How a rocket scientist helped engineer Chinese society *p. 1206*

Daily cycles of transcription in a primate *pp. 1210 & 1232*

Logical reasoning in infants *pp. 1214 & 1263*

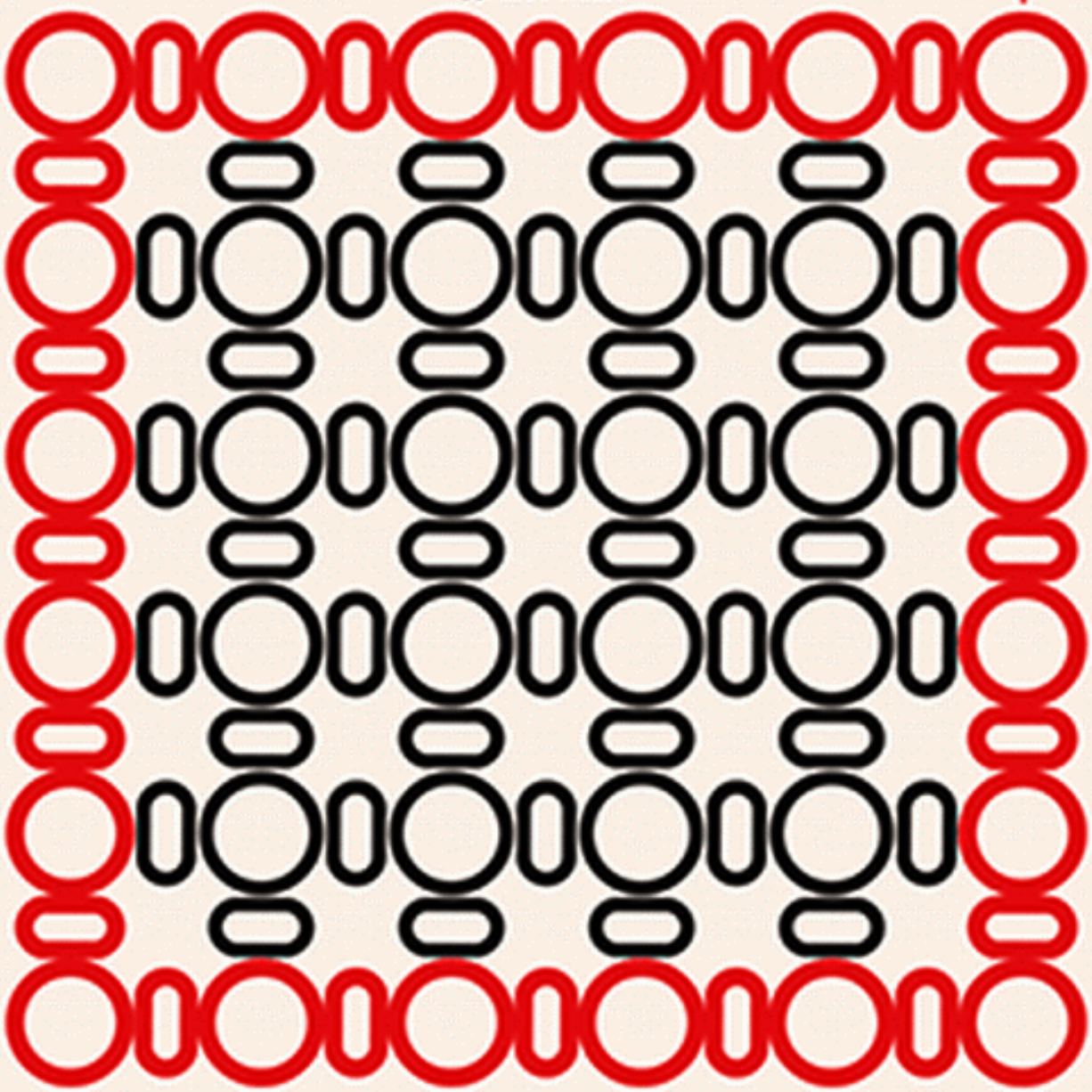
Science

$15
16 MARCH 2018
sciencemag.org

AAAS

LASING ON THE EDGE

pp. 1230 & 1231



CONTENTS

16 MARCH 2018 • VOLUME 359 • ISSUE 6381

Systems engineer Qian Xuesen had fans in high places.

NEWS

IN BRIEF

IN DEPTH

FEATURE

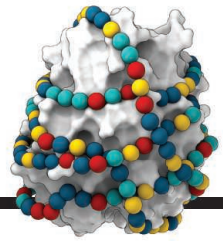
INSIGHTS

PERSPECTIVES

POLICY FORUM

BOOKS *ET AL.*

CONTENTS



1216 & 1239
Plastics helping proteins

16 MARCH 2018 • VOLUME 359 • ISSUE 6381

LETTERS

RESEARCH

IN BRIEF

RESEARCH ARTICLES

TOPOLOGICAL PHOTONICS

REPORTS

DEPARTMENTS

ON THE COVER

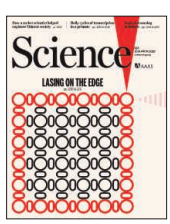


Illustration of a non-magnetic topological insulator laser: an array of microring resonators (circles) coupled by waveguides (ellipses) in a topological fashion. The red perimeter depicts the lasing topological edge mode, which exits through one of the two outputs (solid red triangle). Implementation of this design in an all-dielectric semiconductor enhances lasing performance and makes many diode lasers operate together as a single highly coherent laser. See pages 1230 and 1231.
Illustration: Valerie Altounian/Science

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

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

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

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

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

Published by AAAS

EDITORIAL

Good gun policy needs research

The tragic shooting at a school in Parkland, Florida, last month triggered another round of proposals from local, state, and federal policymakers about controlling firearm-related violence without violating broad interpretations of the rights to keep and bear arms provided by the U.S. Constitution. Unfortunately, there is only very sparse scientific evidence that can help figure out which policies will be effective. Earlier this month, the RAND Corporation released a comprehensive analysis on gun policy in the United States,* and among its conclusions is that too few policies and outcomes have been the subject of rigorous scientific investigation. Even the seemingly popular view that violent crime would be reduced by laws prohibiting the purchase or possession of guns by individuals with mental illness was deemed to have only moderate supporting evidence. If the nation is serious about getting firearm-related violence under control, it must rise above its aversion to providing financial support for firearm-related research, and the scientific community will have to expeditiously carry out the needed research.

"...too few policies and outcomes have been the subject of rigorous scientific investigation."

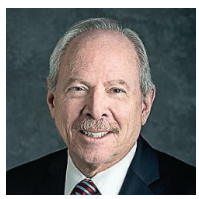
There used to be more federally funded research on firearm-related violence than there is now. Although its program was small relative to other public health issues, the U.S. Centers for Disease Control and Prevention (CDC) did support research on firearm violence through its National Center for Injury Prevention and Control. However, the 1996 "Dickey Amendment" prohibited the CDC from funding activities that promoted or advocated for gun control. Although arguably some research might still have been considered acceptable, the amendment was interpreted as an outright prohibition of CDC support for any gun violence research. In 2011, Congress enacted similar restrictions affecting the Department of Health and Human Services, resulting in a dearth of scientific activity on any aspect of the availability and possession of firearms and the violence that might be related to them.

In 2013, shortly after the Sandy Hook Elementary School shooting in Connecticut, President Obama issued a series of Executive Orders directing federal science agencies to again support research on firearm-related violence. In response, the CDC and the CDC Foundation asked the Institute of Medicine (IOM) and the National Research Council (NRC) (now part of the National Academies of Sciences, Engineering, and Medicine) to draw up a public health research agenda that the CDC might implement. One of us (A.I.L.) chaired the committee that produced the resulting consensus report and research agenda.† Unfortunately, nothing came of it, because although CDC has occasionally requested funds to support relevant research, Congress consistently has said "no." The National Institutes of Health also supported relevant research beginning in 2013, but that program was discontinued in 2017 as other priorities replaced it.

It's time to stop the polarized "debates" that lack a science base and turn our energies toward constructive, informed examination. The IOM-NRC report has spelled out a research path that calls for a closer examination of the characteristics of firearm-related violence; the risk and protective factors (like growing up in violence-prone environments) that increase the probability of firearm violence; the effectiveness of diverse violence prevention and other interventions; and the impact of various gun safety technologies. And the RAND analysis calls for research on the effects of specific firearm policies, such as whether background checks that investigate all types of mental health histories do reduce gun injuries, suicides, and homicides and whether raising the minimum age for purchasing firearms (to 21 years old) reduces firearm suicides among youth.

Without science to drive firearm policy development and implementation, we risk inventing policies based on personal ideology or intuition. If we are serious that gun violence is a major public threat, then let's rise to the moment and take the next science-based steps.

–Alan I. Leshner and Victor J. Dzau



Alan I. Leshner is the chief executive officer emeritus of the American Association for the Advancement of Science and is the former executive publisher of Science. *alan.i.leshner@gmail.com*

Victor J. Dzau is president of the U.S. National Academy of Medicine (formerly the Institute of Medicine), Washington, DC, USA. vdzau@nas.edu

*www.rand.org/pubs/research_reports/RR2088.html; †https://doi.org/10.17226/18319

10.1126/science.aat5127

Ultima Thule

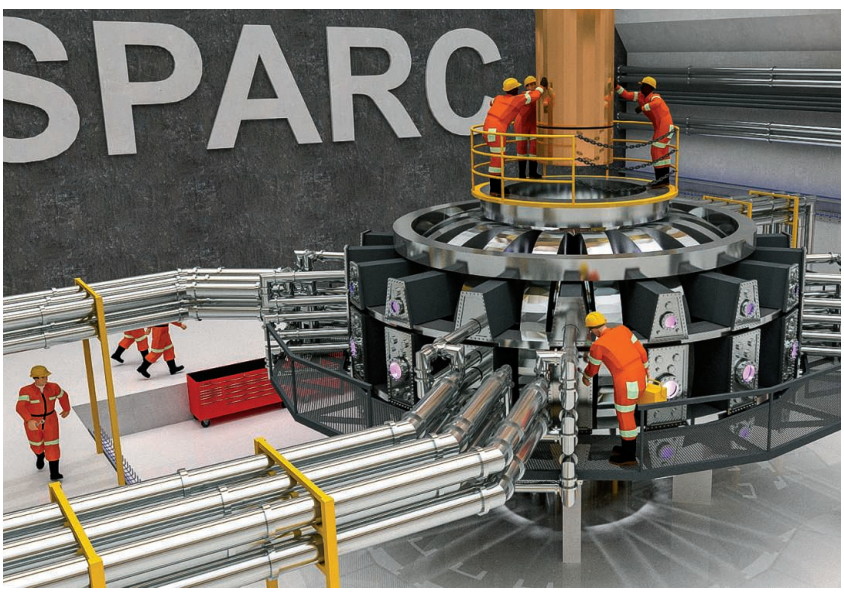
The name chosen by NASA for the next mission by the New Horizons spacecraft, to MU69, a remote object beyond Pluto. Thule is a mythical, far-northern island.

IN BRIEF

Edited by **Jeffrey Brainard**

FUSION ENERGY

Compact reactor vies with ITER project



The SPARC reactor, shown in an illustration, could lower costs of fusion power.

The Massachusetts Institute of Technology (MIT) in Cambridge has spun off a company to develop a compact reactor that could lead to practical fusion power. The Italian oil and gas company Eni agreed to pump $50 million into an MIT startup called Commonwealth Fusion Systems developing the Soonest/Smallest Private-Funded Affordable Robust Compact (SPARC) reactor. A fusion power plant would use powerful magnets to confine a superheated hydrogen plasma, in which nuclei fuse and give off energy. The U.S. government has squeezed its funding of fusion research, including to the international ITER reactor project being built near Cadarache in France. SPARC is designed to generate power in a reactor a fraction of ITER's size and cost. It would rely on magnets made from high-temperature, copper-based superconductors that can handle the strong currents in a high magnetic field. The team aims to make magnets more powerful than ITER's at a smaller scale. Commonwealth Fusion Systems has a rival in Tokamak Energy, a startup in Oxford, U.K., that has been building a similar reactor with high-temperature superconducting magnets (*Science*, 28 April 2017, p. 360).

Columbia neuroscientist ousted

WORKPLACE | Prominent neuroscientist Tom Jessell was fired by Columbia University and the Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland, earlier this month after a university investigation. The probe, Columbia said in a statement, "revealed serious violations of University policies and values governing the behavior of faculty members in an academic environment." HHMI, which had supported Jessell's work at Columbia since he arrived there in 1985, ousted him on 2 March, citing a violation of its policy. Neither Columbia nor HHMI would describe the nature of the violations. HHMI, citing personnel policy, would not say whether it conducted its own investigation or relied on Columbia's findings. A Columbia official said Jessell's lab, which listed 22 staff and students in January, will be disbanded over the next 15 months. Scores of former and current Columbia students and employees protested the university's refusal to describe specific reasons for Jessell's dismissal in an online letter to President Lee Bollinger and the university's trustees dated 10 March. Jessell, 66, identified mechanisms involved in the neural development and organization of the spinal cord, revealing principles now known to apply throughout the brain.

Help wanted at NASA

SPACE EXPLORATION | NASA's acting administrator, Robert Lightfoot, who has led the agency since January 2017, announced this week that he will retire on 30 April. The move places pressure on President Donald Trump's administration and the Senate to resolve a battle over leadership for the agency. Last September, the White House nominated Representative Jim Bridenstine (R–OK) to head NASA, but the nomination has stalled in the Senate because of opposition from Democrats and Florida's two senators, Bill Nelson (D) and Marco Rubio (R). Both have said a "space professional" should lead NASA. Lightfoot's surprise retirement could force action, says John Logsdon, professor emeritus at The George Washington University's Space Policy Institute in Washington, D.C.

PHOTO: KEN FILAR/PSFC RESEARCH AFFILIATE

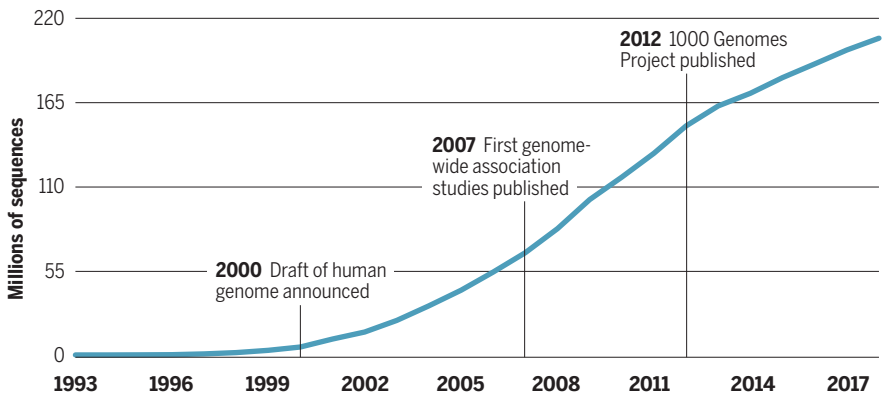
"The Senate should vote up or down on Bridenstine, or the White House should convince him to withdraw" and name a new nominee.

Scientists face the voters

U.S. ELECTIONS | Central Texans embraced a former community college math professor as the slew of scientists running for U.S. Congress this year got their first feedback from voters in primary elections last week. Mary Wilson, now a minister in Austin, won a plurality in the Democratic primary for the 21st congressional district, topping engineer Joseph Kopser, an Army veteran and clean energy entrepreneur from San Antonio. The two will face off on 22 May, with the winner vying for an open seat in November in the solidly Republican district. Houston clinical oncologist Jason Westin fell short in the seventh congressional district, finishing third in a seven-person race. Next week, the nation's attention will shift to Illinois, where Representative Dan Lipinski, a conservative Democrat and longtime science advocate, faces a serious primary challenge from progressive businesswoman Marie Newman. The governor's race is also worth watching: In the Democratic primary, state legislator Daniel Biss, who received a Ph.D. in theoretical mathematics from the Massachusetts Institute of Technology in Cambridge, appears to be closing the gap against hotel heir J. B. Pritzker. The winner will likely face incumbent Republican Bruce Radner. Follow these and other scientists running for Congress with our candidate tracker at http://scim.ag/Sciencecandidates.

Multiplying genomes

The number of DNA sequences from humans and other species stored in GenBank at the U.S. National Center for Biotechnology Information has expanded rapidly in 25 years. By 2025, sequencing projects are expected to have generated at least as many data as the field of astronomy, YouTube, or Twitter—three other major producers of data.



Physics popularizer suspended

WORKPLACE | Arizona State University (ASU) in Tempe placed physicist and outspoken atheist Lawrence Krauss on paid administrative leave and banned him from campus on 6 March, on the heels of a 22 February BuzzFeed article in which several women alleged sexual misconduct by Krauss, including unwanted touching. The university says it has received no formal complaints against Krauss from students, staff, or faculty and began an investigation only after BuzzFeed reporters began asking about him. (An Australian microbiologist not affiliated with ASU filed a complaint about Krauss with the university in 2017.) On 7 March, Krauss tweeted a nine-page rebuttal to the BuzzFeed article to his nearly half-million followers, saying the story presented "false allegations" and "absurd claims." Also last week, Krauss resigned as chair of the board of sponsors of the *Bulletin of the Atomic Scientists*, saying he didn't want the allegations to distract the organization. In addition, the American Humanist

WILDLIFE

French court rules for more bears

The French government hasn't done enough to protect the country's fragile population of European brown bears and must pay €8000 each to two pro-bear lobby groups, a court in the city of Toulouse ruled on 6 March. Almost 40 brown bears (*Ursus arctos*) now live in the French Pyrenees, thanks in part to introductions from Slovenia in 1996 and 2006. But their presence is controversial; sheep farmers claim that bear attacks on their flocks are threatening their livelihoods, despite a system to compensate them for such losses. The two groups said the government had to introduce more bears to help out two lonely males and increase the population's genetic diversity. The court says France's failure to implement a bear conservation strategy constitutes a breach of the European Union's Habitats Directive, a common conservation policy adopted in 1992.

Veterinarians carry a brown bear captured in Slovenia for transport to France in 2006 to help build the population there.

CREDITS: (GRAPHIC) J. BRAINARD AND J. KAISER/*SCIENCE*; (DATA) NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION, U.S. NATIONAL INSTITUTES OF HEALTH; AND Z. STEPHENS *ET AL.*, *PLOS BIOLOGY* 13(7) 2015; (PHOTO) STRINGER/AFP/GETTY IMAGES

Part of Trindade Island, off of Brazil's coast, would be protected.

CONSERVATION

Brazil to create marine reserves

Brazil's government plans to next week create two new marine protected areas, some of the world's largest, but controversy has arisen over their boundaries. The preserves, centered around the remote Saint Peter and Saint Paul Archipelago and Trindade Island, harbor unique marine ecosystems and dozens of endemic species. Each of the protected areas will be about the size of Sweden, and together they will increase the extent of set-asides in Brazil's coastal and oceanic waters to 25%, up from only 1.5%. Scientists protested that boundaries of no-take zones within the protected areas did not appear to include the islands themselves, where Brazil's navy maintains bases, and shallow reef systems around them. The navy says portions of the islands will be included. Scientists in Brazil want the government to also preserve coral reefs and other endangered environments in coastal waters, where conflicts with economic activities are more common than in the protected areas, which contain mostly open ocean.

Association, a group in Washington, D.C., that represents atheists, removed Krauss from its pool of speakers and writers and said it is reviewing his 2015 Humanist of the Year award.

Nerve agent in U.K. attack ID'ed

CHEMICAL WEAPONS | The nerve agent used against a former Russian spy and his daughter was a potent binary chemical weapon, U.K. Prime Minister Theresa May told Parliament this week. Binary nerve agents are comprised of organophosphate compounds that are relatively safe to handle on their own but become highly toxic after being mixed together. May identified the weapon as Novichok, or "newcomer," a label for a once-secret class of nerve agents developed by the Soviet Union in the 1970s and 1980s. She did not reveal which Novichok was used in the 4 March attack in Salisbury, U.K., on Sergei and Yulia Skripal, who remain hospitalized, along with a responding police officer. Unlike other nerve agents that are liquids and vaporize, Novichok compounds—some of which may be the most toxic nerve agents ever created—are believed to exist in the form of an ultrafine powder, says Michelle Carlin, a forensic chemist at Northumbria University in Newcastle upon Tyne, U.K.

Science overhauled in China

RESEARCH POLICY | China this week unveiled a sweeping government overhaul that could have major implications for the government's role in science and public health. Chinese scientists worry about the plan's call to subsume the Natural Science Foundation of China—the nation's main source of peer-reviewed funding—under the science ministry, which has tended to take an opaque approach to doling out its research funds. Experts welcomed other aspects of the plan. It would shift responsibility for antismoking efforts from the industry ministry—which regulates China's tobacco monopoly—to the health ministry. That's "potentially a real breakthrough moment for tobacco control," says the World Health Organization's Angela Pratt in Manila. And the consolidation of policies on pollution and climate change under a new Ministry of Ecological Environment could lead to more "coherent policies and regulations," says Ranping Song of the World Resources Institute in Washington, D.C.

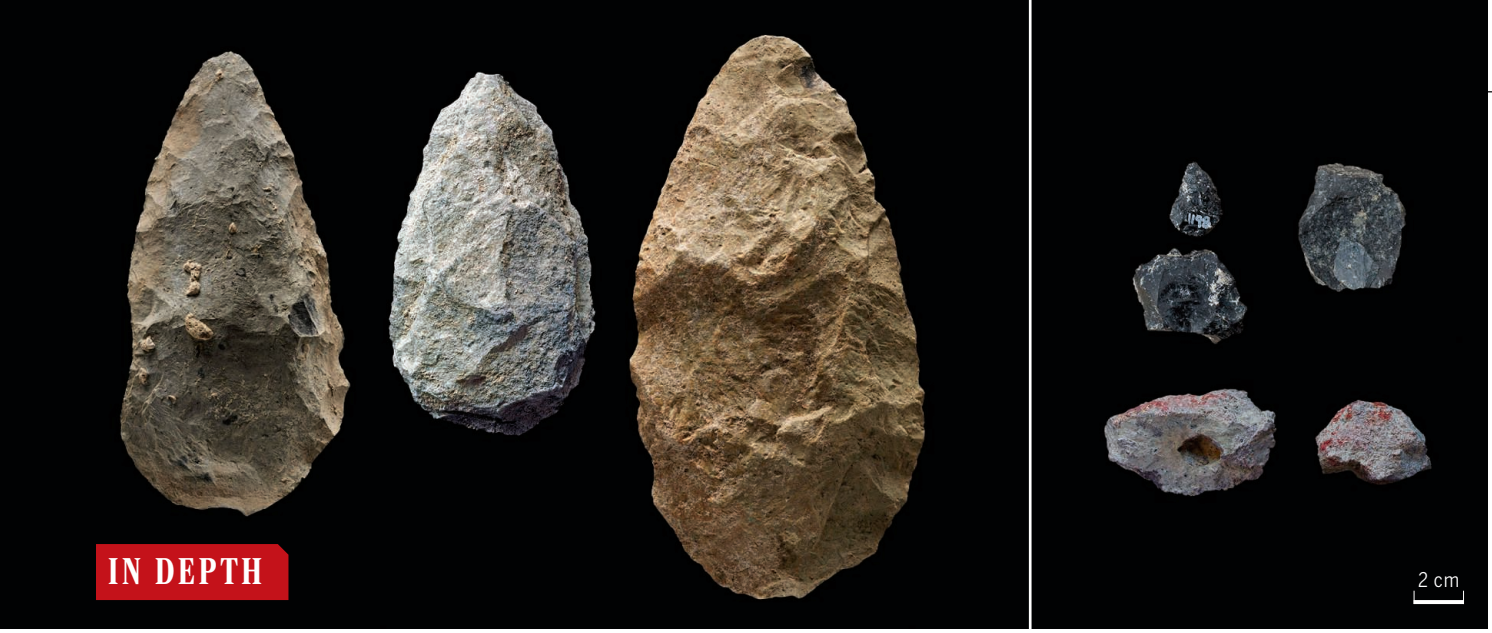
Prepping for disease X

PUBLIC HEALTH | A network of laboratories and clinics across Africa will receive a €10 million starting grant to help prepare for "disease X," the as-yet-unknown cause of the next major global pandemic. In February, the World Health Organization listed disease X as one of its R&D priorities. On 11 March, a consortium of 21 organizations from nine African and four European countries launched the African Coalition for Epidemic Research, Response and Training to coordinate clinical research, surveillance, diagnostics, and treatment in response to outbreaks. The group aims to bolster labs and clinics across the continent and coordinate rules and regulations across countries to remove some of the bureaucratic red tape that hampered the response to the Ebola outbreak in West Africa.

