive startup package. More information about our department can be found at www.bio.tamu.edu. For full consideration, applicants should email a letter of intent, *curriculum vitae*, statement of research and teaching interests, and three letters of recommendation by **November 15, 2018** to **developmentalbiosearch@bio.tamu.edu**

If you have questions about this search, please direct e-mails to Dr. Bruce Riley, Chair of the Search Committee, at developmentalbiosearch@bio.tamu.edu

The Texas A&M System is an Equal Opportunity/ Affirmative Action/ Veterans/ Disability Employer committed to diversity. Texas A&M University, the College of Science and the Department of Biology are dedicated to the goal of building an inclusive and culturally diverse faculty and staff who are committed to teaching and working in an environment of academic freedom and equality of opportunity.

THE UNIVERSITY OF UTAH

The new **Cluster in Evolutionary Genetics and Genomics at the University of Utah** (http://www.ucegg.org), a joint initiative between the Human Genetics and Biology Departments, invites applications for a tenure-track position based within the Department of Human Genetics (School of Medicine). Individuals hired under this initiative will complement related hires in the School of Biological Sciences (College of Science). Successful applicants will join a vibrant intellectual community across the University of Utah working at the interface of genomics, computational biology, and molecular biology. The Cluster will build on substantial, ongoing investments in genome science and experimental basic science at the University of Utah, a unique and inclusive community at the foothills of the Wasatch Mountains with nearby world-class cultural and recreational opportunities.

Competitive candidates will demonstrate an ability to develop independent and vigorous research programs. Applications by investigators, both experimental and theoretical, in all areas of evolutionary genetics and genomics of any organism(s) are encouraged. Examples of areas that will be considered include (but are not limited to): microbial evolutionary genomics, vertebrate population genomics, comparative evolutionary genetics and developmental biology, and human evolutionary genomics. Faculty hired through this initiative will participate in the development and teaching of innovative courses. We expect to make two faculty appointments in the 2018-2019 academic year. Openings are at the Assistant Professor level, but exceptional candidates at the Associate Professor and Professor level will be considered.

All applicants should submit: (1) a cover letter with contact information; (2) a CV, (3) description of current and proposed research (3 pages maximum); (4) a statement of teaching experience and interests (1 page maximum); (5) a description of past and/or potential contributions to advancing diversity, inclusion, and equitable access to education (1 page maximum); and (6) up to three reprints and/or preprints. Applicants at the Assistant Professor level should also arrange for three letters of recommendation to be sent on their behalf before the deadline. Applications should be submitted by **October 1, 2018** but later submissions may be considered. Applicants to other searches at the University of Utah with appropriate experience may be considered in this applicant pool, **https://utah.peopleadmin.com/postings/68353**

ASSISTANT PROFESSOR
MOLECULAR BIOLOGY AND BIOCHEMISTRY
RUTGERS UNIVERSITY

The Department of Molecular Biology and Biochemistry at Rutgers, The State University of New Jersey, New Brunswick (Busch Campus), invites applicants for a tenure-track **Assistant Professor** position. The position requires a Ph.D. degree and involves teaching undergraduate and graduate students and research in Molecular Biology. The Department is especially interested in applicants who use biochemical or molecular approaches for research in functional genomics, epigenetics, chromatin structure and function, and control of growth and differentiation but outstanding researchers in other areas will be considered. The Department is an important part of the expanding program in molecular biology within the Division of Life Sciences on the Busch Campus, where the Center for Advanced Biotechnology and Medicine, the Waksman Institute and the Robert Wood Johnson Medical School are also located. We have strong consolidated, interdepartmental graduate programs in molecular biosciences.

The position is highly competitive with regard to start-up funds, laboratory space and salary. Women and minority candidates are encouraged to apply. Interested individuals must apply online at: **http://jobs.rutgers.edu/postings/75714**

Rutgers University is an Equal Opportunity/Affirmative Action Employer.