Volcanic winter spared humans

ARCHAEOLOGY | Early modern humans in South Africa appear to have survived just fine after a colossal volcanic eruption about 74,000 years ago spewed ash into the atmosphere and triggered a global "volcanic winter." Some researchers have hypothesized that the global effects nearly wiped out early populations of *Homo sapiens*. But new evidence of their apparent resilience comes in a study, published this week in *Nature*, that analyzed the presence of cryptotephra, tiny particles of glass spewed by the volcano, at archaeological sites on the coast of South Africa. Researchers identified a layer of volcanic glass, chemically traced to the explosion of Toba on the Indonesian island of Sumatra, that was probably deposited within a month of the explosion. Tellingly, artifacts found just below and directly above this layer showed no gap in human use of the site. The glass offers a powerful new marker to precisely date and correlate ancient sites across the globe.

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



IN DEPTH

ARCHAEOLOGY

Complex behavior arose at dawn of humans

Advanced stone tools, pigment, and extensive networks emerged as environment changed

By **Ann Gibbons**

More than 320,000 years ago in the Rift Valley of Africa, some early innovators adopted a new technology: They eschewed the clunky, palm-size stone hand axes that their ancestors had used for more than a million years in favor of a sleek new toolkit. Like new generations of cellphones today, their Middle Stone Age (MSA) blades and points were smaller and more precise than the old so-called Acheulean hand axes and scrapers.

These toolmakers in the Olorgesailie Basin in Kenya chose as raw materials shiny black obsidian and white and green chert, rocks they had to get from distant sources or through trade networks. In another first, they chiseled red and black rocks, probably to use as crayons to color their bodies or spears—an early sign of symbolic behavior. "This is indicative of a gear change in behavior, toolmaking, and material culture," says evolutionary psychologist Robin Dunbar at the University of Oxford in the United Kingdom, who studies social networks.

A trio of papers released online in *Science* today documents this remarkable technological transition. Although other sites have yielded MSA tools, the new, securely dated chronology nudges the transition back by at least 20,000 years, matching when our species, *Homo sapiens*, is now thought to have emerged. By analyzing artifacts over time at one site, the papers also show that these behaviors developed as climate swings intensified, supporting the idea that environmental variability drove innovation.

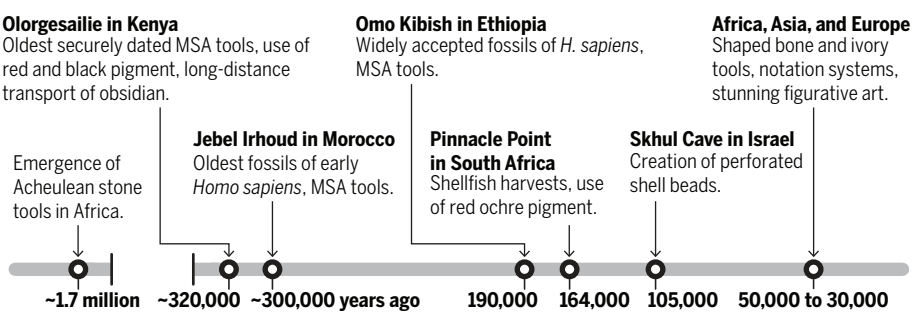
A team led by paleoanthropologists Rick Potts of the Smithsonian Institution's National Museum of Natural History and Alison Brooks of The George Washington University, both in Washington, D.C., gathered artifacts from sediments spanning 1.2 million years at Olorgesailie. Unfortunately, ancient erosion stripped away layers between 499,000 and 320,000 years ago, erasing the time when the MSA was probably invented. But by looking at more than 20,000 animal fossils associated with tens of thousands of stone tools, and multiple clues to the ancient environment, the team provides a detailed picture of life before and after the transition to the MSA. They pin down the timing with what geochronologist Michael Storey of the Natural History Museum of Denmark in Copenhagen calls "very impressive" dating.

About 900,000 years ago, a skull cap shows that the human ancestor *H. erectus* lived at Olorgesailie and used big Acheulean hand axes and scrapers to butcher meat. About 800,000 years ago, the climate began fluctuating more intensely from wet to dry, and the environment became more arid and grassy. At about 615,000 years ago, early humans began to make smaller Acheulean tools they could carry farther, and to more carefully select basalt as raw material, perhaps shifting hunting tactics in a changing environment, Potts says. The last hand ax at the site dates to 499,000 years ago, before the gap.

By the time the archaeological record restarted at 320,000 years ago, the Acheulean tools were gone and the basin had changed dramatically. The wet-dry cycle was even

The roots of modern human behavior

Many complex technologies and signs of symbolic behavior, such as sophisticated Middle Stone Age (MSA) tools and pigment use, appeared first in Africa and then in fits and starts around the world.



CREDITS: (PHOTOS) HUMAN ORIGINS PROGRAM, SMITHSONIAN INSTITUTION; (TIMELINE) N. DESAI/*SCIENCE*

At Olorgesailie in Kenya, big hand axes (left) gave way to smaller, more precise blades and points (right).

more extreme. More than 80% of mammal species had vanished and new kinds of elephants, pigs, foxes, and springboks gathered at tree-lined streams. MSA tools—relatively sophisticated blades and points that would have been hafted onto spears—were plentiful.

The site yielded no human fossils in this key time frame, so researchers can't be sure who the new toolmakers were. But discoveries elsewhere offer a strong hint. For years archaeologists had thought the MSA tools were too old to have been made by our species. Then, last year, fossils resembling *H. sapiens* were found near MSA tools and dated to nearly 300,000 years ago at Jebel Irhoud in Morocco (*Science*, 9 June 2017, p. 993)—timing that fits the Olorgesailie chronology.

Features of the MSA tools also suggest they were the handiwork of sophisticated humans. The toolmakers were highly selective about their raw materials, importing obsidian from up to 90 kilometers away. Such far-flung connections are a "hallmark of human social organization, and an important buffer in forager societies," whose members may move to distant places in hard times, Brooks says. The tools also are smaller, more precise, and more uniform in shape than Acheulean tools. They represent a milestone in abstract thinking: A hand ax preserves the shape of the original lump of rock, but creating a blade from an already prepared core forces toolmakers to visualize its shape in advance, Brooks says.

The team also found chunks of black rock and pieces of red ochre that had been punctured by sharp stone chisels. They propose both were used as pigments to create marks of individual or group identity, suggesting a high degree of social organization.

Expanded social networks are "a key unique feature of modern humans," says archaeologist Curtis Marean of Arizona State University in Tempe, although he suspects there will be some debate about the evidence for long-distance networks, noting that most of the stone came from just 25 to 50 kilometers from the site.

By combining artifacts with environmental data, the papers help explain what drove the MSA, says archaeologist Shannon McPherron of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. "They were able to use the long time-sequence at Olorgesailie to demonstrate how changes in the environment and the fauna correspond to the shift to the MSA."

But these MSA toolmakers hadn't developed the full package of sophisticated behavior, Dunbar cautions. "It's at the bottom of the scale of modern behavior," he says. "We're not talking about Salvador Dalí." ■

PHOTO: REUTERS/STRINGER

INFECTIOUS DISEASE

Nigeria hit by unprecedented Lassa fever outbreak

As efforts to contain it mount, researchers are racing to find out what is driving this year's surge in cases and deaths

By **Leslie Roberts**

By early January, it was clear something "really, really extraordinary" was going on in Nigeria, says Lorenzo Pomarico of the Alliance for International Medical Action (ALIMA). Cases of Lassa fever, a rare viral hemorrhagic disease, were skyrocketing across the country—more were recorded in the first 2 months of 2018 than in any previous year. Unprepared for a disease that has no vaccines or drugs and kills 20% to 30% of those it sickens, eight health care workers were infected early on and three died. "Something was going very wrong with the outbreak," Pomarico says.

Since then, the situation has only gotten worse. The rodent-borne disease is endemic in Nigeria and several other West African countries, fluctuating with the seasons and usually causing "a trickle" of cases a year, says Chikwe Ihekweazu, who heads the Nigeria Centre for Disease Control in Abuja. But as of 11 March, 365 cases and 114 deaths had been confirmed across 19 states, with many more suspected. Ihekweazu says the record-setting figures are sure to be underestimates, because the disease is maddeningly hard to diagnose, and many cases go unreported.

Already, Nigeria's fragile health care system is overwhelmed. The one dedicated Lassa fever ward in the country at Irrua Specialist Teaching Hospital in Edo state has just 24 beds. Without access to proper training and personal protective equipment, health care workers continue to become infected—by now 16 cases have been reported, with one additional death.

As the government and its international partners scramble to set up isolation wards and deliver protective gear to health workers, researchers on three continents are racing to figure out what is driving the unprecedented outbreak. Is it simply better disease surveillance in the wake of Ebola, the similar but more deadly disease that began its rampage across West Africa in 2014? Has the virus changed in some way? Are there more of the rats that carry it, or are more of them infected? Or is another rodent capable of spreading the virus as well?

"There are lots of possible explanations," says Stephan Günther, who heads the virology department at the Bernhard Nocht Institute of Tropical Medicine in Hamburg, Germany, and whose team has long collaborated with Nigerian researchers. Considering how lethal Lassa fever is, shockingly little is known about it, he says. "We don't

This year, the rats that carry Lassa fever may be more numerous, or more likely to harbor the virus.

know why people die. We don't know about the pathophysiology of the disease. We don't know the point of no return."

That could be beginning to change. In 2016, the World Health Organization added Lassa fever to its new list of priority pathogens of epidemic potential, calling for more research. And last week, the recently created Coalition for Epidemic Preparedness Innovations, known as CEPI, awarded its first grant for development of a Lassa fever vaccine to Themis Bioscience in Vienna.

Lassa fever was discovered in 1969, when two missionary nurses died of a mysterious disease in the remote town of Lassa in Borno state in northeastern Nigeria. When a third nurse fell ill, she was evacuated to a hospital in New York City—along with a thermos full of blood and other samples from all three nurses, bound for Yale University's then-new Arbovirus Research Unit. There, a team led by Jordi Casals-Ariet isolated a novel virus from the samples. (He, too, almost died in the process, saved only by an infusion of antibody-rich plasma from the third nurse, who recovered.)

The cause is now known to be an arenavirus, one of a class of rodent-borne pathogens. Its natural reservoir is a multimammate rat, so-called for its rows of mammary glands, that is ubiquitous across West Africa. Cases peak in the dry season, when farmers burn the bushes in preparation for spring planting and rats scurry into houses in search of food. The rodents shed the virus in their urine and droppings, and people contract it by touching contaminated surfaces, inhaling viral particles, or eating contaminated food (including the rats). Like Ebola, the virus can also be spread through contact with bodily fluids from an infected person. Such human-to-human transmission is thought to be rare for Lassa, unlike Ebola, except in hospital settings without proper infection control. However, "The real rate of human-to-human transmission is unknown," says Augustin Augier, secretary general of ALIMA in Paris, which has just launched a Lassa fever research program with the French medical institution INSERM.

No one knows the true incidence of the disease. "Most cases we have found are in places where there are hospitals and labs," Günther says. "There is good reason to assume there are cases that are being overlooked." And because the rat vector lives across a broad swath of the continent, the disease might also be endemic, but unrecognized, outside of West Africa, where it could be responsible for undiagnosed fevers.

Initial symptoms are easily mistaken for malaria or typhoid fever—body aches, sore throat, fever, nausea, diarrhea—before the disease progresses to organ failure, shock, and sometimes internal hemorrhaging. By the time doctors suspect Lassa fever, it's often too late to save the patient. There is no rapid test; accurately diagnosing the disease requires a real-time polymerase chain reaction technique, but just three labs in Nigeria have that capability.

For now, the only treatment is a nonspecific antiviral drug, ribavirin. If it's administered during the first 6 days of the illness, it seems to improve a patient's prognosis, but "no one arrives before day 7," Augier says. Nor is everyone convinced that ribavirin works in Lassa fever, as the only data come from the 1980s, Augier says.

Several potential drugs are on the horizon, in addition to the vaccine. Christian Happi at Redeemer's University in Ede, Nigeria, and the Irrua Specialist Teaching Hospital is developing a rapid diagnostic test with colleagues at Tulane University in New Orleans, Louisiana; the Broad Institute in Cambridge, Massachusetts; and Zalgen, a company in Germantown, Maryland. Happi's group and its partners are also sequencing the virus "around the clock," he says, and trying to figure out whether the genetic changes they have seen could have made it more transmissible or virulent.

> ***"I used to scream and scream that Lassa is important, but no one listened. ... Lassa fever is a disease of the poor."***
>
> **Christian Happi**, Redeemer's University

For Happi, who diagnosed Sierra Leone's first case of Ebola, the new attention to Lassa hasn't come a moment too soon. "I used to scream and scream that Lassa is important, but no one listened," he says. "I wrote so many grants" that were turned down. "Lassa fever is a disease of the poor ... it is confined to a part of West Africa, and it is not viewed as a global threat."

As research scales up, the government and its partners are focusing on training health care workers and providing the basics needed for infection control, as well as educating a frightened public about safety precautions.

Pomarico, who is leading ALIMA's emergency response to the outbreak in the two hardest hit states of Edo and Ondo, hopes cases will subside with the rains and cooler weather, as they usually do. "But this year is different. We are bracing for worst and preparing for the worst." ■

QUANTUM PHYSICS

Vibrations used to talk to quantum circuits

Sound waves could supplant microwaves in controlling quantum computers

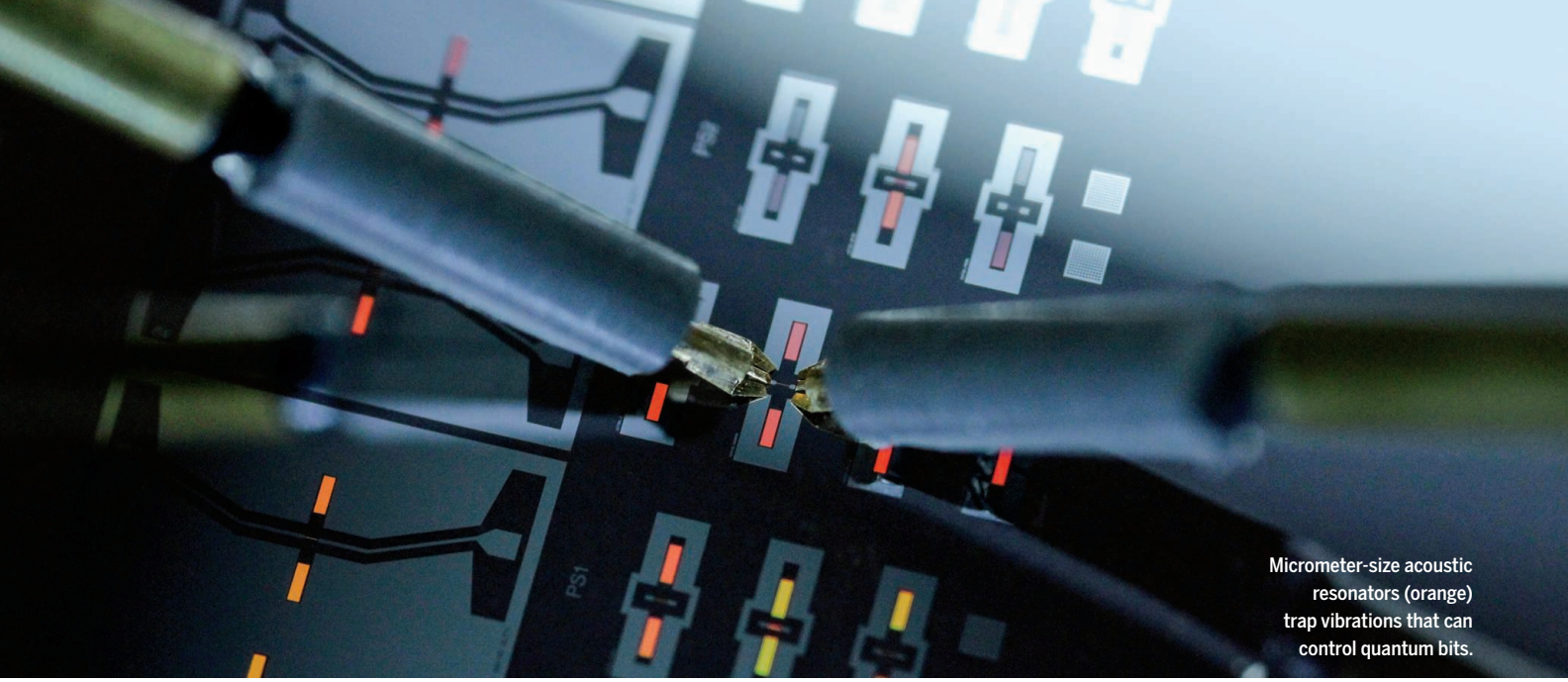
By **Adrian Cho**, *in Los Angeles, California*

For the moment, microwave photons are the keys to many quantum computers: Physicists use them to program, read out, and otherwise manipulate the machines' quantum bits. But microwave technology is bulky, and its quantum states don't last very long. Now, several groups are exploring a new way to talk to a quantum computer: with tiny vibrations, normally carriers of pesky heat and noise.

The budding discipline of quantum acoustics could shake up embryonic quantum computers by miniaturizing technologies and producing longer-lasting quantum memories. "We're right on the cusp" of controlling quantum vibrations, says Andrew Cleland, a physicist at the University of Chicago in Illinois, whose group presented its latest work last week here at the annual March meeting of the American Physical Society.

Whereas an ordinary computer flips bits that can be set to either zero or one, a quantum computer uses qubits that can be set to zero, one, or, bizarrely, zero and one at the same time—potentially enabling huge boosts in speed. Companies such as Google and IBM are racing to demonstrate the superiority of quantum computers for certain tasks (*Science*, 2 December 2016, p. 1090), and many are betting on qubits made of superconducting metal circuits on chips.

To control or read out a superconducting qubit, researchers make it interact with a microwave resonator—typically a strip of metal on the qubit chip or a finger-size cavity surrounding it—which rings with microwave photons the way an organ pipe rings with sound. By adjusting the energy of the qubit, researchers can shuttle its quantum states into the resonator, so that a zero-and-one state of the qubit can be stored as a state of the resonator in which a photon is both present and absent. But some physicists



Micrometer-size acoustic resonators (orange) trap vibrations that can control quantum bits.

see advantages to replacing the microwave resonator with a mechanical one that rings with quantized vibrations, or phonons.

That effort may seem daft, as such vibrations constitute heat, which obliterates delicate quantum states. But when working at temperatures near absolute zero, a well-designed acoustic resonator could ring longer than a microwave one does, enabling it to act as a sort of quantum memory, says Robert Schoelkopf, a physicist at Yale University. The vibrations also have wavelengths less than a thousandth as long as microwaves of the same frequency, so the resonators can be far more compact, he says.

First, physicists must learn to control quantum vibrations. They took a first step in 2010, when Cleland, then at the University of California, Santa Barbara (UCSB), siphoned every phonon out of an oscillating cantilever etched from aluminum nitride, leaving it in its least energetic quantum ground state. However, that simple quantum state persisted for just 5 nanoseconds, too little time to put the device into more complex quantum states of motion.

To push further, several groups are manipulating ripples called surface acoustic waves (SAWs), which travel along a material's surface. On top of a microchip, the researchers etch two gratings of metal stripes just micrometers apart. In the gap between the gratings, the researchers trigger a wave by applying a voltage to a comb-shaped device called a transducer, which causes the material to contract. The gratings act as mirrors, reflecting SAWs of particular wavelengths back and forth so that they resonate in the gap. And by connecting the transducer to a superconducting qubit, researchers link its quantum state to the SAWs.

Using that approach, Cleland and Kevin Satzinger, a UCSB graduate student, fashioned a resonator on a lithium niobate chip that rang for up to 150 nanoseconds. They showed they could create any desired combination of zero and one phonons in the resonator, Satzinger told the meeting. "We can watch the energy going back and forth" between qubit and cavity, he says.

Researchers in Schoelkopf's group are focusing not on waves trapped on a chip's surface, but on vibrations traveling through the chip's bulk material. They exploit vibrations that can bounce between the upper and lower surfaces of the half-millimeter-thick chip beneath their qubit.

Using that geometry, the researchers kept vibrations in their sapphire chip ringing for up to 60 microseconds, Yale's Yiwen Chu told the meeting. Moreover, the researchers could feed up to seven quanta of vibration one by one into the resonator, she reported. Making more complex quantum states is "really the next step," Chu says. For example, she says, they might try to put the resonator into a Schrödinger cat state, in which it would contain a macroscopic sound wave, comprising many vibrational quanta, and at the same time be devoid of vibrations.

Acoustic resonators could offer more flexibility in quantum circuit designs. In some circuits, multiple qubits are linked to the same microwave resonator, which acts as a conduit for the qubit interactions. But most microwave cavities can only host photons of a single frequency. In that case, all the qubits must interact with one another in an interconnected tangle, says Konrad Lehnert, a physicist at JILA, an institute run jointly by the University of Colorado in Boulder and the National Institute of Standards and Technology.

In contrast, acoustic resonators can enable qubits to interact with vibrations of a few different, closely spaced frequencies. That should make it possible to tailor the interactions among the qubits, so that, for example, only nearest neighbors interact—desirable for modeling certain abstract quantum systems, Lehnert says. He and his colleagues have taken a step toward such control, JILA's Bradley Moores told the meeting, by showing that they could simultaneously couple a single qubit to SAWs of several frequencies.

Quantum acoustics might also help solve a major problem for emerging quantum technologies. Microwave cables can shuttle information within a quantum computer. To move it to other experiments or distant locations, however, those signals will likely need to be converted from microwaves to optical photons, which can travel great distances in optical fibers. Acoustic waves rippling at microwave frequencies have wavelengths similar to those of optical photons. So in principle, they could serve as a bridge to translate between the two, researchers say, although nobody yet knows exactly how to do it.

Acoustic resonators might even test the bounds of the quantum realm. Quantum theory allows tiny things like atoms or photons to be in two places at once, but nobody has ever seen such behavior in a macroscopic material object. Some theorists argue that a yet-unknown principle, perhaps involving gravity, would prevent it from happening for large objects. But Chu says it might be possible to make her group's sapphire chip simultaneously vibrate in opposite directions. That would put tens of micrograms of material in two slightly different places at the same time—and test whether quantum weirdness extends almost to the human scale. "You don't know until you try." ■

PHOTO: KEVIN SATZINGER, CLELAND LAB, UNIVERSITY OF CHICAGO

GERMANY

Max Planck Society, at a crossroads, seeks new leaders

A wave of retirements offers a chance to recruit female institute directors and open up new research avenues

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

Erin Schuman and her husband Gilles Laurent left their jobs at the California Institute of Technology in Pasadena a few years ago and moved their family, cat, and dog halfway across the globe to Frankfurt, Germany. It took some courage. She's American, he's French, and two of their daughters were in school; resettling in Europe was "certainly challenging," Schuman says.

But the move came with huge professional benefits. Schuman and her husband, both neuroscientists, moved to the Max Planck Institute for Brain Research as "directors," prestigious positions as the heads of large research groups that come with guaranteed funding until retirement and full academic freedom. They also had the opportunity to design a brand new lab. ("We worked with the architects to maximize the interactions and happiness of our colleagues," Schuman says.) The move turned out to be a "tremendously enriching experience."

Many scientists may soon have a similar chance. The Max Planck Society (MPG) recently took out ads in major scientific journals to recruit 20 new directors in fields ranging from astrophysics to terrestrial microbiology, in one of its biggest talent searches ever. Between now and 2030, roughly 200 of the 300 director posts at the 84 institutes will become vacant, says an MPG spokesperson. Many directors are coming up for retirement at the 20 institutes opened in the eastern part of the country after the German reunification in 1990, and so are many of their counterparts at older institutes in former West Germany.