Faculty Positions in Ecology

The Department of Ecology and Evolutionary Biology (EEB) at UCLA announces faculty searches for Assistant Professor (tenure-track position) or Associate Professor (tenure position) in ecology, broadly defined. We seek ecologists studying processes operating at population to global scales, in either terrestrial or marine habitats. Preference will be given to candidates who study plants, invertebrates, and/ or microbes, as well as those whose research builds connections among existing areas of strength both within the EEB Department and between the department and broader UCLA community. The successful candidate will be expected to develop a rigorous, externally funded research program and teach at both the undergraduate and graduate levels. EEB is a vibrant and expanding department with strengths in evolutionary and conservation genomics, field biology, quantitative biology, and behavior. The department is part of a dynamic interdepartmental research community in computational biology, bioinformatics, genomics, and conservation biology across the greater UCLA campus. The successful candidate would be a part of this rich network, with many opportunities to collaborate between disciplines and departments. Necessary qualifications include a PhD degree in ecology, botany, marine science, or related fields.

Please direct inquires to Associate Professor Nathan Kraft (nkraft@ucla.edu). Submit application packages online through **https://recruit.apo.ucla.edu/apply/JPF04004** and include the following: (1) cover letter (2) curriculum vitae; (3) statement of research interests; (4) statement of teaching expertise; (5) statement of formal and informal activities to promote equity, diversity and inclusion; and (6) names of three referees. All items should be distinct documents. Individuals with a history of mentoring students under-represented in the sciences are encouraged to apply and to describe their experience in a cover letter. The University of California seeks to recruit and retain a diverse workforce as a reflection of our commitment to serve the people of California, to maintain the excellence of the University, and to offer our students richly varied disciplines, perspectives and ways of knowing and learning. Review of applications will begin on November 9, 2018, and continue until position is filled.

The Department of Ecology and Evolutionary Biology has 29 faculty, a large graduate program, over 2000 undergraduates across three majors (Biology; Ecology, Behavior, and Evolution; Marine Biology), and two minors (Conservation Biology and Evolutionary Medicine). The department has close ties with the Institute of the Environment and Sustainability, and departments across the College, David Geffen School of Medicine, and the School of Public Health. EEB is associated with the UCLA La Kretz Center for California Conservation Sciences, UCLA Stunt Ranch UC Reserve, the Mildred E. Mathias Botanical Garden, the Donald E. Dickey Collection of Birds and Mammals, the Center for Education and Innovation and Learning in the Sciences, and the Institute for Quantitative and Computational Biosciences at time of hire.

EEB is administered through the Division of Life Sciences (www.lifesciences.ucla.edu), which includes over 200 faculty, 8000 undergraduates, 500 graduate students, 12 undergraduate majors, and more than 12 PhD programs. The UCLA College through its four academic divisions— Humanities, Life Sciences, Physical Sciences, and Social Sciences – is the academic heart of UCLA, which is California's largest university with an enrollment of nearly 43,000 undergraduate and graduate students.

The University of California is an Equal Opportunity/Affirmative Action Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, age or protected veteran status. For the complete University of California nondiscrimination and affirmative action policy, see: UC Nondiscrimination & Affirmative Action Policy. (http://policy.ucop.edu/doc/4000376/NondiscrimAffirmAct)

Faculty Positions in Translational Neurosciences

As part of a broad institutional initiative in Integrative Biosciences, the Translational Neuroscience Initiative (TNI) at Wayne State University (WSU) is recruiting up to three faculty positions (tenured or tenure-track, open rank). The program in translational neuroscience fosters interdisciplinary, integrative, and collaborative approaches to understand translational neuroscience and its applications to trauma neurobiology including but not limited to post-traumatic stress disorder (PTSD). Primary areas for recruitment are as follows.

- Innovative translational approaches to trauma neurobiology assessment, neurobiological mechanisms, identification of therapeutic targets, and measurement/ prediction of treatment response (biomarkers targeting comorbidity with mood disorders, pain, TBI, and substance use disorders)
- Integration of basic and clinical neuroscience, development of novel pharmacological agents and non-pharmacological interventions for PTSD
- Impact of trauma on brain development using cutting edge neuroimaging, neurostimulation, psychophysiological, or similar techniques in children, adolescents, and adults (including but not limited to MRI, fMRI, rsMRI, DTI, rTMS, deep brain stimulation)

The TNI promotes integrative and collaborative research in basic and clinical neuroscience and is situated in the new 200,000 sq. ft. Integrative Biosciences Center (IBio) that houses coordinated inter- and trans-disciplinary research teams, and a Clinical Research Center. IBio is a fulcrum for leading-edge technology platforms and specialized resources in support of advanced studies in precision medicine. Through research, clinical care, community engagement, and education, the TNI team of researchers and community partners seek to discover, investigate, and solve complex health problems that affect the human nervous system. Successful faculty candidates will have a Ph.D., M.D., or equivalent degree in biomedical sciences relevant to neurosciences with evidence of peer recognition in the field, a commitment to excellence in research education and training, and the ability to engage with broader neuroscience themes for the purpose of achieving transformative and translational research gains. Applicants are expected to have already established extramural research funding or to be on a clear path to secure extramural funding in support of their research programs. Faculty recruits will integrate with departments and colleges or schools consistent with their areas of expertise and shared interests and engage in all aspects of our academic mission, including research, training, instruction and service.

Competitive recruitment packages are available with salary and faculty rank based on qualifications. Applicants are encouraged to apply to posting #043730 through the WSU Online Hiring System **https://jobs.wayne.edu**. Applications will be accepted until positions are filled, but for full consideration this fall, application materials should be submitted by **November 30, 2018**. Applications should include a curriculum vitae and a brief narrative cover letter addressed to the Director of the PTSD and Trauma Neurobiology Program of the TNI, Dr. Tanja Jovanovic, and the Vice President for Research, indicating the applicant's potential for research synergy within the TNI and the broader institutional initiative in integrative biosciences.

Wayne State University is a premier, public, urban research university located in the heart of Detroit where students from all backgrounds are offered a rich, high quality education. Our deep-rooted commitment to excellence, collaboration, integrity, diversity and inclusion creates exceptional educational opportunities preparing students for success in a diverse, global society. WSU encourages applications from women, people of color and other underrepresented people. WSU is an affirmative action/equal opportunity employer.

Founded in 1868, Wayne State University offers a range of academic programs through 13 schools and colleges to nearly 28,000 students. The campus in Midtown Detroit comprises 100 buildings over 200 acres including the School of Medicine, the Eugene Applebaum College of Pharmacy and Health Sciences and the College of Nursing. The university is home to the Perinatology Research Branch of the National Institutes of Health, the Karmanos Cancer Center, a National Cancer Institute-designated comprehensive cancer center, and a National Institute of Environmental Health Sciences Core Center - *Center for Urban Responses to Environmental Stressors (CURES)*

Science Careers Virtual Job Fair

Research in Germany
Land of Ideas

Federal Ministry of Education and Research

Research in Germany

November 7, 2018, Wednesday
9:00 AM - 1:00 PM EST; 3:00 PM – 7:00 PM CET

JOB SEEKERS! ***Science*** **Careers** has teamed up with **Research in Germany** to offer an exciting job fair focused on career opportunities in Germany. Meet online with representatives of German universities and research and funding organizations to learn about positions for Ph.D. students and postdoctoral researchers, funding opportunities, and studying and research in Germany.