For the society it marks a turning point: a chance to hire more women and more foreign researchers, and an opportunity to open up entirely new fields of research, because incoming directors can set their own course. (There are also plans to launch completely new institutes on cybersecurity and the origins of life.) But it's also a huge challenge. Finding directors is a lengthy process, and competition from other institutions seeking top talent—especially women—is fierce. And some argue that the society should use the opportunity to rethink some of its traditions.

Founded in 1948, MPG has a €1.8 billion annual budget and is Germany's leading research powerhouse. The Nature Publishing Index ranks it as the fourth largest contributor to high-quality research in the world; its researchers have won 18 Nobel Prizes.

Most of the institutes are led jointly by three to five directors who each also run their own research departments. As an MPG director, "You have an incredible liberty to research what you want, even changing your field if you like," says Christiane Nüsslein-Volhard, who won a Nobel Prize in Physiology or Medicine as a director at the Max Planck Institute for Developmental Biology in Tübingen in 1995. There are periodic evaluations, but a poor result means losing only a fraction of your funding, says Schuman, who previously held one of the plum positions in U.S. science: as an investigator funded by the Howard Hughes Medical Institute on a 5-year contract. "I did not realize how the renewal clock of 5 years dissuaded me from going for risky ideas until I became a [Max Planck] director," she says.

> ***"You have an incredible liberty to research what you want, even changing your field if you like."***
>
> **Christiane Nüsslein-Volhard**, Max Planck Institute for Developmental Biology

When a director leaves or retires, MPG doesn't just look to "fill a certain slot," says Jürgen Renn, director of the Max Planck Institute for the History of Science here. Instead it seeks excellent scientists and reshapes the institute's mission around them, sometimes altering it completely. When the directors of the Max Planck Institute of Economics in Jena left within a few years of each other, for instance, prominent economists proved hard to lure to that city. So MPG in 2014 refashioned it into the Max Planck Institute for the Science of Human History, which has since done pioneering work on the history of plague.

The hiring process is arduous. Traditionally, an institute's current directors identify potential candidates, who are then reviewed by outside referees. (The recent ads were a new addition.) Settling on a new director can take years. The society currently fills about 10 positions a year, so the rush of openings may be a challenge. And the "explosion of research at universities all around the globe" means that competition "is a lot more intense," says U.S. economist David Audretsch, a former director at the Jena institute.

Finding women is especially hard. The number of female directors has gone up from 4.5% in 2005 to about 15% today; MPG aims to reach 18% by 2020. (Among recent catches is gene-editing pioneer Emmanuelle Charpentier.) But a recent MPG report mentioned an international "bidding war" for senior female talent. To expand the pool, female scientists should receive much more support early in their careers, Renn says. One effort is MPG's Lise Meitner excellence program, launched late last year, which will establish up to 10 new research groups led by women every year. Recruiting directors at a younger age could also help, Schuman says: "I know this younger pool has lots of qualified and amazing women who can join our ranks."

Attracting international talent is another challenge. Under Germany's federal system, MPG institutes are spread around the country, from big cities like Berlin or Munich all the way to places such as Plön, a small lakeside resort in the north, that are less attractive to foreigners who want an international school or an airport close by.

The biggest drawback of MPG institutes is their "outdated" isolation from universities, says Thomas Südhof, a biochemist and Nobel laureate at Stanford University in Palo Alto, California, who was an MPG director in the 1990s but left after falling out with Hubert Markl, then MPG's president. "Science is a team effort and it's important who you meet in the hallway or in the cafeteria," Südhof says. "The Max Plank Society would profit enormously if they integrated their institutes directly into universities." Schuman says he has a point, but warns that integration might lead to a "class system" within universities that favors Max Planck appointees. MPG is trying another solution: bringing institutes in the same city together in campuslike clusters.

As to filling the wave of new director posts, "I think we should dare to be riskier," Schuman says. "We should not be afraid of failing, but rather expect that some failures will accompany our riskier choices, which could lead to game-changing discoveries." In other words, MPG may just need a bit more of the courage it takes for an individual scientist to pull up stakes and move their lab and family to another continent. ■

INFECTIOUS DISEASE

Concern as HIV prevention strategy languishes

Few countries promote prophylactic drug regimen, and many people shun it

By **Jon Cohen**, *in Boston*

There's no question that a simple regimen of a single daily pill can slash HIV infections in people at risk. But although millions of people around the world could benefit, only 200,000 are prescribed what's called pre-exposure prophylaxis, or PrEP, and 75% of them are in the United States. The disparity was the focus of anguished discussions last week at the largest annual U.S. HIV/AIDS conference. "I'm frustrated," says Linda-Gail Bekker, a researcher at the University of Cape Town in South Africa. "We have nearly 1700 infections every week in young women and girls, and this might be a way to throttle this."

PrEP, which uses antiretroviral (ARV) pills made for treatment as preventives, faces multiple obstacles, as Bekker and other discussed at the Conference on Retroviruses and Opportunistic Infections here. Many countries already struggle to provide ARVs to people who are infected and say they can't afford PrEP for the large uninfected populations who need it most. People at substantial risk of becoming infected by HIV may not know about PrEP or may reject it, as several presentations highlighted. And PrEP raises touchy issues including teen sexuality and the appearance of advocating risky behavior.

Yet PrEP works. Its effectiveness in several clinical trials led the U.S. Food and Drug Administration to approve preventive use of ARVs in 2012. The World Health Organization in 2015 recommended PrEP for everyone at "substantial risk" of becoming infected—a total of tens of millions of people, to judge from the nearly 2 million new HIV infections in 2016. In some countries, including the United States, the targeted groups are "key populations" such as men who have sex with men (MSM) and sex workers. In sub-Saharan Africa, where most people become infected from heterosexual sex and many have no "high-risk" behavior, large swaths of the population are candidates for PrEP.

Two reports at the meeting underscored its power. In San Francisco, California, the number of people using PrEP more than tripled from 2013 to 2016, to 12,300. Researchers found that the city's new HIV diagnoses dropped by 43% during that period, although they note that it isn't easy to disentangle the impact of PrEP from other factors. In Australia, the state of New South Wales launched a PrEP project that began enrolling MSM in 2016. As Andrew Grulich, an epidemiologist at the University of New South Wales in Sydney, reported, more than 9000 MSM have now started PrEP. In the first half of 2017, Grulich says the team saw a "very substantial decline" in new infections—a drop of 35%.

Only one hard-hit country in sub-Saharan Africa aggressively promotes PrEP—Kenya, where 13,000 people are now using it. South Africa has an HIV-infected population estimated at 7.2 million, more than any other country, and yet only 5000 or so uninfected people are using PrEP. At a cost of a few hundred dollars a year per person for drugs and monitoring, PrEP would add to the huge economic burden that the country faces in trying to treat all infected people, Bekker acknowledges. But she says the government is also wed to a flawed analysis that suggested PrEP wouldn't be cost effective outside the highest risk groups such as sex workers and MSM.

Even the United States was slow to adopt PrEP, and several studies at the meeting explored why. Many people do not recognize that they are at high risk of infection. Others worry about side effects—long-term use of the drugs can cause bone and kidney problems. Some simply prefer condoms or don't want to take a daily pill. A study in four southern U.S. cities found that of 72 women who were deemed "PrEP eligible," only four had ever heard of it. Another study of 201 transgender black and Latina women—a population that has one of the highest new infection rates in the country—in Washington, D.C., and Baltimore, Maryland, found that 86% knew about PrEP. But only 39% had tried it, in part because of unfounded concerns about interactions with hormones they were taking.

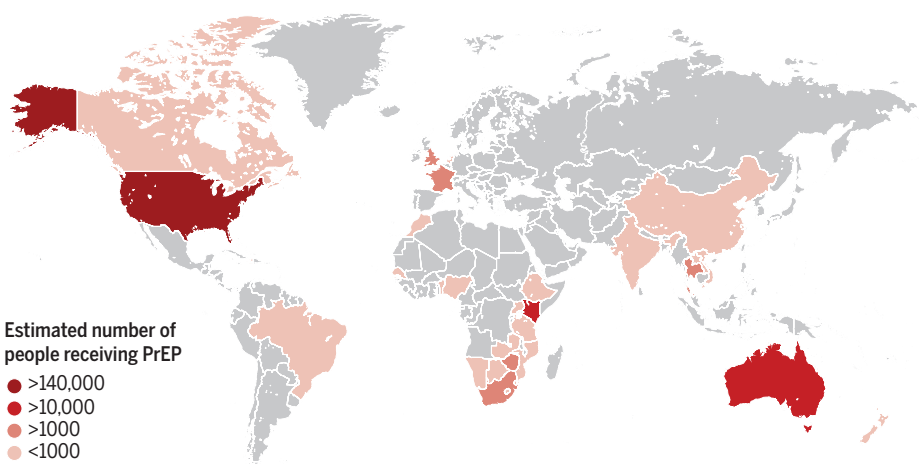
Many researchers and affected communities also worry that PrEP could backfire if it increases risk-taking behavior such as having more partners or not using condoms. But studies have not supported these concerns.

François Venter of the University of the Witwatersrand in Johannesburg, South Africa, who has tested PrEP in sex workers, says his government is squeamish about offering it to teens, who might benefit the most. Ideally, he says, PrEP would be promoted in school sex education classes. But he can't see that happening in South Africa's "inherently conservative society" any time soon. "Everyone wrings their hands and whinges out about young girls, but we can't get into schools," Venter says. "It's a political and programmatic hot potato." Ultimately, he adds, PrEP should be treated like contraception. "You should be able to walk in and buy it over the counter."

Nelly Mugo, an OB-GYN from the Kenya Medical Research Institute in Nairobi, stresses the need to listen to and address the concerns of PrEP skeptics. "We shouldn't be discouraged by low initial uptake," Mugo says. "Maybe we need the slow period for people to be certain the products are good." ■

Poorly PrEPared

Only the United States, Australia, and Kenya aggressively promote pre-exposure prophylaxis (PrEP).



CREDITS: (GRAPHIC) N. CARY/*SCIENCE*; (DATA) AVAC

FEATURES

MASTER PLANNER

China's revered rocket scientist, Qian Xuesen, set in motion a system for engineers to control Chinese society

By **Mara Hvistendahl**, *in Shanghai, China*

PHOTO: IMAGINECHINA VIA AP IMAGES

It's rare that a scientist becomes a folk hero. But in China, Qian Xuesen draws crowds almost a decade after his death. On a Saturday morning in a three-story museum here, tourists admire Qian's faded green sofa set, the worn leather briefcase he carried for 4 decades, and a picture of him shaking hands with opera star Luciano Pavarotti. They file past a relic from a turning point in Qian's life—and in China's rise as a superpower: a framed ticket from his 1955 voyage from San Francisco, California, to Hong Kong in China aboard the *SS President Cleveland*. Once a professor at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California, he had been accused of having communist sympathies in the heat of the Red Scare and placed under virtual house arrest. Upon his release, he and his family set sail for his motherland.

After arriving in China, Qian went on to spearhead the rapid ascent of the country's nuclear weapons program, an achievement that explains some of the adulation. But his legacy is still unfolding in a second area that could have great consequences for China—and for the world. Qian, who died in 2009 at the age of 97, helped lay the groundwork for China's modern surveillance state.

Early in his career, he embraced systems engineering—an interdisciplinary field focused on understanding the general properties common to all physical and societal systems, and using that knowledge to exert control. By mapping a system's dynamics and constraints, including any feedback loops, systems theorists learn how to intervene in it and shape outcomes. Since the field's founding in the 1950s, systems approaches have been applied to areas as varied as biology and transportation infrastructure.

In the West, systems engineering's heyday has long passed. But in China, the discipline is deeply integrated into national planning. The city of Wuhan is preparing to host in August the International Conference on Control Science and Systems Engineering, which focuses on topics such as autonomous transportation and the "control analysis of social and human systems." Systems engineers have had a hand in projects as diverse as hydropower dam construction and China's social credit system, a vast effort aimed at using big data to track citizens' behavior. Systems theory "doesn't just solve natural sciences problems, social science problems, and engineering technology problems," explains Xue Huifeng, director of the China Aerospace Laboratory of Social System Engineering (CALSSE) and president of the China Academy of Aerospace Systems Science and Engineering in Beijing. "It also solves governance problems."

The field has resonated with Chinese President Xi Jinping, who in 2013 said that "comprehensively deepening reform is a complex systems engineering problem." So important is the discipline to the Chinese Communist Party that cadres in its Central Party School in Beijing are required to study it. By applying systems engineering to challenges such as maintaining social stability, the Chinese government aims to "not just understand reality or predict reality, but to control reality," says Rogier Creemers, a scholar of Chinese law at the Leiden University Institute for Area Studies in the Netherlands.

With the discipline now touted at the highest levels of government, Qian has been deified, with biographies, television segments, and symposiums regularly devoted to him. In the 1990s, the Chinese government even spearheaded a "learn from Qian Xuesen" movement. Popular discourse now acknowledges that modern China's first leader, Mao Zedong, "was a human being," says Zhichang Zhu, a systems scientist at the Xiamen University Malaysia in Sepang. "But to a circle of scientists in China, Qian Xuesen is now, in their mind, the new god."

The traditional, or "hard," brand of systems engineering that Qian pioneered has lately come under attack from Zhu and other scholars, both inside and outside

A bust of Qian Xuesen at a museum in Shanghai, China, devoted to the revered scientist. In addition to designing sophisticated weapons systems, Qian laid the groundwork for China's surveillance system.

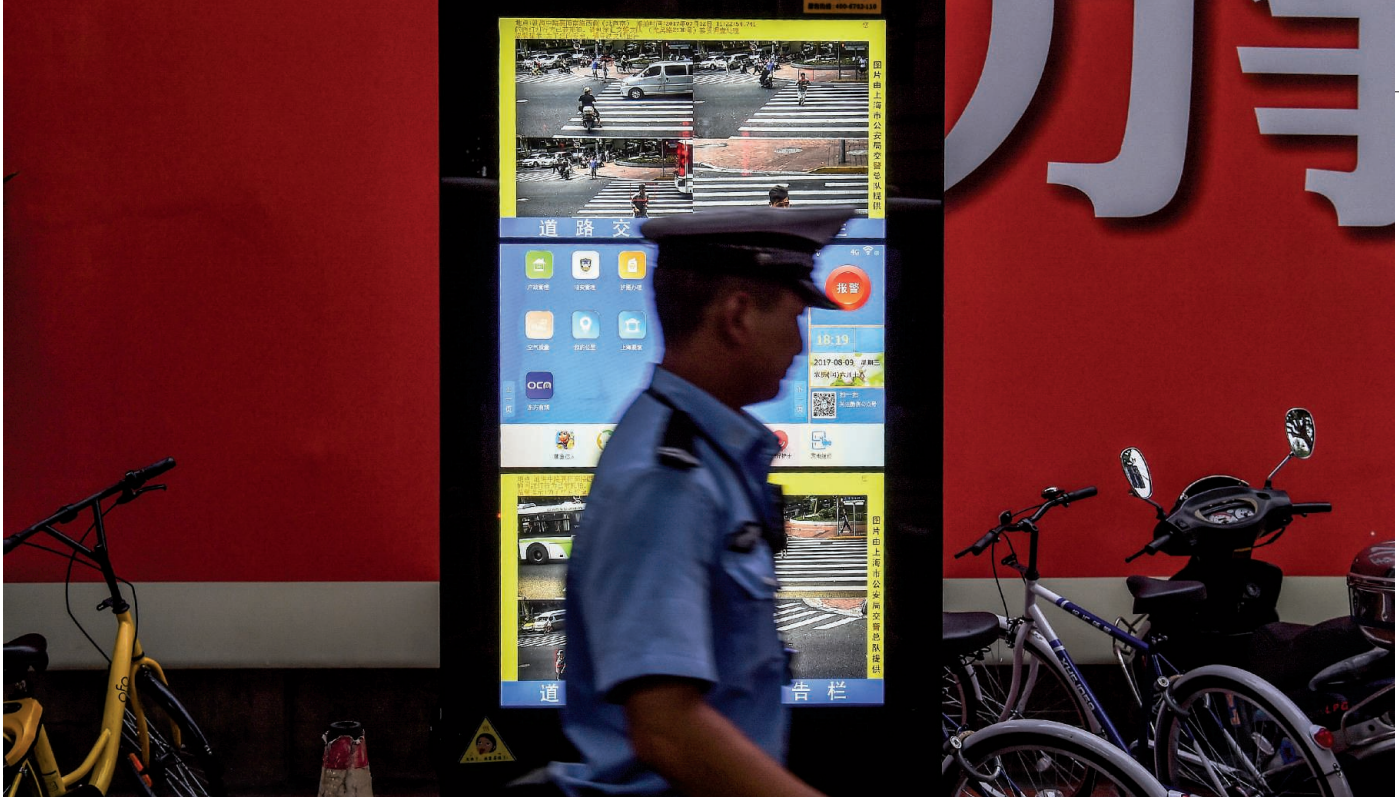
China. They contend that it discounts the experiences of everyday people affected by systems models and values state power above all else. They are trying to carve out an alternative vision for systems science, one less reliant on mathematical formulas and more attuned to civic participation. But that could prove an uphill battle in a country where maintaining stability trumps scholarly debate.

IN A BUILDING flanked by military guards, systems scientists from CALSSE sit around a large conference table, explaining to *Science* the complex diagrams behind their studies on controlling systems. The researchers have helped model resource management and other processes in smart cities powered by artificial intelligence. Xue, who oversees a project named for Qian at CALSSE, traces his work back to the U.S.-educated scientist. "You should not forget your original starting point," he says.

Qian was born in 1911 in Hangzhou, in eastern China. In 1935, a scholarship brought him to the Massachusetts Institute of Technology in Cambridge. He then went on to the California Institute of Technology in Pasadena, where he worked with the Hungarian mathematician Theodore von Kármán. When Von Kármán and others founded JPL in 1944 to develop rocket technology, Qian was given a security clearance and brought on to work on classified weapons research.

As the Red Scare took hold in the 1950s, the scientist came into the Federal Bureau of Investigation's crosshairs. His security clearance was revoked, and after years of bilateral negotiations, Qian was allowed to return to China. Back in Beijing, his previous experience designing complicated weapons systems and rockets for the United States became integral to China's budding efforts. On 16 October 1964, at 3 p.m. local time, China detonated its first atomic bomb. Xue says the program succeeded in part because Qian modeled a complex weapons system down to its most unpredictable parts. He automated China's weapons command and control system, enabling planners to direct the activities of thousands of people at once.

As Qian honed China's weapons systems, scientists in North America and Europe began applying systems approaches to intractable policy problems, modeling them as a collection of inputs and variables linked by direct or inverse relationships and feedback loops. In the 1960s, for example, school districts across California tried systems approaches. To help educators set budget priorities, multimillion-dollar data processing programs designed by JPL and the rocket manufacturer Aerojet General Corporation



A traffic officer walks past a facial recognition screen at an intersection in Shanghai, China. The Chinese government uses systems engineering to monitor crowd behavior.

PHOTO: AFP PHOTO/CHANDAN KHANNA

collated children's academic records, IQ scores, and attendance.

According to critics, such efforts wasted money that could have gone toward hiring teachers and reduced to rational analysis what should have been a complex political process. "When the policymakers came in and started asking questions, [they] were talking about variables that weren't in the models," says Gerald Midgley, a systems scientist at the University of Hull in the United Kingdom. California's data-driven approach to education was eventually scrapped. Those and other embarrassments brought the field into disrepute in the West.

BACK IN CHINA, though, the notion that scientists could neatly model societal endeavors resonated with leaders reared on central planning. An early major contribution of social systems scientists occurred in the late 1970s, when Qian's protégé, missile scientist Song Jian, led a team whose computer-generated projections showed China's population rising to 4 billion by 2080 (*Science*, 17 September 2010, p. 1460). That work helped justify extreme restrictions on births after the government implemented the one-child policy in 1980.

Soon after, systems scientists began assessing the feasibility of building the titanic Three Gorges Dam on the Yangtze River. The goal was to determine the optimal dam height and water level in the reservoir, balancing the demands of power generation with other factors, including the massive project's negative impacts. One group working on the project took stock of 14 "subsystems," including geology, ecology, and human migration. The researchers then analyzed how various water levels would affect outputs such as seismic activity or the number of people forced to relocate. Ultimately, the group arrived at an ideal water impoundment level, 175 meters. The dam's operator hewed to that advice, raising water levels to 175 meters by 2012.

As with the one-child policy, though, the systems scientists entrusted with studying the Three Gorges Dam devoted little time to consulting people whom the project would affect most. (Dam building and the reservoir that formed behind the structure displaced 1.3 million people in southwestern China.) Because the dam's construction was a foregone conclusion, the feasibility study was limited to outcomes that reinforced government plans. Researchers in China often approach megaprojects like Three Gorges "from the perspective of how to successfully implement the project whose execution has already been decided politically," says Yoshiteru Nakamori, a systems scientist and former dean of the School of Knowledge Science at Japan Advanced Institute of Science and Technology in Nomi.

More recently, the involvement of China's systems scientists in designing the country's digital infrastructure has raised similar questions about whether the scientists are aiding the state at the expense of the public. Take China's smart cities initiative. The Chinese government claims to have wired hundreds of cities with sensors that collect data on topics including city service usage and crime. At the opening ceremony of China's 19th Party Congress last fall, Xi said smart cities were part of a "deep integration of the internet, big data, and artificial intelligence with the real economy."

The initiative, which has received funding from the United Nations Development Programme, has benign components. Xue and colleagues, for example, are working on how smart cities can manage water resources. In Guangdong province, the researchers are evaluating how to develop a standardized approach for monitoring water use that might be extended to other smart cities.

But Xue says that smart cities are as much about preserving societal stability as streamlining transportation flows and mitigating air pollution. Samantha Hoffman, a consultant with the International Institute for Strategic Studies in London, says the program is tied to long-standing efforts to build a digital surveillance infrastructure and is "specifically there for social control reasons" (*Science*, 9 February, p. 628). The smart cities initiative builds on 1990s systems engineering projects—the "golden" projects—aimed at dividing cities into geographic grids for monitoring, she adds.

Layered onto the smart cities project is another systems engineering effort: China's social credit system. In 2014, the country's State Council outlined a plan to compile data on individuals, government officials, and companies into a nationwide tracking system by 2020. The goal is to shape behavior by using a mixture of carrots and sticks. In some citywide and commercial pilot projects already underway, individu-

als can be dinged for transgressions such as spreading rumors online. People who receive poor marks in the national system may eventually be barred from travel and denied access to social services, according to government documents.

Civil liberties groups charge that the system will deepen monitoring of the citizenry, especially if combined with the Chinese state's growing biometrics capabilities (*Science*, 21 September 2012, p. 1448). Social credit is aimed at "further tightening the web of social control," says Maya Wang, a researcher with Human Rights Watch in Hong Kong. (The Chinese government maintains that the system is about building trust and accountability as well as helping law enforcement identify criminals.)

Government documents refer to the social credit system as a "social systems engineering project." Details about which systems engineers consulted on the project are scant. But one theory that may have proved useful is Qian's "open complex giant system," Zhu says. A quarter-century ago, Qian proposed that society is a system comprising millions of subsystems: individual persons, in human parlance. Maintaining control in such a system is challenging because people have diverse backgrounds, hold a broad spectrum of opinions, and communicate using a variety of media, he wrote in 1993 in the *Journal of Systems Engineering and Electronics*. His answer sounds like an early road map for the social credit system: to use then-embryonic tools such as artificial intelligence to collect and synthesize reams of data.

According to published papers, China's hard systems scientists also use approaches derived from Qian's work to monitor public opinion and gauge crowd behavior. And systems science approaches are on display in Xinjiang, a region in northwest China with a high percentage of Muslims. According to Human Rights Watch, Xinjiang's public security bureau is aggregating data from sources such as closed-circuit TV cameras, security checkpoints, and residents' networked devices. Authorities then use a form of systems analysis adapted from People's Liberation Army doctrine to flag people seen as potentially disruptive.

AFTER SYSTEMS ENGINEERING fell from grace in the West, researchers spearheaded a fundamentally new approach. In 1981, Peter Checkland of Lancaster University in the United Kingdom called for a "soft" systems science that valued input from stakeholders over mathematical modeling. In Checkland's vision, experts exist not to impose their values, but instead to learn from people involved in the problem at hand. When Zhu left China in 1988 to earn a master's degree at the University of Hull, that was the approach he encountered. Zhu studied there under Midgley, just as systems engineering was transforming in the West.

A few years later, Zhu began collaborating with Gu Jifa in Beijing, who had worked on the Three Gorges Dam assessment and become head of the Systems Engineering Society of China. Zhu had come to see Qian's brand of systems science as "brain without soul." Gu, meanwhile, had learned firsthand the importance of what he called *renli*, or human relations, in shaping project outcomes. For example, his recommendations for an urban development plan in Beijing were not adopted because his team neglected to engage key stakeholders. Hard systems engineering worked well for rocket science, but not for more complex social problems, Gu says: "We realized we needed to change our approach." He felt strongly that any methods used in China had to be grounded in Chinese culture.

The duo came up with what it called the WSR approach: It integrated *wuli*, an investigation of facts and future scenarios; *shili*, the mathematical and conceptual models used to organize systems; and *renli*. Though influenced by U.K. systems thinking, the approach was decidedly eastern, its precepts inspired by the emphasis on social relationships in Chinese culture. Instead of shunning mathematical approaches, WSR tried to integrate them with softer inquiries, such as taking stock of what groups a project would benefit or harm. WSR has since been used to calculate wait times for large events in China and to determine how China's universities perform, among other projects.

Despite the efforts of Gu and others, systems science in China today remains rooted in the hard systems engineering approaches that rocket scientists pioneered decades ago. That emphasis is apparent at the cinder block apartment in Beijing where Qian lived for nearly a half-century, which is now an unofficial museum for state visitors.

China's systems engineers helped determine the optimal water level for the controversial Three Gorges Dam.

PHOTO: CHINAFOTOPRESS VIA GETTY IMAGES

Giving a tour of the apartment, Qian Yonggang, Xuesen's 69-year-old son, gestures to a living room decorated with blond wood accents. "Many Chinese leaders have sat here," he says. He moves on to the bedroom, indicating the twin bed in which his father spent his final days. Hanging above it is a framed photo of a somber-looking Qian Xuesen.

Zhu contends that the time has come to bring the god down to Earth. He recently wrote that systems science in China is "under a rationalistic grip, with the 'scientific' leg long and the democratic leg short." Zhu says he has no doubt that systems scientists can make projects such as the social credit system more effective. However, he cautions, "Systems approaches should not be just a convenient tool in the expert's hands for realizing the party's wills. They should be a powerful weapon in people's hands for building a fair, just, prosperous society." ■

PERSPECTIVES

PHYSIOLOGY

Rhythms: The dark side meets the light

Researchers examine the 24-hour biological clock in daytime-active baboons

By **Arthur Millius**[1] *and* **Hiroki Ueda**[1,2]

An internal biological rhythm, the circadian clock—which can be measured by changes in rhythmic gene expression, cellular activity, or physiological behavior—enables an organism to anticipate daily cyclic changes in the environment. For years, researchers have examined mammalian behavioral and physiological rhythms primarily in nocturnal species such as mice and rats; but in what ways are the genes, neural circuits, and rhythms of nocturnal animals related to those of humans, who (mostly) operate during the day? On page 1232 of this issue, Mure *et al.* (*1*) examine our close, evolutionarily related cousin—the wild olive baboon (*Papio anubis*) (see the photo)—to understand the internal clock of a diurnal primate, which has implications for understanding human biological rhythms and chronotherapy (treatment with drugs based on the time of day).

PHOTO: MARC GUITARD/GETTY IMAGES

The internal oscillations of the wild olive baboon may provide insights into human biological rhythms.

A vast empire of knowledge on mammalian biological rhythms is built primarily on research on nocturnal rodents, but how does this information translate to animals that are active during the day? Nocturnal and diurnal rodents both have similar brain tissues, neurotransmitters, and neuronal pacemaker cells (*2*–*4*), and diurnal rodents also have rhythms in a few core clock genes and proteins (*5*, *6*). Researchers have sampled various tissues throughout the day in a nocturnal mouse and have reported that about half of all genes have oscillatory expression (*7*). This suggests that widespread rhythmic changes in gene expression accompany daily changes in activity, from getting sleepy to feeling hungry. However, there is no corresponding timetable of gene expression in neuronal and peripheral tissues of a diurnal organism, or one closely related to humans, so that we can begin to draw connections between the timing of gene expression, cellular function, and behavior.

Mure *et al.* addressed this gap by examining physiology and gene expression every 2 hours over a 24-hour period in the daytime-active wild olive baboon and found that nearly 82% of transcripts had rhythmic oscillation in at least one tissue. This means that time-dependent gene expression is not a paradigm limited to a few core genes that regulate the circadian clock but rather is a pervasive element of primate physiology. In fact, about half of the most prevalent drugs in the United States target the products of rhythmic genes (*7*), which provides an opportunity to increase drug efficacy by timing delivery with their expression (chronotherapy).

Approximately 11,000 transcripts were expressed in all 64 sampled tissues, which the researchers called ubiquitously expressed genes, including many involved in basic cellular functions such as DNA repair, transcription, and protein homeostasis. Most of these ubiquitously expressed genes were rhythmic in at least one tissue, but there was little overlap in rhythmic genes between tissues, which suggests that tissue-specific mechanisms control oscillatory expression. For example, a gene that had rhythmic expression in the liver was constitutively expressed in the heart. Because ubiquitously expressed genes control fundamental biological processes, timing their expression can affect the overall function of a tissue. For example, diurnal regulation of exocytosis in the thyroid or adrenal glands may enable rhythmic release of endocrine factors, compared with other organs in which the timing of exocytosis is less important for function.

Only a few genes had rhythmic expression in more than one tissue—including many involved in DNA processing, metabolism, and circadian rhythms—but there were a few surprises. For example, the core clock transcriptional repressors cryptochrome 1 and cryptochrome 2 were expressed at unequal levels in different tissues. This suggests that each tissue has a distinct composition of circadian clock components, like an hourglass counting time in sand or salt. Nowhere was the disparity in tissue-specific expression of core circadian clock genes more apparent than between the expression of two transcriptional activators that form a heterodimer: BMAL1 (brain and muscle ARNT-like 1) and CLOCK (circadian locomotor output cycles protein kaput). The *BMAL1* gene was rhythmically expressed in more than 40 sampled tissues, whereas *CLOCK* was rhythmic in less than 10 tissues (mostly overlapping with *BMAL1*). In addition, the phase of *CLOCK* expression varied wildly over 12 hours, but the phase of *BMAL1* expression was tightly limited to a 4-hour period. If BMAL1 and CLOCK are like copilots driving circadian gene expression, then it seems BMAL1 is mostly at the wheel. The gene that was rhythmically expressed in the greatest number of tissues encodes a relative newcomer to the circadian field: CHRONO (ChIP-derived repressor of network oscillator), which counteracts BMAL1 and CLOCK transcriptional activation in a manner similar to but independent of the cryptochrome paralogs and down-regulates anti-inflammatory responses through glucocorticoids (*8*, *9*). The widespread rhythmicity of CHRONO expression suggests a more prominent role for circadian regulation of glucocorticoid signaling in diurnal organisms.

Is a diurnal animal simply a nocturnal one, with the gene expression phases flipped by 12 hours? When Mure *et al.* compared wild olive baboon transcriptional oscillations of a few core circadian clock components with those of nocturnal mice, the phases were indeed shifted by 12 hours in most tissues, with one notable exception. Phases in gene expression between mice and wild olive baboons were identical in a specialized brain structure called the suprachiasmatic nucleus (SCN). The SCN is called the central pacemaker because removal of the SCN abolishes an organism's circadian rhythms (*10*, *11*). Importantly, the SCN responds to light via light-sensitive cells in the retina, so it makes sense to assume that light synchronizes transcriptional oscillations in both nocturnal and diurnal animals by means of a similar mechanism. However, this finding means that nocturnal and diurnal behavior is regulated downstream of the SCN and is perhaps coordinated by individual organs and tissues according to their particular metabolic requirements. For example, the phases of rhythmic transcripts between mouse and baboon in the heart, which presumably has similar metabolic requirements during night and day, were more closely aligned than those in the cerebellum, where sleep and wake may have a stronger impact on metabolic load.

Mure *et al.* found that rhythmic gene expression was grouped into two phases—early afternoon and just before dawn—with a marked absence of expression around midnight. This biphasic circadian rhythm is reminiscent of fruitfly behavior but differs from the nocturnal mouse, in which phases of gene expression are spread evenly throughout the day (*7*, *12*). Unfortunately, Mure *et al.* did not determine how rhythmic gene expression was influenced by light, food, or temperature, which may be the subject of future work. For example, nocturnal rodents can be coerced into diurnal activity by restricting food access (*13*) or by increasing nighttime workload for food (*14*). Nevertheless, the data add an important reference atlas for a daytime-active organism—the closest to humans and most complete yet—and reiterate that biological rhythms are not soloed by an SCN overlord but rather arise from a plucky band of tissue-specific mechanisms. ■

"...the internal clock of a diurnal primate...has implications for understanding human biological rhythms and chronotherapy..."

[1]*Laboratory for Synthetic Biology, RIKEN Quantitative Biology Center, Osaka, Japan.* [2]*World Premier International Research Center Initiative–International Research Center for Neurointelligence, University of Tokyo Institutes for Advanced Study, and Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. Email: uedah-tky@umin.ac.jp*

REFERENCES

1. L. S. Mure *et al.*, *Science* **359**, eaao0318 (2018).
2. M. D. Schwartz *et al.*, *Neuroscience* **127**, 13 (2004).
3. R. Cohen *et al.*, *Brain. Behav. Evol.* **75**, 9 (2010).
4. M. D. Schwartz *et al.*, *Brain Res.* **1367**, 146 (2011).
5. N. Mrosovsky *et al.*, *J. Biol. Rhythms* **16**, 471 (2001).
6. C. Ramanathan *et al.*, *Brain Res.* **1073**, 348 (2006).
7. R. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16219 (2014).
8. R. C. Anafi *et al.*, *PLOS Biol.* **12**, e1001840 (2014).
9. A. Goriki *et al.*, *PLOS Biol.* **12**, e1001839 (2014).
10. F. K. Stephan, I. Zucker, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1583 (1972).
11. R. Y. Moore, V. B. Eichler, *Brain Res.* **42**, 201 (1972).
12. M. E. Hughes *et al.*, *PLOS Genet.* **5**, e1000442 (2009).
13. K.-A. Stokkan *et al.*, *Science* **291**, 490 (2001).
14. R. A. Hut *et al.*, *PLOS ONE* **6**, e17527 (2011).

10.1126/science.aat3211

GENETICS

Circadian organization of the genome

The clock protein Rev-erbα regulates genome folding to establish circadian gene repression

By **Carolina Dietrich Mallet de Lima** *and* **Anita Göndör**

Phenotypic plasticity, the potential for phenotypic change in response to external signals, drives adaptation to environmental fluctuations and requires flexible gene regulation (*1*). A seminal example of adaptive plasticity is represented by the circadian clock (*2*), which establishes 24-hour rhythmicity in physiology, metabolic activities, and behavior. As external time cues, such as light and food intake, can reset the phase of oscillations, circadian homeostasis enables light-sensitive organisms to both anticipate and adapt to daily environmental cycles (*2*). On page 1274 of this issue, Kim *et al.* (*3*) provide a glimpse into the genome-wide complexity of transcriptional plasticity during the physiological circadian cycle in mice, with implications for our understanding of diseases linked with deregulation of the circadian clock (*2*).

Chromatin provides a platform for the integration and propagation of signaling and metabolic cues to influence gene expression (*4*–*6*). How combinations of chromatin modifications and transcription factors fine-tune the degree of transcriptional plasticity is central to our understanding of differentiation and diseases with epigenomic deregulation, such as cancer (*7*). Circadian clock–regulated plasticity in chromatin states at loci that are transcribed in a circadian manner is well-documented and correlates with gene expression (*5, 8*). Chromatin, however, possesses regulatory information beyond histone and DNA modifications by influencing how regulatory elements, such as enhancers and promoters, communicate with each other over large genomic distances (*9*). Maps of physical contact probabilities between distant regions have earlier revealed that the genome is organized into topologically associating domains (TADs) displaying high local, intradomain chromatin-fiber contact frequencies. Given that TADs constrain and thereby increase the specificity of enhancer-promoter (E-P) contacts, the mechanisms and dynamics of TAD formation are intensely investigated (*9*).

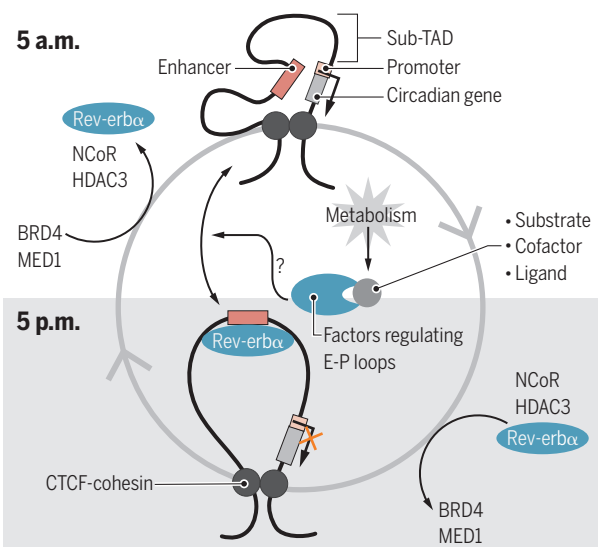
Kim *et al.* have revealed a surprising plasticity in genome-wide chromatin folding during the circadian cycle and documented its role in oscillating transcription. They generated three-dimensional contact maps between distant regions in living cells at 5 a.m. and 5 p.m. to determine changes in genome organization. Consistent with the previously reported cell type–independence of most TAD boundaries, at the two time points examined, Kim *et al.* found stable TAD-boundary locations which correlated with the binding of the chromatin architectural proteins CCCTC-binding factor (CTCF) and cohesin, factors known to regulate TAD-boundary formation (*9*). In addition to stable TADs, the authors identified extensive remodeling of E-P interactions within hundreds of sub-TADs (smaller units within TADs with high interaction frequencies) containing circadian genes with expression peaks at 5 a.m. or 5 p.m.

To identify the mechanism and relevance of circadian fluctuations in E-P contacts, they focused on cryptochrome 1 (CRY1), a major repressor in the core clockwork (*10*). The phase of *Cry1* gene expression is critical for clock function and is fine-tuned by the initial combinatorial binding of transcriptional activators and repressors to *Cry1* enhancers, followed by its derepression upon the removal of its major transcriptional repressor, Rev-erbα (*10*). Analysis of chromatin interactions complemented with mapping of Rev-erbα binding demonstrated that *Cry1* E-P contacts were circadian, in phase with *Cry1* expression, and antiphase to Rev-erbα binding. Accordingly, genetic depletion of Rev-erbα attenuated the amplitude of the frequency of E-P contacts, whereas in vivo overexpression of Rev-erbα reduced the frequency of E-P loops and *Cry1* expression. The mechanism of Rev-erbα action involved the recruitment of nuclear receptor corepressor (NCoR) complex, which has histone deacetylase (HDAC) activity, followed by the eviction of bromodomain-containing protein 4 (BRD4) and its partner mediator complex subunit 1 (MED1)—factors implicated in loop formation (*9*) (see the figure).

Interestingly, E-P loops that are present at 5 a.m. and involve sites that bind Rev-erbα at 5 p.m. could be classified either as "engaged" (gene transcription is repressed at 5 p.m. upon Rev-erbα binding) or "passive" (not repressed at 5 p.m. despite Rev-erbα binding). The observation that Rev-erbα binding was associated with rhythmic change in enhancer and gene activity only at engaged sites raises the questions of what functions the passive sites might have and how the effects of Rev-erbα are buffered against at these sites, at the time points examined. Interestingly, Rev-erbα can repress its target genes through either direct DNA binding or indirect binding, tethered by cell type–specific transcription factors to orchestrate cell type–specific transcriptional rhythms (*11*). Unlike engaged sites, passive sites were not enriched in direct Rev-erbα binding motifs. It might therefore be interesting to explore how local regulatory elements and the various mediators of indirect Rev-erbα binding at passive sites might influence Rev-erbα function(s). They might thus affect the recruitment of corepressors or alter the localization of target regions in the compartmentalized nucleus, where active and repressive nuclear environments tend to

Circadian plasticity of chromatin folding

Enhancer-promoter (E-P) interactions in circadian sub-TADs are under circadian control and antagonized by Rev-erbα. Factors regulating E-P loops that act as metabolic sensors might, in general, mediate cross-talk between metabolism and genome folding.



GRAPHIC: N. DESAI/*SCIENCE*

Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. Email: anita.gondor@ki.se

be spatially separated, to potentially reduce transcriptional noise (*1*).

Kim *et al.* have uncovered a new principle of circadian gene repression by Rev-erbα that involves the dynamic reorganization of genome folding at circadian loci. Previously, it was difficult to predict to what extent genome organization undergoes circadian oscillations. Although these experiments suggest that circadian chromatin plasticity is restricted to sub-TADs enriched in circadian genes, it cannot be ruled out that rhythmic changes in chromatin folding do not leave nearby non-circadian genes unaffected. Given the role of transcriptional noise in cell-state transitions and phenotypic heterogeneity driving tumor evolution (*7*), it will be important to decipher any potential cross-talk between the clockwork and the machinery that regulates stochastic transcriptional fluctuations (*1*).

The findings of Kim *et al.* also support the emerging theme that metabolic states might directly affect not only chromatin marks (*4*, *12*) but also genome folding via "metabolic sensors" with functions in genome organization. Rev-erbα binds to heme (*13*), which is synthesized in a circadian- and Rev-erbα–dependent manner (*14*). Interestingly, heme facilitates Rev-erbα binding to the NCoR complex and its repressor function at metabolic genes (*13*), suggesting potential cross-talk between circadian genome folding and energy homeostasis. Another example is poly(ADP-ribose) polymerase 1 (PARP1), which senses the amount of oxidized nicotinamide adenine dinucleotide (NAD^+) (*4*) and regulates, with CTCF, the rhythmic mobility of circadian genes between transcriptionally permissive and repressive nuclear compartments (*15*). As E-P contacts and their subnuclear localization are dynamic, E-P loop formation and chromatin mobility might be among the first responders to metabolic cues, for example, in response to diet. Further deciphering the cross-talk between metabolic sensors and genome organization will likely shed new light on diseases with deregulated phenotypic plasticity, such as cancer (*7*). ■

REFERENCES

1. E. Pujadas *et al.*, *Cell* **148**, 1123 (2012).
2. J. Bass *et al.*, *Science* **354**, 994 (2016).
3. Y. H. Kim *et al.*, *Science* **359**, 1274 (2018).
4. J. A. van der Knaap *et al.*, *Genes Dev.* **30**, 2345 (2016).
5. S. L. Berger *et al.*, *Cold Spring Harb. Perspect. Biol.* **8**, a019463 (2016).
6. A. I. Badeaux, Y. Shi, *Nat. Rev. Mol. Cell Biol.* **14**, 211 (2013).
7. A. P. Feinberg *et al.*, *Nat. Rev. Genet.* **17**, 284 (2016).
8. N. Koike *et al.*, *Science* **338**, 349 (2012).
9. A. S. Hansen *et al.*, *Nucleus* **9**, 20 (2018).
10. M. Ukai-Tadenuma *et al.*, *Cell* **144**, 268 (2011).
11. B. Fang, M. A. Lazar, *Cold Spring Harb. Symp. Quant. Biol.* **80**, 233 (2015).
12. J. P. Etchegaray *et al.*, *Mol. Cell* **62**, 695 (2016).
13. L. Yin *et al.*, *Science* **318**, 1786 (2007).
14. N. Wu *et al.*, *Genes Dev.* **23**, 2201 (2009).
15. H. Zhao *et al.*, *Mol. Cell* **59**, 984 (2015).

10.1126/science.aat0934

GRAPHIC: A. KITTERMAN/*SCIENCE*

ECOLOGY

A landscape of disgust

Parasite avoidance behavior affects ecology and evolution in ways similar to predator avoidance

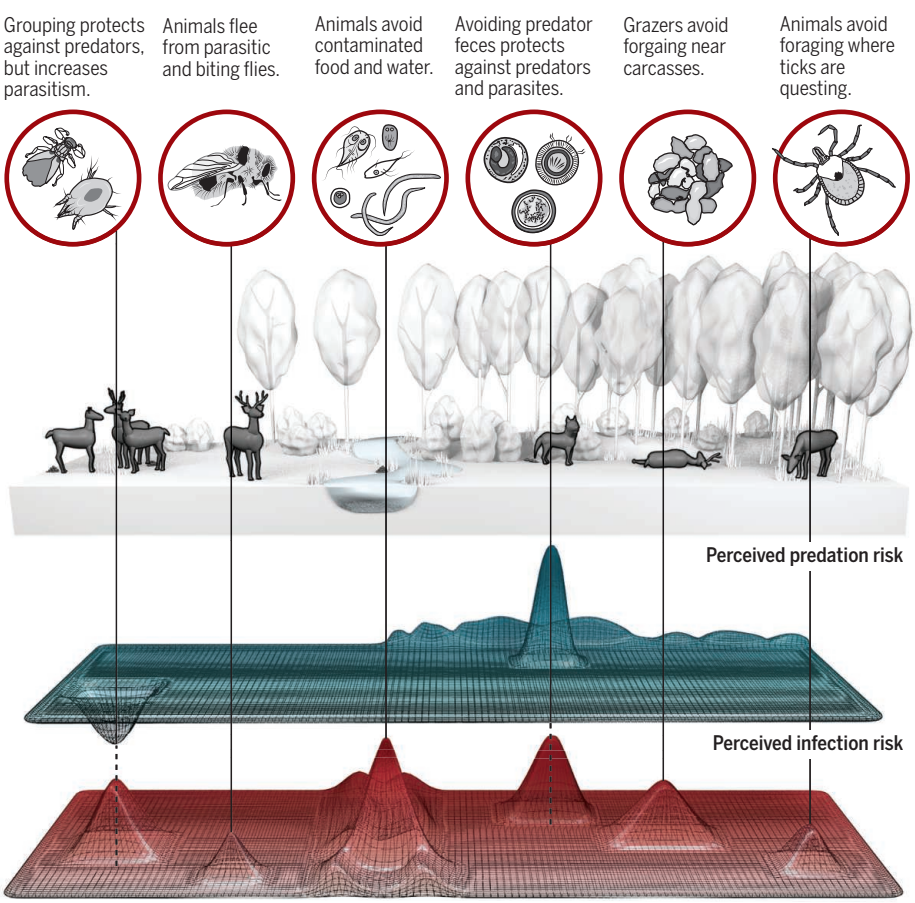
By **Sara B. Weinstein,[1] Julia C. Buck,[2] Hillary S. Young[2]**

A rancid meal, a moist handshake, a pile of feces: These phenomena elicit disgust and avoidance that protect humans from our most pervasive consumer—infectious agents. This avoidance is not specific to humans. Various animals alter their behavior to avoid infection (*1*). For instance, Poirotte *et al.* recently showed that mandrills avoid parasite-contaminated feces and refrain from grooming infected individuals (*2*). These primates' nuanced ability to detect and alter their behavior in response to differential exposure risk suggests close parallels to the "landscape of fear" elicited by predators (see the figure), with perceived peaks and valleys driven by parasite abundance and exposure risk.