For more details and to register, visit: *ScienceCareers.org/virtualcareerfair*

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Special Job Focus:

Cell Biology

Issue date: November 16
Book ad by November 1
Ads accepted until November 9 if space allows

To book your ad:
advertise@sciencecareers.org

The Americas
+ 202 326 6577
Europe
+44 (0) 1223 326527
Japan
+81 3 6459 4174
China/Korea/Singapore/Taiwan
+86 131 4114 0012

SCIENCECAREERS.ORG

Why choose this Cell Biology focus for your advertisement?

- Relevant ads lead off these career sections with a special "cell biology" banner
- Bonus distribution to: American Society for Cell Biology, December 8–12, San Diego, CA.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Produced by the *Science*/AAAS Custom Publishing Office.

FOR RECRUITMENT IN SCIENCE, THERE'S ONLY ONE *SCIENCE*.

By **Jet-Sing M. Lee**

Leaving my comfort zone

I never thought I would venture far from home. I was a mediocre high school student from a working-class family, and I didn't think traveling the world was in the cards for me. For my undergrad degree, I only applied to universities within an hour of my hometown in the United Kingdom. When I decided to pursue a Ph.D., I stayed at the same university. I still wasn't convinced I was a great student (even though I was near the top of my class), and staying seemed like the safe option. It allowed me to remain close to friends, and I was familiar with the department, which put me in a good position to pick my supervisor wisely. A few years later, as I was considering a postdoc, I was tempted to once again stay close to home, in the comfort of my Ph.D. lab. My experience there had been very positive, and previous doctoral students had stayed on, so why shouldn't I? It would have been so easy. But I hope to run my own lab one day, and I knew that seeing a different approach to science would be valuable.

I heard about a fellowship program in Japan that sounded like a good fit. I had always been interested in Japan, and the flexibility of the program appealed to me. I could choose the institution where I would work, the research I would conduct, and how long I would be there—from as little as a month to as much as a year.

I decided that, if I were accepted, I would stay for 3 months—long enough to experience a new place, learn, and get some research done, but short enough to make the prospect less intimidating. I thought of it almost as a holiday, somewhat like going to a conference in an interesting far-off place.

But when I was accepted, the head of the group I was going to join wanted me to stay for the full year. I hesitated. I was excited about the research and the opportunity to learn and experience a new environment. But the thought of being so far from home for so long made me anxious. I didn't speak any Japanese. I wouldn't know anyone. Would I be miserable?

After much deliberation, I worked up my courage and signed on for the year—knowing that, if things went badly, I could come home earlier.

Almost everything was different from what I was used to. Even commuting to lab and shopping for groceries were so unfamiliar that I found myself constantly taking out my phone to snap pictures, like a tourist. In the lab, too, the practices and culture were new to me. The expected working hours were different. Lab technicians did much of the work that I was accustomed to doing myself. In group meetings, questions were restricted to the end, as opposed to the more open structure of my old lab. Many of my new colleagues were not native English speakers and came to me for language help, which was a new experience for me.

"I knew that seeing a different approach to science would be valuable."

But getting used to these differences was easier than I had expected. One of my favorite new routines was going out to lunch with my labmates. Back in the United Kingdom, going out was too expensive, so I usually packed my lunch and ate it in my office. In Japan, on the other hand, eating out is common and offers a great bonding opportunity.

My new co-workers came from all over the world, primarily across Asia. My previous lab had also been international, but most of my colleagues there were from Europe and China. Talking with my new colleagues about their experiences conducting research and garnering funding in other countries and scientific cultures helped open my eyes to new ways of doing things.

After my fellowship, I thought I would leave Japan to do a postdoc elsewhere. But when my fellowship adviser offered me a longer-term position in his lab, I couldn't say no. This time, though, it wasn't because I was afraid to go somewhere new. It was because I wanted to take advantage of an exciting opportunity. After the leap I took with my fellowship, I now feel I can do anything. ■

Jet-Sing M. Lee is an assistant professor of chemistry at Kyoto University in Japan. Do you have an interesting career story? Send it to SciCareerEditor@aaas.org.