Owing to the high cumulative fitness costs of parasitism over an individual's lifetime, there will be strong selection for parasite avoidance. Yet, because most parasites are difficult to detect, avoidance often relies on indirect cues that are driven by long-term associations. Many species avoid feces and carcasses, regardless of infection status (*1*). Infection can also alter the chemical composition of sweat, breath, and feces, allowing more targeted avoidance of infected individuals (*3*). For instance, mandrills identify

Overlapping landscapes of fear and disgust

Animals detect predation threats, creating a three-dimensional fear landscape with mountains of risk and valleys of safety. Animals can also detect infection threats, which form a landscape of disgust. The resulting overlapping landscapes have far-reaching ecological and evolutionary consequences.



and avoid infected conspecifics on the basis of olfactory cues associated with the altered chemical composition of infected feces (*2*). Thus, whereas prey "fear" predators, parasite avoidance is more akin to disgust—an innate aversion to cues associated with parasites (*1*). As with predator avoidance, parasite avoidance incurs costs such as altered social interactions, reduced reproductive opportunities, diminished foraging, and increased energy expenditure (*4*). To predict when parasite avoidance is most likely to occur and understand its consequences, it must be integrated into the same trade-off framework that is widely applied to predator avoidance.

Beyond its direct effects on hosts, parasite avoidance has cascading effects at a range of scales and in various systems (*5*). Ants and their parasitoids provide a useful model system for studying the cascading effects of parasite avoidance on herbivory. Azteca ants experimentally exposed to parasitoid flies dramatically reduce their activity levels to avoid infection. This reduces the ants' ability to attack their herbivorous prey, the coffee berry borer, and allows the latter to consume more coffee berries (*6*). Such ecological impacts of parasite avoidance are not limited to top-down effects on vegetation; they also alter mutualism, competition, and predation.

Like the predator-induced landscape of fear, a parasite-induced landscape of disgust (see the figure) has far-reaching consequences for population dynamics, species interactions, and ecosystem structure. However, whereas direct effects of predators are naturally restricted to prey species (*7*), a parasite-induced landscape of disgust extends to all animals, including very large species and top predators that are typically immune to predation (*8*). For example, a recent synthesis by Moleón *et al*. showed that carnivores avoid foraging on other carnivore carcasses and rarely engage in cannibalism (*9*). Feeding on conspecifics increases disease transmission, and forgoing this food resource reduces exposure to shared parasites (*10*). At a landscape level, this avoidance behavior dramatically alters the amount of carcass material that is available in an ecosystem, with important downstream effects. For instance, carcasses not consumed by mammalian scavengers may support more diverse and abundant arthropod communities (*9*). Furthermore, when carcasses decompose, their nutrient inputs increase plant growth (*11*). Like the ecosystem-level impacts of predator avoidance, the ecosystem-level impacts of parasite avoidance are likely to be substantial.

Parasite and predator avoidance can exert opposing pressures on host and prey species. For example, tadpoles increase their activity levels to protect themselves from trematode infection, but this increases their susceptibility to visual predators (*12*). The tadpoles fear and avoid both predators and parasites, but do not perceive these threats as equal. When experimentally forced to choose between predation and parasitism risk, tadpoles opt for infection (*13*). This makes sense because predation has a larger short-term fitness effect than does parasitism, and the optimal behavior avoids the greatest immediate threat.

However, parasite and predator avoidance are not always mutually exclusive. For instance, avoiding predator feces protects prey against both predators and their parasites (*14*), and, when animals group together, they dilute their risk of attack by both predators and some parasites (for example, ticks, botflies) (*15*). Although predator avoidance has received more attention, parasite avoidance also has strong effects on foraging, movement, and social interactions (*1*).

Parasite avoidance is likely to have important consequences. For example, grazing animals such as cows, sheep, and horses avoid fecal contaminated forage (*4*), and changing livestock practices to recognize the needs and capabilities of animals to avoid parasites will reduce the cost associated with avoidable livestock diseases. Furthermore, conservation planning and projections may benefit from knowledge of how shrinking and fragmented habitat limits the ability of animals to avoid risky sites and conspecifics. As populations of large predators continue to decline, parasite avoidance may become a relatively stronger driver of ecological processes. Integrating parasites into the ecology of fear is thus crucial to understanding the general ecological and evolutionary effects of parasites on ecosystems and predicting how ecosystems are likely to respond to the rapidly changing environments of the Anthropocene. ■

REFERENCES

1. V. A. Curtis, *Trends Immunol.* **35**, 457 (2014).
2. C. Poirotte *et al.*, *Sci. Adv.* **3**, e1601721 (2017).
3. M. Shirasu, K. Touhara, *J. Biochem.* **150**, 257 (2011).
4. M. R. Hutchings, J. Judge, I. J. Gordon, S. Athanasiadou, I. Kyriazakis, *Mammal Rev.* **36**, 1 (2006).
5. J. C. Buck, W. J. Ripple, *Trends Ecol. Evol.* **32**, 681 (2017).
6. G. L. Pardee, S. M. Philpott, *Environ. Entomol.* **40**, 581 (2011).
7. J. S. Brown, J. W. Laundré, M. Gurung, *J. Mammal.* **80**, 385 (1999).
8. A. R. Sinclair, S. Mduma, J. S. Brashares, *Nature* **425**, 288 (2003).
9. M. Moleón *et al.*, *J. Animal Ecol.* **86**, 1179 (2017).
10. D. W. Pfennig, S. G. Ho, E. A. Hoffman, *Animal Behav.* **55**, 1255 (1998).
11. W. C. Turner *et al.*, *Proc. R. Soc. B* **281**, 1 (2014).
12. J. A. Marino Jr., E. E. Werner, *Ecology* **94**, 2697 (2013).
13. J. Koprivnikar, L. Penalva, *PLOS ONE* **10**, e0116569 (2015).
14. S. B. Weinstein, C. W. Moura, J. F. Mendez, K. D. Lafferty, *Oikos* 10.1111/oik.04866 (2018).
15. M. S. Mooring, B. L. Hart, *Behaviour* **123**, 173 (1992).

10.1126/science.aas8694

[1]Department of Biology, University of Utah, Salt Lake City, UT 84112, USA. [2]Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA 93106, USA. Email: weinsteins@si.edu

PSYCHOLOGY

Logic in babies

12-month-olds spontaneously reason using process of elimination

By **Justin Halberda**

The success of science rests on our human ability to reason logically. But where does this ability come from? Is it an inherent attribute of mind that even a young child might possess, or is it a hard-won accomplishment mastered later in life? On page 1263 of this issue, Cesana-Arlotti *et al*. (*1*) reveal that one essential form of logical inference, process of elimination, is within the toolkit of 12-month-old infants. This reveals the earliest known foundations of our human ability to reason logically. The race to document the range of early logical abilities shared by infants, adults, and nonhuman animals, and to determine how these foundational abilities empower our broader capacities to reason, has begun.

Every scientific method requires a supporting logic. For Francis Bacon (*2*), this was unfettered empirical observation followed by induction (reasoning from many cases to form a general principle). Karl Popper (*3*) stressed the importance of hypothesis testing and the ability to refute hypotheses found to be false (science as an extended instance of process of elimination). And Thomas Kuhn (*4*) highlighted the dramatic changes that occur during scientific revolutions, in which wholly new models of phenomena are created through model building and abduction (sometimes called "inference to the best explanation"). Examples of such revolutions are the Copernican Revolution (*5*) producing the heliocentric model of the solar system, and the Einsteinian Revolution of special relativity in which space and time become one (*6*). In each of these cases (induction, hypothesis testing, abduction), the work of science is supported by an underlying logic. No logic, no science.

Where does our human faculty to reason logically come from? At the end of the previous century, Susan Carey suggested that such logic might be deep in our developmental roots: A young child might reason using intuitive theories, models, and abduc-

Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA. Email: halberda@jhu.edu

An infant at play is also an infant reasoning logically.

PHOTO: OKSANA KUZMINA/SHUTTERSTOCK.COM

tion, much like a professional scientist (*7*). Other developmental psychologists pushed this further and suggested that even infants might be understood to be little scientists drawing inferences in the cradle about objects and causes (*8*). But suggestions like these are frequently met with incredulity (*9*): After all, it often feels like logical reasoning is effortful, conscious, and even linguistically based (*10*). These characteristics, if accurate, would seem to preclude the possibility that preverbal infants could engage in any such process.

Cesana-Arlotti *et al.* asked whether prelinguistic 12- and 19-month-old infants would spontaneously reason using process of elimination. This is a form of inference also known as disjunctive syllogism or modus tollendo ponens—it is any argument of the form: A or B, not A, therefore B. Cesana-Arlotti *et al.* relied on one of the few behaviors babies voluntarily engage in—looking at whatever they find most interesting (*11*). They measured infants' looking at computerized vignettes in which two different objects (A and B) were shown being hidden behind a wall. Infants watched as a cup scooped one of the objects from behind the wall, and then came to rest next to the wall—critically, only the topmost edge of the contained object could be seen peeking out of the cup, such that infants could not tell for sure whether the object was A or B. At this moment, infants could have formed a disjunctive thought—for example, "either the object in the cup is object A or it is object B." Next, this ambiguity was resolved: The wall dropped to reveal that object A was behind the wall, but the contents of the cup remained hidden. This is the moment of potential elimination, and an opportunity for infants to draw a key inference—"because object A is not in the cup, object B must be in the cup." Finally, infants' expectations for the cup's contents were tested: Either the expected object (object B) or, surprisingly, another object A emerged from the cup. Infants looked longer at the surprising outcome—an indication that their expectations were violated and a hint that they were seeking further information to resolve the conflict (*12*).

Potentially even more exciting, infants also showed signs of making the necessary inferences along the way—for example, upon seeing which of the two objects was behind the wall, but before the cup's contents were revealed, infants' pupils dilated and they tended to shift their fixation to the cup (consistent with them inferring which object must be inside). This pattern suggests that infants used the information they had seen to reason through a disjunctive syllogism (A or B, not A, therefore B). There were also additional versions of these vignettes that manipulated the precise sequence of hiding and revealing, which allowed the authors to determine what specifically the infants were remembering and expecting during each moment.

The careful crafting of stimuli and clever analyses of infants' spontaneous looking behavior by Cesana-Arlotti *et al.* show us that infants have the capacity to reason by process of elimination. By contrast, whereas nonhuman animals such as dogs facing similar situations of ambiguity may ultimately form the right conclusion, they appear to arrive at this hunch using an associative rather than logical process (*13*). Unwinding the similarities and differences among infant, child, adult, and animal reasoning abilities is an important task for the future, as is determining why, despite being deployed by infants, reasoning feels effortful and why humans often "go with our gut" and rely on heuristics and biases rather than logical inference (*14*).

A new field, studying the foundations of logical abilities, is emerging. As with other successes of developmental psychology in recent decades [as in work on the psychology of number, language, and theory of mind (*15*)], this new field will make use of behavior and brain studies with infants, children, and adults; studies of nonhuman animal abilities; studies of the correlations between reasoning and other abilities; and training studies to determine where we all begin and how maturation and experience build upon these foundations. It is a thrilling time for us as scientists—using logical reasoning to understand how we reason logically. ■

REFERENCES

1. N. Cesana-Arlotti *et al.*, *Science* **359**, 1263 (2018).
2. F. Bacon, *Novum Organum Scientiarum* (1620).
3. K. Popper, *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft* (1934).
4. T. S. Kuhn, *The Structure of Scientific Revolutions* (Univ. of Chicago Press, Chicago, IL, ed. 3, 1962).
5. N. Copernicus, *De revolutionibus orbium coelestium* (1543).
6. A. Einstein, *Annalen der Physik* **17**, 891 (1905) [English transl. by G. Barker Jeffery, W. Perrett, in *The Principle of Relativity* (Methuen, London, 1923)].
7. S. Carey, *Conceptual Change in Childhood* (Bradford Books, MIT Press, Cambridge, MA, 1985).
8. A. Gopnik, A. N. Meltzoff, P. K. Kuhl, *The Scientist in the Crib: Minds, Brains, and How Children Learn* (Harper Collins, New York, 1999).
9. M. M. Haith, *Infant Behav. Dev.* **21**, 167 (1998).
10. Aristotle, *Rhetoric* [transl. by W. Rhys Roberts (Courier Corporation, 2004)].
11. R. L. Fantz, *Science* **140**, 296 (1963).
12. A. E. Stahl, L. Feigenson, *Science* **348**, 91 (2015).
13. J. S. Watson *et al.*, *J. Comp. Psychol.* **115**, 219 (2001).
14. A. Tversky, D. Kahneman, *Science* **185**, 1124 (1974).
15. S. Carey, E. Spelke, in *Mapping the Mind: Domain Specificity in Cognition and Culture*, L. A. Hirschfeld, S. A. Gelman, Eds. (Cambridge Univ. Press, Cambridge, MA, 1994).

10.1126/science.aas9183

POLYMERS

Random copolymers that protect proteins

Synthetic polymers are designed to stabilize proteins in polar and nonpolar solvents

By **Alfredo Alexander-Katz**[1] *and* **Reid C. Van Lehn**[2]

Scientists have tried and in some limited cases succeeded to harness proteins to do chemistry (*1*) or use them in functional materials. However, most proteins only function correctly if they fold into specific conformations, which typically occurs with the assistance of other proteins (such as chaperones, translocons, or transporters) that mediate structure formation, membrane insertion, and intracellular trafficking (*2*, *3*). Several methods have been used to improve protein stability in nonbiological environments—including micelle encapsulation, polymer conjugation, and sol-gel trapping (*4*)—but for most intended applications, they suffer from low levels of functionality, difficult chemical postfunctionalization, or the requirement of very specific solvent environments. On page 1239 of this issue, Panganiban *et al.* (*5*) introduce an approach for stabilizing proteins in disparate solvent environments that does not suffer from these drawbacks.

The surfaces of folded proteins are chemically diverse, with small nonpolar, polar, and charged "patches" of the surface in close proximity. By characterizing the distribution of patch size and separation for several exemplary proteins, Panganiban *et al.* developed a statistical representation of the chemical diversity expected of protein surfaces. They then synthesized random heteropolymers (RHPs) composed of four monomers with chemical properties designed to interact with the chemical species present on protein surfaces. Critically, the RHPs mimicked the statistical distribution of protein surface patches rather than the specific surface properties of a protein of interest (see the figure).

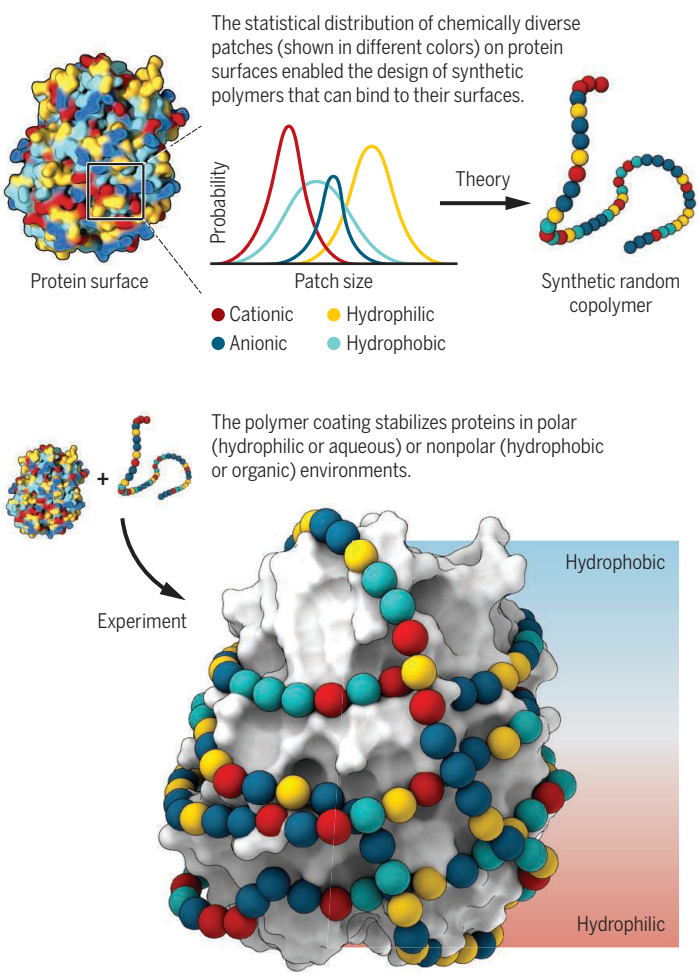
Using this design strategy, they identified a RHP composition that strongly binds to protein surfaces in organic solvents yet weakly binds in water. Molecular simulations at multiple length scales demonstrated that the RHP adopts local conformations to maximize favorable interactions with the protein surface, leveraging the polymer's conformational flexibility and the complementary chemical functionalities engineered into the polymer backbone. Experimentally, this interaction leads to astonishing protein folding and stabilization in a variety of non-native environments compared with alternative methods. The authors show that the same RHP composition increases membrane protein folding in water and increases water-soluble protein activity in organic solvent by an order of magnitude (or more). These results demonstrate that RHPs can act like omnifunctional chaperones that can stabilize both lipophilic proteins in aqueous solution and hydrophilic proteins in organic solvents, without compromising protein function.

Researchers had previously proposed (and computationally proved) that statistically controlled random copolymers (such as unfolded proteins) could recognize "random" patterns on surfaces (*6*, *7*). Panganiban *et al.* validated this conjecture even for the curved surfaces of proteins, a remarkable result given the complexity of protein interfaces. This work provides a blueprint for matching RHPs to protein surfaces, further blurring the line between the biological and the synthetic world. It also raises new questions. How many monomer chemistries are necessary to bind a given protein surface? Can other chemistries be developed along these lines? How do the physical properties of proteins, such as curvature, affect the necessary monomer statistical distribution?

Expanding on the strategy used in this work, it may be possible to elucidate the rules used by biology to ensure protein function by analyzing the properties of effective RHPs. For example, a key feature of the RHPs is their conformational flexibility. Intriguingly, large conformational fluctuations and transient unfolding

A coat that fits off the rack

A statistical analysis of protein surfaces enabled Panganiban *et al.* to design random copolymers as coatings. These polymers stabilize proteins in a wide variety of solvents.



GRAPHIC: V. ALTOUNIAN/*SCIENCE* AND REID C. VAN LEHN/UNIVERSITY OF WISCONSIN-MADISON; PDB ID 1CF3

[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [2]Department of Chemical and Biological Engineering, University of Wisconsin–Madison, 1415 Engineering Drive, Madison, WI 53706, USA. Email: aalexand@mit.edu

are necessary for the function of water-soluble and membrane protein chaperones, suggesting that conformational flexibility may be a general feature of chaperones (*8*, *9*). Comparing the chemical features of effective RHPs with the sequences of similarly flexible intrinsically disordered proteins may further enable the identification of proteins with chaperone activity. Optimizing RHP compositions based on the chemical diversity found in separate classes of proteins could also reveal the minimal chemical features necessary for chaperone action. Last, analyzing the thermodynamics underlying the insertion of RHP-bound membrane proteins into a lipid bilayer, possibly aided by simulation techniques such as those used in this work, could resolve long-standing questions regarding the forces driving membrane protein folding and assembly (*10*).

The demonstration by Panganiban *et al.* that RHPs can stabilize proteins in different media without compromising their function should open the door for the generation of new technologies as well as scientific findings. This result could enable the use of many proteins for diverse chemical processes, but the prospects of this work extend beyond chemistry. For example, one can envision stabilizing proteins in healthy fats in order to solve food problems in regions where refrigeration is not available, or using biodegradable random copolymers to neutralize viruses or toxic agents to prevent infection or biological attacks. ■

"...[random heteropolymers] can act like omnifunctional chaperones that can stabilize both lipophilic... and hydrophilic proteins..."

REFERENCES

1. U. T. Bornscheuer *et al.*, *Nature* **485**, 185 (2012).
2. F. U. Hartl, M. Hayer-Hartl, *Science* **295**, 1852 (2002).
3. S. C. Shao, R. S. Hegde, *Annu. Rev. Cell Dev. Biol.* **27**, 25 (2011).
4. P. V. Iyer, L. Ananthanarayan, *Process Biochem.* **43**, 1019 (2008).
5. B. Panganiban *et al.*, *Science* **359**, 1239 (2018).
6. D. Bratko *et al.*, *Chem. Phys. Lett.* **280**, 46 (1997).
7. A. J. Golumbfskie *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11707 (1999).
8. F. C. Liang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1615 (2016).
9. J. C. A. Bardwell, U. Jakob, *Trends Biochem. Sci.* **37**, 517 (2012).
10. F. Cymer *et al.*, *J. Mol. Biol.* **427**, 999 (2015).

10.1126/science.aat0155

CANCER THERAPY

Beyond PARP–POLθ as an anticancer target

Targeting cancers dependent on DNA polymerase θ has considerable clinical potential

By **Geoff S. Higgins**[1] *and* **Simon J. Boulton**[2,3]

Since the discovery that DNA is unstable and prone to decay, targeting DNA repair deficiencies has become a proven and effective strategy in the fight against cancer. Over the past decades, it has become increasingly apparent that selective loss of DNA repair pathways is an early and frequent event in tumorigenesis, occurring in 40 to 50% of many cancer types. Loss of DNA repair likely provides a selective growth advantage to tumor cells as this results in genetic instability and/or enhanced mutation rates, which can drive tumor evolution (*1*). However, DNA repair–deficient cancers often become critically dependent on backup DNA repair pathways, which present an "Achilles heel" that can be targeted to eliminate cancer cells. This is the basis of synthetic lethality and is exemplified by the success of poly(ADP-ribose) polymerase (PARP) inhibitors in treating *BRCA*-deficient breast and ovarian cancers. However, new therapeutic strategies are urgently needed to overcome acquired and innate PARP inhibitor (PARPi) resistance and to exploit other DNA repair deficiencies. Optimism is increasing that targeting DNA polymerase θ (POLθ), which is involved in DNA repair, will not only synergize with PARPi's but may have broader utility in cancer treatment.

The realization that tumor-specific alterations in DNA repair pathways could be exploited therapeutically was first established in 2005 (*2*, *3*). Inactivating *BRCA* mutations, which account for ~10% of hereditary breast and ovarian cancers, are defective for homologous recombination (HR), a key pathway of DNA double-strand break (DSB) repair. BRCA-deficient cancer cells, but not the wild-type controls, were found to be exquisitely sensitive to inhibitors of the single-strand break repair enzyme PARP owing to the inability to deal with the lesions induced by PARPi in the cancer cells (*2*, *3*). Recent studies have established that PARPi's act by trapping PARP on DNA. This stalls replication forks, which require HR for stabilization and replication restart (*4*). Hence, cells deficient in HR (for example, with *BRCA* mutations) are unable to respond to trapping of PARP on DNA, leading to cell death.

Currently, there are five PARP inhibitors in clinical trials, which have clear benefit in some patients with HR-deficient (HRD) cancers. *BRCA* mutations represent the most widely studied HRD subtype and have previously been shown to be associated with the greatest progression-free survival benefits from PARPi therapy (*5*). Unfortunately, of the HRD cancers that do respond to PARPi's, ongoing studies suggest that these tumors ultimately progress because of acquired PARPi resistance. It is also clear that PARPi's do not work in all DNA repair–deficient tumors, such as gastric cancers defective for DNA damage signaling by ataxia-telangiectasia mutated (ATM) (*6*), nor do they target increased spontaneous DNA damage caused by DNA replication stress, which is a hallmark of many tumors. POLθ has emerged as a promising target for the treatment of various DNA repair–deficient cancers, including those that are HRD. Evidence suggests that POLθ inhibitors could synergize with PARPi's (*7*, *8*), as they act by a different mechanism, and may also provide a means to selectively sensitize tumors to radiotherapy (*9*).

DSBs are predominantly repaired by the nonhomologous end joining (NHEJ) DNA repair pathway in G_1 phase of the cell cycle and by HR in S–G_2 phases. Microhomology-mediated end joining (MMEJ) is a "backup" DSB repair pathway in the event that NHEJ or HR are compromised (see the figure). HR and MMEJ share the same substrate, namely, a resected DSB that contains a 3′ single-stranded DNA (ssDNA) overhang bound by replication protein A (RPA). Hence, if HR fails or is deficient, then MMEJ is the favored option to repair resected DSBs. POLθ possesses both DNA polymerase and helicase domains, which are important for MMEJ in vivo (*10*). Biochemical evidence

[1]*Cancer Research UK–Medical Research Council Oxford Institute for Radiation Oncology, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK.* [2]*The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK.* [3]*Artios Pharma Limited, Maia, Babraham Research Campus, Cambridge CB22 3AT, UK. Email: geoffrey.higgins@oncology.ox.ac.uk; simon.boulton@crick.ac.uk; simon.boulton@artiospharma.com*

suggests that the helicase domain displaces RPA at resected DSBs to facilitate annealing of ssDNA containing microhomologies (short segments of complementary DNA sequence). The polymerase activity then completes DNA synthesis to fill in the resected gap before ligation completes the MMEJ repair reaction (*10*). MMEJ is error-prone as it often results in loss of genetic information at the break site and can promote repair between DSBs from different parts of the genome, leading to translocations and other complex rearrangements. Although other proteins such as RPA and CtBP-interacting protein (CtIP) are important for MMEJ, their roles in DNA damage recognition and in other DNA repair processes arguably make them less attractive targets for drug development because of their lack of specificity.

One of the principle reasons why POLθ has become a major focus for drug discovery is that its expression is largely absent in normal cells but is increased in many cancers, both with and without HRD (*7*). Indeed, POLθ overexpression in multiple HR-proficient tumor types such as lung, gastric, and colorectal cancer, is associated with adverse clinical outcomes (*11*). Critically, tumor cells harboring DSB repair defects develop a pronounced dependency on MMEJ to the extent that synthetic lethality has been shown to occur when POLθ is disrupted in cells deficient in either HR, NHEJ, or ATM (*7, 8, 12*). The reasons why POLθ is upregulated and associated with poor outcomes in many tumors is not well understood, but one possibility is that POLθ repairs spontaneous DNA damage present in cancer cells and therefore affords them a growth advantage (*13*). Alternatively, error-prone DNA repair associated with POLθ activity could also contribute to greater genomic instability and lead tumors to develop greater genetic diversity and evolutionary plasticity, which can drive resistance to cytotoxic therapies.

Despite the substantial clinical benefits observed with PARPi monotherapy in *BRCA*-mutated cancers, treatment resistance is a common occurrence, and POLθ inhibition might be beneficial in treating those tumors that progress on PARPi's. PARPi resistance can arise from a variety of genetic alterations, including those that reactivate HR. Although POLθ inhibition would not be expected to cause tumor response in cases where secondary reversion *BRCA* mutations cause restoration of HR activity, over half of recurrences arise from other changes, including decreased expression of DNA repair proteins such as PARP1 or p53-binding protein 1 (53BP1) (*14*). It is possible that POLθ inhibitors would induce tumor response in these instances of PARPi resistance. Tumors may eventually develop resistance to POLθ inhibitors, although the mechanisms likely to be involved are not yet clear. It is, however, anticipated that a primary mechanism would be secondary *BRCA* mutations causing reinstatement of HR function.

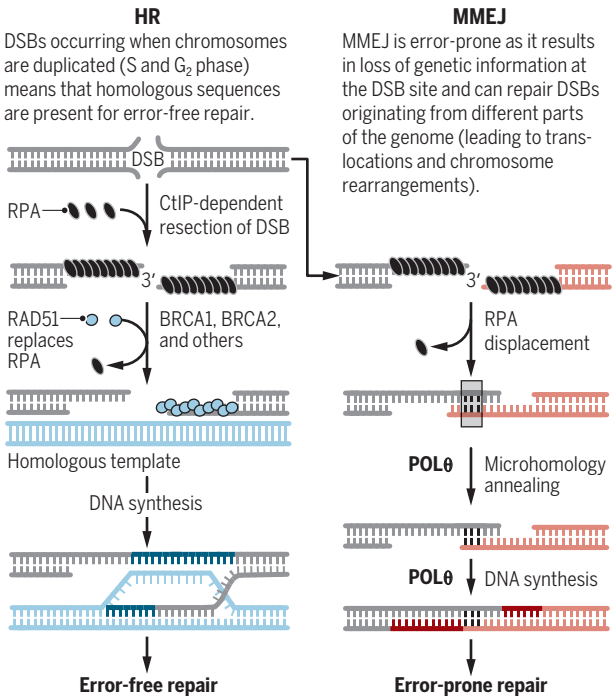
Because of their distinct mechanisms of action, combined PARPi and POLθ inhibitor therapy might be expected to have a synergistic effect; simultaneously driving DSB formation while inhibiting DSB repair. Dual POLθ and PARP inhibition might therefore lead to more prolonged, clinically meaningful responses and possibly delay the emergence of resistant disease compared with either therapy alone. In vitro data support this optimism, because PARPi treatment, combined with reduced POLθ expression, has been shown to synergistically reduce colony formation in BRCA1-deficient cells (*7*). The dramatic tumor growth delay seen in xenograft studies, whereby human HRD tumor cells are subcutaneously engrafted in mice, also points to a synergistic effect of POLθ and PARP deficiency in tumors with impaired HR (*7*).

In addition to the synthetic lethal effects, disruption of POLθ has also been shown to induce sensitization of tumors to radiotherapy in HR-proficient tumor cells at a level that would be expected to have considerable clinical benefit (*9*). Radiotherapy is widely used in the curative management of cancer patients, either as an adjunct to surgery, or as definitive treatment alone or in combination with cytotoxic chemotherapy. POLθ is overexpressed in several common cancers that are routinely treated with radiotherapy, such as lung and rectal cancer (*11*). Despite developments in the precision and accuracy of radiotherapy, the unavoidable irradiation of healthy tissues surrounding tumors can induce side effects that limit the dose of radiation that can be delivered and thus the tumor cell killing that can be induced. The marked disparity in POLθ expression between tumor cells and most normal tissues has raised the exciting prospect that specific POLθ inhibitors, delivered concurrently with radiotherapy, could improve tumor control without exacerbating the associated side effects of radiotherapy.

The clinical potential for POLθ inhibition to prolong the survival of patients with *BRCA*-mutated tumors, or HRD tumors in general, and to improve the cure rates of patients receiving radiotherapy are substantial. It is hoped that the ability to define those patients likely to benefit from POLθ inhibitor treatment should enable this emerging class of anticancer therapy to be rapidly translated into the clinic. ■

DSB repair by HR and MMEJ

In HRD cancer cells, MMEJ becomes active, making POLθ a cancer-specific therapeutic target that induces synthetic lethality in HRD tumors.



GRAPHIC: K. SUTLIFF/*SCIENCE*

REFERENCES AND NOTES

1. E. M. Kass *et al.*, *Mol. Cell* **62**, 777 (2016).
2. H. E. Bryant *et al.*, *Nature* **434**, 913 (2005).
3. H. Farmer *et al.*, *Nature* **434**, 917 (2005).
4. Y. Pommier *et al.*, *Sci. Transl. Med.* **8**, 362ps17 (2016).
5. J. Ledermann *et al.*, *Lancet Oncol.* **15**, 852 (2014).
6. Y.-J. Bang *et al.*, *Lancet Oncol.* **18**, 1637 (2017).
7. R. Ceccaldi *et al.*, *Nature* **518**, 258 (2015).
8. P. A. Mateos-Gomez *et al.*, *Nature* **518**, 254 (2015).
9. G. S. Higgins *et al.*, *Cancer Res.* **70**, 2984 (2010).
10. P. A. Mateos-Gomez *et al.*, *Nat. Struct. Mol. Biol.* **24**, 1116 (2017).
11. K. Kawamura *et al.*, *Int. J. Cancer* **109**, 9 (2004).
12. D. W. Wyatt *et al.*, *Mol. Cell* **63**, 662 (2016).
13. T. Goullet de Rugy *et al.*, *Biol. Open* **5**, 1485 (2016).
14. C. J. Lord, A. Ashworth, *Science* **355**, 1152 (2017).

ACKNOWLEDGMENTS

G.S.H. is an associate professor in the Cancer Research UK–Medical Research Council Oxford Institute for Radiation Oncology, an honorary consultant clinical oncologist in Oxford University Hospitals National Health Service trust, and a medical research fellow at Corpus Christi College. S.J.B. is a senior group leader at the Francis Crick Institute in London and senior vice president of Science Strategy at Artios Pharma Ltd., Babraham, UK, which develops cancer therapy targeting DNA damage response pathways.

10.1126/science.aar5149

POLICY FORUM

EDUCATION

Societal inequalities amplify gender gaps in math

Egalitarian countries cultivate high-performing girls

By **Thomas Breda,[1,2] Elyès Jouini,[2,3] Clotilde Napp[2,3]**

While gender gaps in average math performance are now close to zero in developed countries, women are still strongly underrepresented among math high performers (*1*). This gender gap contributes to the underrepresentation of women in math and science in higher education and to their subsequent worse position in the labor market (*2*, *3*). With the roles of nature and nurture (*4*–*6*) on gender performance gaps having been debated for more than a century, research in the 1990s and 2000s (*7*–*9*) suggested a cultural origin, relating gender gaps in math to measures of countries' gender inequality. However, with more recent studies (*10*–*12*) having shown that this relation is weak, today we have no clearly identified relationship between countries' socioeconomic or cultural environment and the gender gap in math. We relate below gender gaps in math to societal inequalities that are not directly related to gender. We find a strong and robust relationship and provide tests suggesting that it is causal: Countries that are generally more egalitarian, or that have institutions more conductive to equality, have a lower gender performance gap in math, suggesting that this gap is partly shaped by more general societal inequalities.

According to the Programme for International Student Assessment (PISA), there are on average only seven girls for ten boys in the top decile of the math performance distribution among the 35 countries belonging to the Organisation for Economic Co-operation and Development (OECD). Underrepresentation of girls at high levels of performance is a common feature of all OECD countries (table S1a) and has remained remarkably stable since 2000 (table S1b). Gender gaps of the same magnitude are also observed in science and reading, the latter in favor of girls (table S1). Much former research has been based on the gender stratification hypothesis, according to which gender differences in opportunities and status shape numerous socialization processes that in turn may affect performance (*7*). We elaborate on this idea, assuming that processes that transform differences in status between social groups into differences in performance depend on the degree of countries' inclusiveness; inclusive countries are likely to mitigate the impact of status differences in general. We hypothesize that women have a lower status than men in virtually all countries but that this lower status is more likely to be detrimental to girls' performance in countries that are in general less fair and inclusive; the more unequal a country, the more the status difference between boys and girls should translate into differences in performance.

INEQUALITIES AND PERFORMANCE

We analyze data from five successive waves of PISA, an every-three-year international assessment of the knowledge and skills of 15-year-old students in mathematics, reading, and science taking place in more than 70 countries (see details on all data and analyses in the supplementary materials (SM)]. We relate gender gaps in PISA performance to measures of countries' societal inequalities that are not directly related to gender but reflect (i) general socioeconomic and cultural inequalities, including pure income inequalities, or (ii) educational inequalities, including inequalities in performance and in learning opportunities (*13*).

We focus primarily on "high performers" at levels 5 or 6 from the six PISA proficiency levels (10.7% of all students in OECD countries in math in 2015). We use as our main measure of the gender performance gap in math the ratio of girls to boys among these high performers. This is because (i) this ratio is very unfavorable to girls and has not narrowed over time and (ii) performance at high levels is related to underrepresentation of women among STEM college graduates.

Among the 35 OECD countries observed in 2015, this ratio is negatively and significantly correlated with socioeconomic and cultural inequalities such as the income Gini index (r = −0.55) (see fig. S1) or the variance in the socioeconomic and cultural background of a country's students (r = −0.66), with inequalities in school performance such as the share of students from a low socioeconomic and cultural background among high performers (r = 0.70) (see fig. S1), or with inequalities in learning opportunities at school such as the between-school variation in students' socioeconomic background (r = −0.61).

Consistent with a link between math gender gap and the way countries perpetuate or reduce initial status differences, we find a strong correlation between the gender gap in math and intergenerational earnings elasticity (r = −0.63), education mobility measured as parents-child correlations in years of schooling (r = −0.54), or the index of inequality of economic opportunity (r = −0.64), which measures income inequality due to predetermined circumstances beyond individual control (such as region of birth or parental background).

A more systematic study of those relationships using linear regression models shows that for a large variety of measures of societal inequalities, a higher level of inequality is associated with a significantly lower girls-to-boys ratio among high performers in math (see the table and table S2, column 1). About 30% of the cross-country variance in this ratio can be accounted for by a single measure of inequality, and the explained variance reaches 60% when three variables reflecting different aspects of inequalities are jointly included.

Several robustness checks show that the relation is still observed when (i) controlling for countries' gross domestic product (GDP) and extent of gender stratification (table S3); (ii) including 37 partner non-OECD countries (table S4); (iii) considering the gender gap in average performance, although the relation is weaker (table S5); (iv) using different measures of the gender gap in math among high performers (table S6); (v) using former PISA surveys or other data sources (see, e.g., table S7); and (vi) using lagged measures of income inequalities [in the 1980s and to some extent in the 1950s and 1960s (see SM)]. We also show that inequalities are associated with lower school performance of both boys and girls, with the estimated association systematically larger for girls (table S8).

We conclude that the relationship between the math gender performance gap and societal inequality is larger and more robust than other relationships already documented with more obvious country characteristics such as gender stratification or economic development.

[1]*Paris School of Economics, Paris, France.* [2]*Centre National de la Recherche Scientifique, Paris, France.* [3]*Université Paris-Dauphine, PSL Research University, Paris, France. All authors contributed equally to this paper. Email: thomas.breda@ens.fr; jouini@ceremade.dauphine.fr; clotilde.napp@dauphine.fr*

To back up this point, we run "horse races" and also apply three standard machine-learning techniques for model selection to compare the explanatory power of measures of gender inequality and of countries' development to our measures of non-gender-related inequalities (see SM). For instance, the effect of the Gender Gap Index (GGI) or of the Female Labor Force Participation on the gender gap in math becomes statistically nonsignificant as soon as a measure of (non-gender-related) societal inequalities is introduced as a competing explanatory variable in regression models (table S9).

Societal inequalities are also associated with lower girls-to-boys ratios in science (increasing the gender gap, as in math) and higher boys-to-girls ratio in reading (reducing the gender gap) (tables S2, S4, S5, and S8). Consistent with our hypothesis, inequalities are detrimental to the performance of girls relative to boys in the three topics math, reading, and science.

Inequalities and gender gaps among high performers in math

Estimates in column 1 are obtained from separate simple linear country-level regression models on 35 OECD countries in PISA 2015. The dependent variable is the ratio of the numbers of girls and boys among high math performers. Column 2 provides similar estimates based on individual-level regressions with several control variables and adjusted standard errors (see SM). Column 3 is a country fixed-effect version of column 1 based on PISA 2003, 2006, 2009, 2012, and 2015.

EXPLANATORY VARIABLE	COUNTRY-LEVEL REGRESSIONS	INDIVIDUAL-LEVEL LOGISTIC MODELS	COUNTRY FIXED-EFFECT MODELS (2003-2015)
Income inequalities (GINI index)	0.067**	−0.094**	−0.094**
Inequalities in socioeconomic and cultural background	−0.081**	−0.169**	−0.043*
Socioeconomic inequalities in performance	−0.085**	−0.157**	−0.027*
Inequalities in learning opportunities	−0.075**	−0.154**	−0.075

*** P < 0.01, * P < 0.05. See SM for methodological details, the exact definition of inequality variables, sample sizes, standard errors, and R-squares.*

UNDERSTANDING ORIGINS

To suggest causality and better understand the origin of the relationship between non–gender-related inequality measures and gender performance gaps, we offer three different strategies: (i) individual-level regressions, (ii) panel analysis, and (iii) instrumental variables.

We first check that our results are not driven by cross-country differences in observable students' characteristics. We replicate our analysis using the full PISA student-level data, which contains information on students' grade repetition, parents' education, and households' economic and cultural resources. Conditional on these controls and their interaction with students' gender, girls still have a lower probability relative to boys to score above the high-performance cutoff (or increase relatively less) socioeconomic inequalities also reduce relatively more the gender performance gap in math (see the table, column 3, table S11, and fig. S2).

To assess further a possible causal link, we exploit institutional differences between countries' labor markets as instruments for their extent of income inequality. We argue that institutional features such as bargaining coverage, union density, or the value of the minimum wage are not likely to have a direct link with the gender performance gap at school, such that any impact they may have on the math gap would likely be through their impact on social inequalities. To obtain estimates less likely to reflect reverse causality or omitted variables biases, we retrieve variations in the Gini index that are solely explained by those institutional variables. These "instrumented" inequalities, solely driven by institutional factors, affect the gender gap in math to the same extent as noninstrumented inequalities (table S12).

We also study whether institutional features of education systems affect gender performance gaps by considering measures of inequalities in learning opportunities across schools or across students' socioeconomic background, measures of vertical stratification at school, such as the extent of grade repetition, and measures of the quality of education. All those measures, known to affect socioeconomic inequalities at school (*14*, *15*), directly relate to gender performance gaps across countries (tables S2 and S4 to S7).

It is striking that general indicators of inequalities can explain so well the patterns of gender differences in math, science, and reading performance across countries (whereas other indicators directly related to gender stratification have limited explanatory power). In more egalitarian countries, differences in initial status seem less likely to translate into differences in performance, and girls are more represented among high performers as are, for example, students from a low socioeconomic and cultural background. This suggests that the gender gap in math is a form of social inequality like many others.

This is consistent with our results that gender performance gaps at school are linked to countries' institutions that more generally reduce social and economic inequalities. As a consequence, gender equality may not only be a matter of gender norms and stereotypes. General policies in favor of more inclusive, less vertically stratified, and more standardized education systems may also have a positive impact on girls' performance. ■

REFERENCES AND NOTES

1. J. S. Hyde, J. E. Mertz, *Proc. Natl. Acad. Sci.* U.S.A. **106**, 8801 (2009).
2. C. J. Weinberger, *Industrial Relations* **38**, 407 (1999).
3. J. Hunt, J. P. Garant, H. Herman, D. J. Munroe, *Res. Pol.* **42**, 831 (2013).
4. H. Ellis, *Man and Woman: A Study of Secondary and Tertiary Sexual Characters* (Heinemann, London, 1894).
5. S. J. Ceci, W. M. Williams, *Proc. Natl. Acad. Sci.* U.S.A. **108**, 3157 (2011).
6. J. E. Parsons, T. Adler, J. L. Meece, *J. Pers. Soc. Psychol.* **46**, 26 (1984).
7. D. P. Baker, D. Perkins Jones, *Sociol. Educ.* **66**, 91 (1993).
8. L. Guiso, F. Monte, P. Sapienza, L. Zingales, *Science* **320**, 1164 (2008).
9. N. M. Else-Quest, J. S. Hyde, M. C. Linn, *Psychol. Bull.* **136**, 103 (2010).
10. G. Stoet, D. C. Geary, *Intelligence* **48**, 137 (2015).
11. J. M. Kane, J. E. Mertz, *Notices Am. Math. Soc.* **59**, 10 (2012).
12. R. G. Fryer Jr., S. D. Levitt, *Am. Econ. J.: Appl. Econ.* **2**, 210 (2010).
13. H. Ayalon, I. Livneh, *Soc. Sci. Res.* **42**, 432 (2013).
14. W. H. Schmidt, N. A. Burroughs, P. Zoido, R. T. Houang, *Educ. Res.* **44**, 371 (2015).
15. H. G. Van de Werfhorst, J. J. B. Mijs, *Annu. Rev. Sociol.* **36**, 407 (2010)

ACKNOWLEDGMENTS

We are very grateful to J.-M. Marin for his tremendous help with machine-learning techniques; to F. Avvisati, F. Borgonovi, and F. Keslair at the OECD for their support and help regarding how to handle the PISA data; and to A. Aguilar, G. Piaton, and F. Riva for their suggestions on the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1219/suppl/DC1

10.1126/science.aar2307

BOOKS *et al.*

Renowned as a paleontologist, Marie Stopes also championed scientifically sound sex education.

PHOTO: MARIE STOPES INTERNATIONAL AUSTRALIA/WIKIMEDIA COMMONS

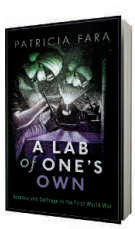
HISTORY OF SCIENCE

Women's work

Stepping into scientific roles left vacant during World War I brought gains for women in and out of the laboratory

By **Laura Micheletti Puaca**

The ability of war to disrupt gender roles is well established. Amid the chaos of fighting, the enlistment of men, and the need to keep home fires burning, women have historically stepped into roles previously considered masculine domains. In World War I–era Britain, women could be found in a variety of new places, such as the munitions factories and military auxiliaries that have attracted the most attention by scholars to date. Less well known is women's work in scientific fields. With *A Lab of One's Own*, Patricia Fara aims to correct this imbalance, masterfully bringing to light women's wartime contributions to these areas, as well as their efforts to improve women's status in both science and society.



A Lab of One's Own
Science and Suffrage in the First World War
Patricia Fara
Oxford University Press, 2018. 350 pp.

During World War I, countless British women designed airplane equipment, invented drugs, tested bombs, studied the effects of tear gas, and cracked codes. They also ran hospitals and science museums and taught at universities. Even seemingly "feminine" fields, such as nutrition, assumed increased importance during the war. As Fara argues, "in wartime, the need to find out why bread goes mouldy was as pressing as the invention of a new explosive."

Although some women received recognition for their work, many did not, a situation that raises much larger questions about the nature of women's wartime gains. On the one hand, many enjoyed increased educational and professional opportunities that had not been made available to them before the war. Yet those opportunities were also circumscribed by gender stereotypes and traditional conventions regarding women's work.

Throughout the war, female scientists were overwhelmingly relegated to the lowest-paid and least-prestigious positions. Much of their work, moreover, involved tasks that were not only tedious but also dangerous, such as in the case of industrial chemists, where women made up almost 90% of their ranks.

Although outright hostility to female scientists seemed to diminish during the war, women continued to be marked as outsiders. Their appearance was routinely scrutinized, and some scientific women also felt pressure to reassure anxious observers of their femininity, as if to offset the masculine nature of their work. The assumption that women in science were temporary participants was made all the more apparent when many male veterans returned to reclaim their jobs.

As Fara demonstrates, however, women's wartime participation was not all for naught. Some were in fact successful in applying their wartime training to postwar careers. Their wartime experiences, coupled with a postwar influx of government money into science, education, and industry, allowed them to enjoy expanded opportunities once again. "They were still not getting their fair share," Fara concedes, "but at least the cake was larger."

Others, regardless of whether they remained in scientific fields, found that their self-confidence and expectations had been irrevocably heightened by their participation in wartime science. Compared with the short-term nature of some gains, this transformation was long-lasting.

Some of the most fascinating parts of the book include analyses of women who blended their work and their commitment to women's rights. During the war, for example, Louisa Garrett Anderson and Flora Murray—who were medical pioneers, suffragettes, and long-term companions—set up a number of hospitals that were staffed entirely by women. Their most famous venture, the Endell Street Hospital in the Bloomsbury section of London, was an explicitly political organization with clear ideological and financial ties to the suffrage movement.

Meanwhile, other scientifically trained women gravitated toward the National Union of Women's Suffrage Societies (NUWSS), which took on important wartime work. For much of the conflict, the NUWSS assumed the role of a patriotic employment agency, which involved, among other things, training women to carry out scientific and technical jobs usually held by men. Unfortunately, despite figuring prominently in the title, suffrage does not get as much attention as one might expect in *A Lab of One's Own*. More of these stories would have enriched the book.

As Fara is careful to note, the degree to which women's wartime contributions influenced the 1918 decision to grant suffrage to women over 30 (subject to other restrictions) is the subject of much debate. What can be said with more certainty is that "public opinion was nudged towards approving female suffrage" in recognition of women's work.

Perhaps a more important issue, though, is the role that suffrage played in improving women's status. Many professional women viewed the vote as the first step in a much longer battle. For them, being paid, promoted, and respected in their fields was just as, if not more, important. Indeed, these concerns remain pressing ones today. ■

The reviewer is in the Department of History, Christopher Newport University, Newport News, VA 23606, USA. Email: laura.puaca@cnu.edu

10.1126/science.aar8211

RESEARCH ETHICS

Life in triplicate

A tale of triplets separated at birth raises red flags and questions about the role of nature and nurture

By **Gabrielle Kardon**

How much are we shaped by our genetic makeup, and how much does the environment in which we are raised contribute to our physical appearance, our personalities, and our preferences? The relative importance of nature versus nurture is a fundamental question in biology and is notably difficult to study in humans, due to ethical concerns.

Three Identical Strangers
Tim Wardle, director
RAW, 2018.
96 minutes.

Analyzing identical twins is one way to examine this question, but most twins are raised in a single household, where the effects of nature and nurture are inextricably intertwined. The story of not two, but three, identical brothers separated at birth is the subject of an insightful and disturbing new documentary, *Three Identical Strangers*, that premiered this January at the 2018 Sundance Film Festival.

In 1980, Robert Shafran arrived at Sullivan County Community College in upstate New York as a new freshman and was greeted by unexpected backslaps from fellow students welcoming him back to school. He realized he was being mistaken for a former student, with whom he quickly arranged a meeting.

Indeed, he and Eddy Galland looked exactly the same, but what's more, they shared a birthday and were both adopted, leading the two to conclude that they must be twins. Newspaper headlines and photos of the pair attracted the attention of yet another curly-haired 19-year-old, David Kellman, and the three discovered that they were triplets separated at birth.

A media frenzy ensued—complete with television interviews with Tom Brokaw and Phil Donahue—and a giddy period of discovery began. Not only were the trio identical physically, they shared many specific behaviors: All were wrestlers, had a preference for dating older women, and (apropos of the time) smoked the same brand of cigarettes. The long-lost brothers bonded, joyfully exploring the 1980s New York City party scene together, and even started a business together—a restaurant aptly named Triplets.

But why were the triplets separated at birth? Their adoption, we learn, was arranged by the preeminent (now defunct) New York–based Louise Wise Adoption Agency, founded in the early 20th century to match Jewish orphans with adoptive families.

In the 1960s, some psychologists believed that separating orphaned twins or triplets afforded a better opportunity for each to forge an independent identity. The triplets, however, were deliberately assigned to particular families as part of a secret study designed to examine the effects of nature versus nurture.

The study, led by child psychoanalyst Peter Neubauer, placed multiple sets of twins and the triplets with families whose socioeconomic status, and even family structure, were carefully controlled. The families were aware of neither the other identical siblings nor their participation in the study.

Yearly comprehensive questionnaires and film documentation of the siblings' milestones were couched as standard follow-up studies of adopted children. As the film unfolds, even more sinister details of the study and its effects on the triplets' lives are revealed.

Five years in the making, *Three Identical Strangers* deftly interweaves interviews with the triplets, their family members, and the few surviving researchers associated with the research. The film tackles both the questionable underpinnings of the study and the fundamental questions it attempted to address.

Current rules of informed consent would now preclude the existence of such a study, and Wardle implies that the moral and ethical issues of the study were likely of concern to Neubauer, who died in 2008, as well as to the Louise Wise Agency. The irony of a Jewish researcher and a Jewish adoption agency conducting a twin study after the atrocities waged against Jewish people in Nazi Germany is clear and perhaps the reason that Neubauer never published the study and sealed his data in a Yale archive until 2066.

Although the study was clearly unpardonable, its questions are enduring. How much are we the product of genetic inheritance? How does the environment shape who we become? To what extent do we have free will?

Eddy Galland (left), David Kellman, and Robert Shafran (right) were placed in separate households as infants.

PHOTO: COURTESY OF NEON

The striking divergence of Shafran, Galland, and Kellman in physical appearance after 1980 confirms that genetics control just part of our makeup, but many questions remain unanswered. Winner of a special Sundance jury award for storytelling, *Three Identical Strangers* will stimulate viewers to think about these important issues. ■

The reviewer is in the Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA. Email: gkardon@genetics.utah.edu

10.1126/science.aat0954

LETTERS

In this watercolor, Clovis Paleoindians journey eastward toward the Americas.

CREDIT: GREG HARLIN/PRIVATE COLLECTION/WOOD RONSAVILLE HARLIN, INC. USA/BRIDGEMAN IMAGES

Edited by **Jennifer Sills**

Arrival routes of first Americans uncertain

In their Perspective "Finding the first Americans" (3 November 2017, p. 592), T. J. Braje *et al.* argue that people first entered the Americas about 25,000 to 15,000 years ago by way of the Pacific coast. We believe that current evidence yields far less certainty than Braje *et al.* suggest—and more likely supports a later arrival by inland and/or coastal routes.

Contrary to Braje *et al.*, genetic evidence indicates that Native American ancestors diverged from Siberian populations ~24,900 to ~18,400 years ago, followed by population expansion into the Americas ~16,000 to ~13,500 years ago (*1*). Only two Native American lineages have been identified south of Beringia: a northern lineage constrained to northern North America, and a southern lineage directly linked with Clovis—the earliest unequivocal widespread cultural manifestation south of the ice sheets (*2*). Thus, a Native American lineage in the Americas between 25,000 and 15,000 years ago is inconsistent with current data. There is no consensus on the validity of purported pre–16,000-year-old sites, which vary in accurate dating, unambiguous artifacts, and clear association between them (*3*). Moreover, there are few technological connections among 16,000- to 13,500-year-old pre-Clovis sites, or with later Paleoindian artifacts; thus, the relationships between these sites and later Native Americans remain ambiguous.

The coastal colonization route Braje *et al.* advocate remains a viable hypothesis for a later arrival date, but several issues should be addressed (*4*). Despite the rise of sea levels in the Holocene, much of the late Pleistocene coast from Puget Sound to Alaska remains above sea level (*5*), yet surveys have failed to discover coastal sites securely dated older than 12,500 years ago (*4*). Furthermore, widespread empirical patterns are inconsistent with the coastal hypothesis: All known populations in Siberia, Russian Far East, and Beringia had terrestrial-oriented economies and technologies, as did widespread Paleoindian groups (*6*, *7*), with limited evidence of coastal exploitation in lower latitudes (*8*). This empirical patterning suggests that they were more likely to follow a land route through Siberia, Beringia, and the Americas south of the ice sheets. Current studies indicate that deglaciation began 19,000 years ago and that an ice-free corridor, largely vegetated and free of proglacial lakes, existed by at least 15,000 to 14,000 years ago (*9*, *10*).

Braje *et al.* suggest that stemmed projectile points in different contexts provide evidence for a coastal expansion before 16,000 years ago, but this is not a consensus view. Stemming is a widespread form of weapon design that was innovated numerous times and thus cannot be used to argue for cultural affiliation. No technological analysis has established a valid connection between these disparate assemblages, and there remains debate on the dating of North American stemmed points (*11*) because most securely dated sites are younger than Clovis.

Braje *et al.* assert that there is near complete agreement among archaeologists on these issues. However, the most recent survey (*12*) showed that archaeologists are divided, with many thinking that both interior and coastal routes were used, and expressing skepticism about several proposed pre-Clovis sites. Genetic and archaeological data suggest expansion from Siberia into the Americas around 16,000 to 13,500 years ago, consistent with terrestrial and/or coastal migrations. This evidence base explains the absence of consensus among scientists regarding both routes and timing of the peopling of the Americas.

Ben A. Potter,[1]* Alwynne B. Beaudoin,[2] C. Vance Haynes,[3] Vance T. Holliday,[3] Charles E. Holmes,[1] John W. Ives,[4] Robert Kelly,[5] Bastien Llamas,[6] Ripan Malhi,[7] Shane Miller,[8] David Reich,[9,10,11] Joshua D. Reuther,[1,12] Stephan Schiffels,[13] Todd Surovell[5]

[1]Department of Anthropology, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. [2]Royal Alberta Museum, Edmonton, AB T5J 0G2, Canada. [3]School of Anthropology and Department of Geosciences, University of Arizona, Tucson, AZ 85721, USA. [4]Institute of Prairie Archaeology, University of Alberta, Edmonton, AB T6G 2R3, Canada. [5]Department of Anthropology, University of Wyoming, Laramie, WY 82071, USA. [6]Australian Centre for Ancient DNA, Environment Institute, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia. [7]Department of Anthropology, University of Illinois, Urbana, IL 61801, USA. [8]Department of Anthropology and Middle Eastern Cultures, Mississippi State University, Starkville, MS 39759, USA. [9]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [10]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. [11]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [12]Archaeology Department, University of Alaska Museum of the North, Fairbanks, AK 99775, USA. [13]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany.
*Corresponding author. Email: bapotter@alaska.edu

REFERENCES

1. B. Llamas *et al.*, *Sci. Adv.* **2**, e1501385 (2016).
2. M. Rasmussen *et al.*, *Nature* **506**, 225 (2014).
3. T. Goebel, M. R. Waters, D. H. O'Rourke, *Science* **319**, 1497 (2008).
4. B. A. Potter *et al.*, *Quat. Int.* **444**, 36 (2017).
5. D. H. Shugar *et al.*, *Quat. Sci. Rev.* **97**, 170 (2014).

6. T. Goebel, *Archeol. Pap. Am. Anthropol. Assoc.* **12**, 117 (2002).
7. L. C. Bement, B. J. Carter, in *Clovis: On the Edge of a New Understanding*, A. M. Smallwood, T. A. Jennings, Eds. (Texas A&M Press, College Station, 2015), pp. 263–276.
8. J. M. Erlandson, *Early Hunter-Gatherers of the California Coast* (Springer, 2013).
9. K. Munyikawa, T. M. Rittenour, J. K. Feathers, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **470**, 147 (2017).
10. D. H. Huntley, A. S. Hickin, O. B. Lian, *Can. J. Earth Sci.* **54**, 52 (2017).
11. T. Goebel, J. L. Keene, in *Archaeology in the Great Basin and Southwest: Papers in Honor of Don D. Fowler*, J. Parezo, J. C. Janetski, Eds. (University of Utah Press, Salt Lake City, UT, 2014), pp. 35–60.
12. A. Wheat, *SAA Rec.* **12**, 10 (2012).

10.1126/science.aar8233

Response

In our Perspective, we argue that the earliest Americans followed a coastal migration route. Potter *et al.* counter that an inland, ice-free corridor route was more likely. The genetic and archaeological evidence that Potter *et al.* discuss support both the timing and coastal route for the peopling of the Americas presented in our Perspective. Potter *et al.* are disappointed, however, that we did not also emphasize the possibility of an interior ice-free corridor migration and more tightly constrain colonization to after 16,000 years ago. This misses the main points of our Perspective: A Pleistocene coastal migration into the Americas has been bolstered tremendously by the discovery of early coastal and near coastal sites in the Americas, and more research, particularly underwater archaeology, is badly needed.

Potter *et al.* imply that in the most recent survey "archaeologists were divided, with many thinking both interior and coastal routes were used." In fact, 86% of respondents selected "coastal migration," compared with 65% for "interior passage migration" (*1*). This represents a sea change from 30 years ago when funding dollars and research efforts were funneled toward ice-free corridor models of New World colonization, resulting in decades of marginalization of research into early coastal migrations and adaptations (*2*).

Potter *et al.* underestimate the complexity of finding intact late Pleistocene coastal sites. Even in geological contexts where Pleistocene shorelines remain aboveground (such as in Alaska and British Columbia), coastal erosion and other forces tend to ravage or obscure early sites. In contrast to the sustained and intensive efforts to identify early sites in interior regions, the search for Pleistocene sites in coastal settings is just getting started. A growing number of such sites have been identified in places where systematic efforts to find late Pleistocene to early Holocene coastal sites have been made: the Pacific Northwest, the islands of Alta and Baja California, and coastal Peru and Chile (*3*–*6*).

Recent genomic evidence suggesting a migration into the Americas from Northeast Asia between ~16,000 and 13,500 years ago does not preclude a coastal migration, and ages based on molecular clocks are approximations due to unpredictable variation in mutation rates. Potter *et al.* suggest that we argue for a migration earlier than 16,000 years ago based on the distribution of stemmed projectile points. We did argue that northeast Asian stemmed point technologies may support Native American origins sometime in the past 20,000 years—like genetic data—but that more work is needed to link these Asian technologies to similar but younger terminal Pleistocene toolkits in the Americas.

Potter *et al.* subscribe to a window of colonization between about 16,000 and 13,500 years ago, the latter half of which is almost certainly incorrect given the widely accepted 14,500-year-old occupation at Monte Verde (*7*). A coastal route was open by ~17,000 years ago, whereas luminescence dates suggest that the ice-free corridor was open and viable by 15,000 to 14,000 years ago (*8*), or slightly later (*9*). The latter route would have allowed little time for human migrants to reach and establish residency at terminal Pleistocene sites in Oregon (*10*), Florida (*11*), Peru (*4*), Chile (*7*), and other far-flung regions.

The timing and viability of the ice-free corridor route remains uncertain, and arguments that the initial peopling of the Americas followed a terrestrial route hinge on its reconstruction. Questions remain about the coastal migration theory as well, and current evidence suggests that there likely were multiple dispersals, routes, and time frames for the peopling of the Americas.

Todd J. Braje,[1]* Torben C. Rick,[2] Tom D. Dillehay,[3] Jon M. Erlandson,[4] Richard G. Klein[5]

[1]Department of Anthropology, San Diego State University, San Diego, CA 92182, USA. [2]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA. [3]Department of Anthropology, Vanderbilt University, Nashville, TN 37240, USA. [4]Department of Anthropology and Museum of Natural and Cultural History, University of Oregon, Eugene, OR 97403, USA. [5]Departments of Anthropology and Biology, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: tbraje@mail.sdsu.edu

REFERENCES

1. A. Wheat, *SAA Archaeol. Rec.* **12**, 10 (2012).
2. J. M. Erlandson, *J. Archaeol. Res.* **9**, 287 (2001).
3. M. R. Des Lauriers, *Island of Fogs: Archaeological and Ethnohistorical Investigations of Isla Cedros, Baja California* (University of Utah Press, Salt Lake City, UT, 2010).
4. T. D. Dillehay *et al.*, *Sci. Adv.* **3**, e1602778 (2017).
5. J. M. Erlandson *et al.*, *Science* **331**, 1181 (2011).
6. D. H. Sandweiss, in *Handbook of South American Archaeology*, H. Silverman, W. Isbell, Eds. (Springer Press, Berlin, 2008), pp. 145–156.
7. T. D. Dillehay *et al.*, *PLoS ONE* **10**, https://doi.org/10.1371/journal.pone.0141923 (2015).
8. K. Munyikawa, T. M. Rittenour, J. K. Feathers, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **470**, 147 (2017).
9. M. W. Pedersen *et al.*, *Nature* **537**, 45 (2016).
10. D. L. Jenkins *et al.*, *Science* **337**, 223 (2012).
11. J. J. Halligan *et al.*, *Sci. Adv.* **2**, e1600375 (2016).

10.1126/science. aar8645

Postpublication peer review: A crucial tool

The current peer-review model used throughout science is not perfect (*1*). Whether it be the result of poor experimental design, accident, or academic misconduct, publication of irreproducible, incorrect, or fabricated results occurs more frequently than it should [check Retraction Watch for recent examples (*2*)]. This leads not only to a waste of precious time and financial resources as scientists try to replicate or build on flawed research but also to damage to the reputation of science and to much larger societal impacts (such as the loss of public trust in science and loss of federal funding).

An emerging online tool for combating these issues is postpublication peer review (PPPR). PPPR sites such as F1000, ResearchGate, PubPeer, and PubMed Commons, as well as *Science*'s own eLetters, provide environments for user comments and discussion and are responsible for catching flawed research that has slipped through traditional peer review (*3*). In addition to identifying fraudulent data, pointing out errors, and providing criticism (which generally take the form of negative comments), PPPR also enables positive feedback (such as verifying the reproducibility of results), which is valuable but is currently provided much less frequently (*4*). This disparity likely stems from the reality that overworked scientists do not have time for activities that provide little to no recognition (*5*). However, most scientists already participate in informal (offline) PPPR. We discuss the results of papers with our colleagues, present papers in group meetings, and critically analyze papers in journal clubs. With a little more effort, a formal record of our reviews (negative and positive) can be made online for the betterment of science.

There is a risk to publicly challenging the work of established scientists (in particular, younger scientists may face retribution), but it can be mitigated by providing feedback in a respectful, positive, and professional manner (*4*, *6*). These are risks that need to be taken. The scientific community needs to take action to maintain the integrity of our published work. With continued implementation and development (*7*),

PPPR can become a new cornerstone in the self-correcting mechanism of science.

Gregory Isaac Peterson
Seoul National University, Seoul, Republic of Korea.
Email: gpeterson@snu.ac.kr

REFERENCES

1. R. Smith, *J. R. Soc. Med.* ***99***, 178 (2006).
2. Retraction Watch (http://retractionwatch.com).
3. P. Knoepfler, *Trends Gent.* ***31***, 221 (2015).
4. H. Bastian, *PLOS Med.* ***11***, e1001772 (2014).
5. M. Markie, *Insights* ***28***, 107 (2015).
6. E. Pain, *Science Careers* (2013); www.sciencemag.org/careers/2013/04/interactive-peer-review-what-s-it-reviewers.
7. J. A. Teixeira da Silva, A. Al-Khativ, J. Dobranszki, *Sci. Eng. Ethics* ***23***, 1213 (2017).

10.1126/science.aas9490

TECHNICAL COMMENT ABSTRACTS

Editor's note: The Comment by DeSilva and Response by Rosas *et al.* were published online in the 9 March issue.

Comment on "The growth pattern of Neandertals, reconstructed from a juvenile skeleton from El Sidrón (Spain)"

Jeremy M. DeSilva

Rosas *et al.* (Reports, 22 September 2017, p. 1282) calculate El Sidrón J1 to have reached only 87.5% of its adult brain size. This finding is based on an overestimation of Neandertal brain size. Pairwise comparisons with a larger sample of Neandertal fossils reveal that it is unlikely that the brain of El Sidrón would have grown appreciably larger.

Full text: dx.doi.org/10.1126/science.aar3611

Response to Comment on "The growth pattern of Neandertals, reconstructed from a juvenile skeleton from El Sidrón (Spain)"

Antonio Rosas, Luis Ríos, Almudena Estalrrich, Helen Liversidge, Antonio García-Tabernero, Rosa Huguet, Hugo Cardoso, Markus Bastir, Carles Lalueza-Fox, Marco de la Rasilla, Christopher Dean

The comment by DeSilva challenges our suggestion that brain growth of the El Sidrón J1 Neandertal was still incomplete at 7.7 years of age. Evidence suggests that endocranial volume is likely to represent less than 90% adult size at El Sidrón as well as Neandertal male plus Krapina samples, in line with further evidence from endocranial surface histology and dural sinus groove size.

Full text: dx.doi.org/10.1126/science.aar3820

Comment on "Synthesis and characterization of the pentazolate anion *cyclo*-N_5^- in $(N_5)_6(H_3O)_3(NH_4)_4Cl$"

Rong-Yi Huang and Heng Xu

Zhang *et al.* (Reports, 27 January 2017, p. 374) reported synthesis of a *cyclo*-N_5^- ion putatively stabilized in a solid-state salt by hydrogen bonding from surrounding counterions. We performed theoretical calculations suggesting that HN_5 would be favored over the anion in the reported pentazolate salt via proton transfer.

Full text: dx.doi.org/10.1126/science.aao3672

Response to Comment on "Synthesis and characterization of the pentazolate anion *cyclo*-N_5^- in $(N_5)_6(H_3O)_3(NH_4)_4Cl$"

Chao Jiang, Lei Zhang, Chengguo Sun, Chong Zhang, Chen Yang, Jun Chen, Bingcheng Hu

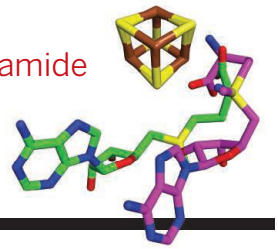
Huang and Xu argue that the *cyclo*-N_5^- ion in $(N_5)_6(H_3O)_3(NH_4)_4Cl$ we described in our Report is theoretically unfavorable and is instead protonated. Their conclusion is invalid, as they use an improper method to assess the proton transfer in a solid crystal structure. We present an in-depth experimental and theoretical analysis of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ that supports the results in the original paper.

Full text: dx.doi.org/10.1126/science.aas8953

RESEARCH

Dissecting diphthamide biosynthesis

Dong et al., *p. 1247*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

FISHERIES

Healthy fisheries can reduce bycatch

Bycatch of marine mammals, turtles, and birds during commercial fishing is a considerable threat. Activities intended to reduce bycatch are often thought to conflict with commercial fishing. However, Burgess *et al.* show that in the majority of cases, managing fishery stocks to best promote long-term sustainability would also reduce bycatch. Rebuilding fish stocks will naturally promote lower bycatch, and these factors together will facilitate sustainable profit generation from fish harvest. —SNV

Science, this issue p. 1255

Fishing may accidentally capture endangered species such as this green sea turtle.

CREDITS: (TOP TO BOTTOM) M. DONG *ET AL.*; JEFF ROTMAN/ALAMY STOCK PHOTO

INDUCED SEISMICITY

Injection depth matters for induced earthquakes

Wastewater injection has induced earthquakes in Oklahoma, but the relative importance of operational and geologic parameters in triggering such earthquakes is unclear. Hincks *et al*. developed an advanced Bayesian network to determine the interplay between these parameters in Oklahoma. The injection depth above the crystalline basement was the most important parameter when considering the potential for release of seismic energy. This modeling strategy may provide a way to improve forecasts of the impact of proposed regulatory changes on induced seismicity. —BG

Science, this issue p. 1251

SURFACE CHEMISTRY

Watching graphene grow

The growth of graphene on metal surfaces can be catalyzed by mobile surface metal atoms. Patera *et al*. used a high-speed scanning tunneling microscope to image the growth of graphene islands on a nickel surface. High temperatures caused carbon to diffuse to the surface, where mobile nickel atoms catalyzed graphene growth on the edges of islands. Molecular dynamics and density functional theory calculations provide mechanistic insights into the reaction steps. —PDS

Science, this issue p. 1243

POLYMERS

Mimicking the designs found in proteins

Natural proteins combine a range of useful features, including chemical diversity, the ability to rapidly switch between preprogrammed shapes, and a hierarchy of structures. Panganiban *et al*. designed random copolymers with polar and nonpolar groups, using many of the features found in proteins (see the Perspective by Alexander-Katz and Van Lehn). Their structures could serve as "broad spectrum" surfactants, able to promote the solubilization of proteins in organic solvents and help preserve the functionality of proteins in aqueous environments. —MSL

Science, this issue p. 1239; see also p. 1216

COGNITIVE SCIENCE

The infant as philosopher

Visual behaviors, such as a shift in one's gaze or a prolonged stare, can be diagnostic of internal thoughts. Cesana-Arlotti *et al*. used these measures to

demonstrate that preverbal infants can formulate a logical structure called a disjunctive syllogism (see the Perspective by Halberda). That is, if A or B is true, and A is false, then B must be true. Presenting infants with scenes where the outcome revealed B to be false evoked looks of surprise. —GJC and AMS

Science, this issue p. 1263; see also p. 1214

NEURODEVELOPMENT

Call to action

The developing brain initially makes more synapses than it needs. With further development, excess synapses are pruned away, leaving mature circuits. Synapses can be eliminated by microglia, which engulf and destroy them. Vainchtein *et al*. found that the microglia are called into action by astrocytes, supportive cells on which neurons rely. Astrocytes near a redundant synapse release the cytokine interleukin-33 (IL-33), which recruits microglia to the site. In mice, disruptions in this process, as caused by deficiency in IL-33, led to too many excitatory synapses and overactive brain circuitry. —PJH

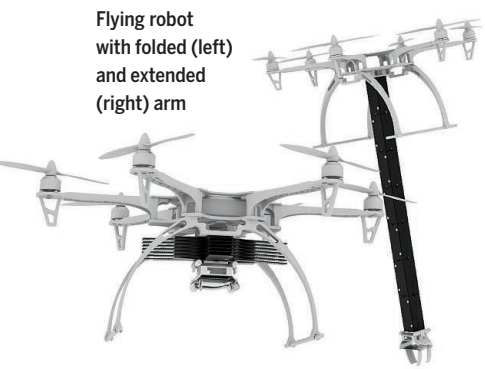
Science, this issue p. 1269

SOFT ROBOTS

Robots reach out

Unmanned aerial vehicles (UAVs) sometimes need to reach into confined spaces. Previous foldable robotic arms have been too flexible to be practical. Kim *et al.* created a modular, origami-inspired robotic arm with a locking mechanism based on the idea that when two folds intersect at a right angle, only one can be folded at a time. This allowed tendon-based actuation by a single electric motor and enabled a UAV to grasp objects in a narrow space and insert a camera among tree branches. —RLK

Sci. Robot. **3**, eaar2915 (2018).



Flying robot with folded (left) and extended (right) arm

AGING

Lysosomes keep neuronal stem cells young

An important consequence of aging is loss of regenerative capacity in stem cells, particularly those of the nervous system. Leeman *et al*. isolated quiescent and activated stem cells from mice and compared their transcriptomes. The findings emphasize the role of large lysosomes in quiescent neuronal stem cells in which aggregated proteins accumulate. Treatments that stimulated lysosomal function allowed aged quiescent stem cells to clear protein aggregates and restored the cells' ability to be activated. Such restoration of stem cell function might alleviate compromised proteostasis in aging. —LBR

Science, this issue p. 1277

NEUROTECHNOLOGY

Good vibrations for movement perception

Because prostheses do not provide physical feedback during movement, amputees may not feel that they are in full control of their bodily action. Marasco *et al.* developed an automated neural-machine interface that vibrates the muscles used for the control of prosthetic hands. This system instilled kinesthetic sense in amputees, allowing them to control prosthetic hand movements in the absence of visual feedback and increasing their sense of control. —MM

Sci. Transl. Med. **10**, eaao6990 (2018).

IN OTHER JOURNALS

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

VOCAL BEHAVIOR

Speaking in parts

The complexity of human speech is an anomaly among primates, where calls have typically been thought of as stereotyped and not variable. Recent work has shown that there is more learning and control present in nonhuman primate vocalizations than previously believed, but adult calls have nevertheless seemed to be stable. Pomberger *et al.* interrupted calling marmosets with acoustic playbacks triggered by their own calls and found that the animals could extend their "phee" calls beyond normal lengths by restarting them rapidly after perturbation, but they could only alter their calls at periodic time points. These results suggest that the phee call is not one long stereotyped series, but rather is composed of small sequential segments, like human speech. —SNV

Curr. Biol. 10.1016/j.cub.2018.01.070 (2018).

Marmoset phee calls occur in units, like human speech.

MICROBIOTA

Predicting fecal transplant success

Consumption of fecal matter is taboo, unless you have a life-threatening disease, such as *Clostridium difficile* infection. In these cases, fecal microbiota transplantation (FMT) is extraordinarily effective at restoring gut function to a healthier state. However, FMT does not work so well for other indications that are less strongly associated with the microbiota. We have a weak understanding of what factors allow microbial engraftment in the gut. Smillie *et al.* took genomic data from FMT-treated *C. difficile* patients to build a machine-learning statistical model that tells us which bacterial strains will engraft. The main predictive factors are abundance of a strain in the donor and the species present in the patient. Donor strains whose species are also present in the patient always engraft. If the taxa are discordant, then it appears that microbial interactions prevent engraftment. —CA

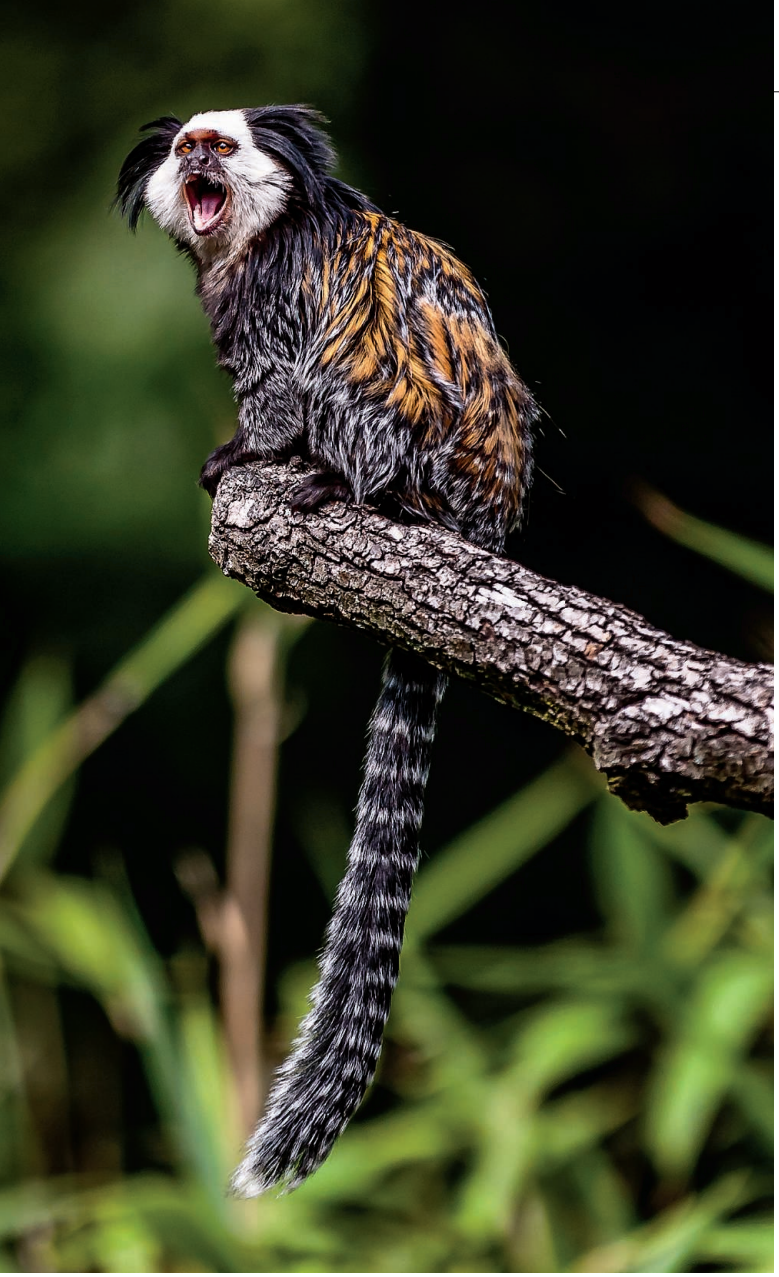
Cell Host Microbe 10.1016/j.chom.2018.01.003 (2018).

MOLECULAR BIOLOGY

Secrets of CRISPR enzymes revealed

CRISPR-Cas9 cleaves DNA, enabling genome editing. Surprisingly, three studies have

PHOTOS: (LEFT TO RIGHT) JAMIE JUNG-RYUL SONG/SNU SOFT ROBOTICS RESEARCH CENTER; SAAD315/SHUTTERSTOCK



concurrently uncovered an intrinsic RNA cutting activity of Cas9. Cas9 cuts endogenous RNA in some bacteria and, when programed, targets RNA in test tubes as well. Strutt *et al.* showed that Cas9 expression in a heterologous bacterium protects against infection by an RNA virus. Rousseau *et al.* showed that the activity can be switched off by virus-derived anti-Cas9 proteins. Dugar *et al.* profiled the cellular RNAs that are subjected to endogenous Cas9 targeting. Although how it is used to fend off intruders or regulate gene expression by bacteria remains to be investigated, this newly identified activity of Cas9 provides an exciting tool in our increasing arsenal for modulating RNA metabolism. —SYM

eLife 10.7554/eLife.32724 (2018); *Mol. Cell.* 10.1016/j.molcel.2018.01.025, 10.1016/j.molcel.2018.01.032 (2018).

LINGUISTICS

Tracking expanding dimensions of words

Language can evolve by adding new meanings or "senses" to old words. Thus, the word "face" can refer to the body part or the front surface of a cliff. There are several theories regarding how new senses of words arise historically. Ramiro *et al.* used an algorithmic approach to study 5000 English words taken from a digital historical dictionary that covers word usage from Anglo-Saxon times to the present. They found that one model—nearest-neighbor chaining—predominates, possibly because it minimizes the cognitive costs of communicating new ideas. —BJ

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1714730115 (2018).

TUMOR IMMUNOLOGY

Tumor sensing by the immune system

The idea that innate lymphoid cells (ILCs) play a role in cancer is of interest because these immune cells respond to IL-33, a cytokine proposed to regulate cancer development. Saranchova *et al*. report that the subgroup called ILC2s may limit tumor growth rate and metastasis. In a mouse model engineered to lack ILC2s, lung cancer cells implanted into the leg enlarged more rapidly than in mice expressing ILC2s. Moreover, in the absence of ILC2, cancer cells spread to other tissues, including the lung, brain, and adrenal gland. The authors propose that IL-33 released by the cancer cells locally stimulates ILC2s, which in turn produce other cytokines that promote an immune response to cancer, limiting its growth and spread. —LC

Sci. Rep. 10.1038/s41598-018-20608-6 (2018).

PLANT MORPHOLOGY

Plant puzzle patterns

The epidermal cells of plants have a jigsaw-like appearance when viewed microscopically. What is the function of this intricate pattern, and how does it arise? Sapala *et al.* suggest that the interlocking arrangement reduces mechanical stress in the cell wall—stress that arises from the turgor pressure from within. Simulation modeling and observations of several plant species and organs show that as growth proceeds, stress causes localized cellulose deposition, leading to small surface indentations in cell walls that gradually develop into the jigsaw pattern. —AMS

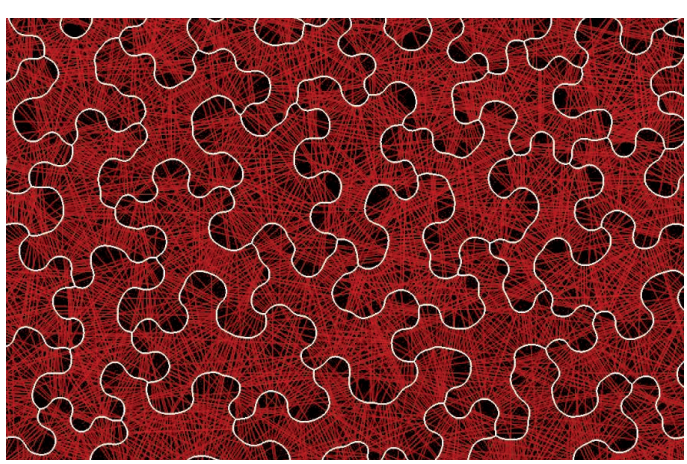
eLife 10.7554/eLife.32794 (2018).

CLIMATE CHANGE

Seeing the human hand

Environmental temperature changes occur both naturally and because of human activity (anthropogenic global warming), and distinguishing those two types is difficult. Wills *et al.* present a method that can be used to make that distinction for Pacific sea surface temperatures, which allows them to identify the separate effects of global warming, the Pacific Decadal Oscillation (PDO), and the El Niño–Southern Oscillation. They do this on the basis of differences in time scale and spatial pattern, which enables them to circumvent the use of climate models. Among other results, they find that midlatitude PDO anomalies are more persistent than previously thought. —HJS

Geophys. Res. Lett. 10.1002/2017GL076327 (2018).



Jigsaw patterning in plant cells counters mechanical stress.

CREDIT: A. SAPALA *ET AL.*, *ELIFE* 10.7554/ELIFE.32794 (2018)

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

TOPOLOGICAL PHOTONICS

Topological protection for lasers

Ideas based on topology, initially developed in mathematics to describe the properties of geometric space under deformations, are now finding application in materials, electronics, and optics. The main driver is topological protection, a property that provides stability to a system even in the presence of defects. Harari *et al*. outline a theoretical proposal that carries such ideas over to geometrically designed laser cavities. The lasing mode is confined to the topological edge state of the cavity structure. Bandres *et al*. implemented those ideas to fabricate a topological insulator laser with an array of ring resonators. The results demonstrate a powerful platform for developing new laser systems. —ISO

Science, this issue p. 1230, p. 1231

CIRCADIAN RHYTHMS

Daily transcription cycling in the baboon

Much of our knowledge about the important effects of circadian rhythms in physiology comes from studies of mice, which are nocturnal. Mure *et al*. report transcriptional profiles from many tissues and brain regions in baboons over a 24-hour period (see the Perspective by Millius and Ueda). The results emphasize how extensive rhythmic expression is, with more than 80% of protein-coding genes involved. They also highlight unanticipated differences between the mouse and baboon in the cycling of transcripts in various tissues. The findings provide a comprehensive analysis of circadian variation in gene expression for a diurnal animal closely related to humans. —LBR

Science, this issue p. 1232; see also p. 1210

HUMAN GENOMICS

Hidden effects of Mendelian inheritance

Identifying the determinate factors of genetic disease has been quite successful for Mendelian inheritance of large-effect pathogenic variants. In these cases, two non- or low-functioning genes contribute to disease. However, Mendelian effects of lesser strength have generally been ignored when looking at genomic consequences in human health. Bastarache *et al*. used electronic records to identify the phenotypic effects of previously unidentified Mendelian variations. Their analysis suggests that individuals with undiagnosed Mendelian diseases may be more prevalent in the general population than assumed. Because of this, genetic analysis may be able to assist clinicians in arriving at a diagnosis. —LMZ

Science, this issue p. 1233

ENZYMOLOGY

A quick freeze shows an enzyme's secrets

Organic radicals are chemically useful in enzymatic reactions but are often hard to observe, owing to their short lifetimes. Dong *et al*. used rapid freeze-quench methods to trap two intermediates formed by a noncanonical radical S-adenosylmethionine (SAM) enzyme: a fragmented SAM molecule bound to the iron-sulfur cluster through an iron-carbon bond and a product-like radical. The structure of the SAM-bound enzyme reveals a noncolinear arrangement of carbon, sulfur, and iron atoms. The arrangement of bonds suggests that the organometallic intermediate may be created through a two-electron nucleophilic mechanism. A subsequent radical intermediate is formed on the protein substrate and resolves by oxidation to form the amino acid product diphthamide. —MAF

Science, this issue p. 1247

NITROGEN CYCLE

Stream physics set the limits

A combination of physical transport processes and biologically mediated reactions in streams and their sediments removes dissolved inorganic nitrogen (DIN) from the water. Although stream chemistry and biology have been considered the dominant controls on how quickly DIN is removed, Grant *et al*. show that physics is what sets the limits on removal rates of nitrate (a component of DIN). Residence time in the hyporheic zone (the region below the sediment surface where groundwater and surface water mix) determines the maximum rate at which nitrate can be removed from stream water. Nevertheless, at local scales, chemistry and biology modify how closely to that maximum rate removal occurs. —HJS

Science, this issue p. 1266

MALARIA

Sexual development in *Plasmodium*

Malaria-causing parasites (*Plasmodium*) have complex life histories in the tissues of humans. For the most part, the parasites focus their efforts on replication within the human host cells. However, occasionally, some replicating cells release gametes into the bloodstream, which are picked up by biting mosquitoes. Filarsky *et al*. discovered that the *Plasmodium* parasite keeps the production of gametes under tight epigenetic control using heterochromatin protein 1 (HP1). *Plasmodium* gametocytogenesis is initiated when HP1 is evicted from upstream of gamete-specific genes by gametocyte development 1 (GDV1) protein. GDV1 is in turn regulated by its antisense RNA. What triggers GDV1 expression remains unclear. Elucidating this pathway could provide a target for interrupting malaria transmission. —CA

Science, this issue p. 1259

CIRCADIAN RHYTHMS

Chromosome dynamics and cellular clocks

Many genes undergo daily or circadian changes in their rate of transcription. Kim *et al*. explored the mechanism by which the circadian clock is linked to chromosome dynamics (see the Perspective by Diettrich Mallet de Lima and Göndör). They used a chromosome conformation capture technique (Hi-C) to identify the interactions of adjacent DNA fragments and determine how DNA looping that altered such interactions changed over daily cycles. The repressive transcription factor Rev-erbα, which functions as part of the mammalian clock mechanism, appears to bind to chromatin and recruit a protein complex that evicts other proteins that enhance looping, thus favoring enhancer-promoter interactions. —LBR

Science, this issue p. 1274; see also p. 1212

SINGLE-CELL ANALYSIS

Mutation rates and effects in single cells

Understanding the dynamics of mutations and the distribution of fitness effects is critical for most evolutionary models. Robert *et al*. used a single-cell technology to visualize the accumulation of new mutations. The method identifies DNA sequences with mispaired bases and small insertions or deletions caused by DNA replication errors in living *Escherichia coli* cells. Following the fates of cells after mutation allowed for a precise quantification of the effects of new mutations.

A smaller fraction of mutations were found to be deleterious than predicted previously from indirect observations. —LMZ

Science, this issue p. 1283

CANCER THERAPY

A new target induces synthetic lethality

Identifying mutations that can be targeted to cause tumor cell death selectively in a process known as synthetic lethality has become a paradigm of cancer research. Loss of DNA repair can provide a selective growth advantage to tumor cells by enhancing mutation rates and driving tumor evolution. In a Perspective, Higgins and Boulton discuss how the DNA damage response protein known as DNA polymerase θ is increased in certain tumors, such as those found in breast and ovarian cancers. This pathway represents an exciting anticancer target that could perhaps be combined with radiotherapy to improve therapeutic response. —GKA

Science, this issue p. 1217

ECOLOGY

How animals navigate infection risk

Ecologists have long recognized how fear of predators affects animal behavior. In a Perspective, Weinstein *et al.* highlight the importance of another factor: infection risk. Using examples from a range of animals, they show how the disgust triggered by potential sources of infection affects animal behavior. Understanding how animals respond to the overlapping landscapes of predation and infection risk is crucial for predicting how ecosystems respond to changing environments. —JFU

Science, this issue p. 1213

IMMUNOLOGY

A psoriasis target in Langerhans cells

Psoriasis is an autoimmune skin condition that is linked to the proinflammatory cytokine interleukin-17 (IL-17), which is produced by T cells in the skin. Zheng *et al.* found that p38α signaling in skin-resident dendritic cells (Langerhans cells) was important for the pathogenesis of psoriasis in a mouse model of the disease. p38α signaling in Langerhans cells stimulated the production of IL-23, which is critical for the development of IL-17–producing T cells. Genetic deletion or pharmacological inhibition of p38α reduced skin inflammation in mice with established psoriatic disease. —ERW

Sci. Signal. **11**, eaao1685 (2018).

INFECTIOUS DISEASE

Structural insights into RSV

Respiratory syncytial virus (RSV) causes respiratory infections associated with severe morbidity and mortality in infants, young children, and the elderly. No licensed vaccine against this disease exists. Fedechkin *et al.* describe cocrystal structures of the RSV G glycoprotein conserved central domain (CCD) bound by two different broadly neutralizing monoclonal antibodies (mAbs). Both mAbs bind to conformational epitopes on this highly conserved region. The RSV G CCD can activate the chemokine receptor CX3CR1, and this activity can be blocked by binding of the mAbs. These findings provide insight into how neutralizing mAbs interact with the RSV G glycoprotein and may advance the development of RSV vaccines and therapeutics. —CNF

Sci. Immunol. **3**, eaar3534 (2018).

NEUROSCIENCE

MHCI suppresses relapse

Relapse is a persistent problem in the treatment of drug addiction and carries a great personal and societal cost. Murakami *et al.* discovered a molecular mechanism that suppresses relapse to cocaine addiction in mice. In a transgenic mouse model, cocaine self-administration reduced major histocompatibility complex class I (MHCI) levels in the brain's reward system, enhanced glutamate signaling, and led to behavioral relapse. Up- or down-regulating genes that influence the function of MHCI in dopamine neurons prevented or promoted this addictive profile. —KSL

Sci. Adv. 10.1126/sciadv.aap7388 (2018).

RESEARCH ARTICLE SUMMARY

TOPOLOGICAL PHOTONICS

Topological insulator laser: Theory

Gal Harari,* Miguel A. Bandres,* Yaakov Lumer, Mikael C. Rechtsman, Y. D. Chong, Mercedeh Khajavikhan, Demetrios N. Christodoulides, Mordechai Segev†

INTRODUCTION: Topological insulators emerged in condensed matter physics and constitute a new phase of matter, with insulating bulk and robust edge conductance that is immune to imperfections and disorder. To date, topological protection is known to be a ubiquitous phenomenon, occurring in many physical settings, ranging from photonics and cold atoms to acoustic, mechanical, and elastic systems. So far, however, most of these studies were carried out in entirely passive, linear, and conservative settings.

RATIONALE: We propose topological insulator lasers: lasers whose lasing mode exhibits topologically protected transport without magnetic fields. Extending topological physics to lasers is far from natural. In fact, lasers are built on foundations that are seemingly inconsistent with the essence of topological insulators: They require gain (and thus are non-Hermitian), they are nonlinear entities because the gain must be saturable, and they are open systems because they emit light. These properties, common to all lasers, cast major doubts on the possibility of harnessing topological features to make a topological insulator laser. Despite this common mindset, we show that the use of topological properties leads to highly efficient lasers, robust to defects and disorder, with single-mode lasing even at conditions high above the laser threshold.

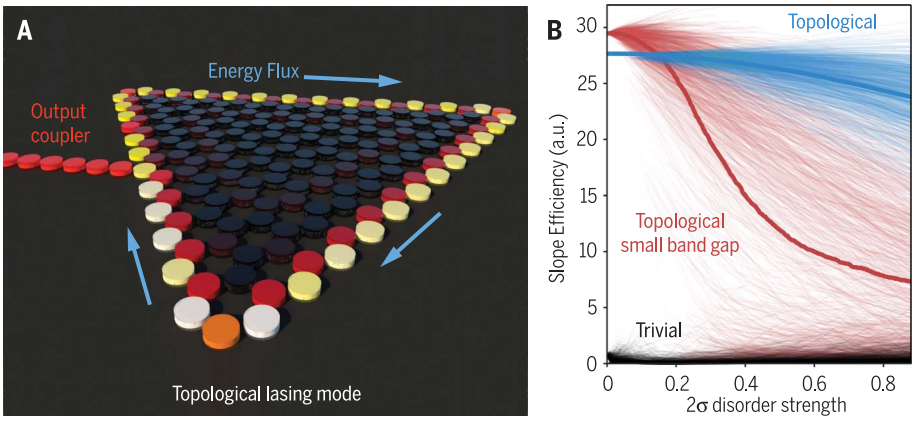
RESULTS: We demonstrate that topological insulator lasers are theoretically possible and experimentally feasible. We consider two configurations involving planar arrays of coupled active resonators. The first is based on the Haldane model, archetypical for topological systems. The second model, geared toward experiment, constitutes an aperiodic array architecture creating an artificial magnetic field. We show that by introducing saturable gain and loss, it is possible to make these systems lase in a topological edge state. In this way, the lasing mode exhibits topologically protected transport; the light propagates unidirectionally along the edges of the cavity, immune to scattering and disorder, unaffected by the shape of the edges. Moreover, we show that the underlying topological properties not only make the system robust to fabrication and operational disorder and defects, they also lead to a highly efficient single-mode lasing that remains single-mode even at gain values high above the laser threshold.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aar4003

The figure describes the geometry and features of a topological insulator laser based on the Haldane model while adding saturable gain, loss, and an output port. The cavity is a planar honeycomb lattice of coupled microring resonators, pumped at the perimeter with a lossy interior. We show that under these conditions, lasing occurs at the topological edge mode, which has unidirectional flux and is extended around the perimeter with almost-uniform intensity. The topological cavities exhibit higher efficiency than the trivial cavity, even under strong disorder. For the topological laser with a small gap, the topological protection holds as long as the disorder level is smaller than the gap size.

DISCUSSION: The concept of the topological insulator laser alters current understanding of the interplay between disorder and lasing, and opens exciting possibilities at the interface of topological physics and laser science, such as topologically protected transport in systems with gain. We show here that the laser system based on the archetypal Haldane model exhibits topologically protected transport, with features similar to those of its passive counterpart. This behavior means that this system is likely to have topological invariants, despite the non-hermiticity. Technologically, the topological insulator laser offers an avenue to make many semiconductor lasers operate as one single-mode high-power laser. The topological insulator laser constructed from an aperiodic array of resonators was realized experimentally in an all-dielectric platform, as described in the accompanying experimental paper by Bandres *et al.* ■



Topological insulator laser based on the Haldane model and its efficiency. (**A**) Planar honeycomb lattice of coupled microring resonators pumped at the perimeter. The topological lasing mode has unidirectional flux with almost-uniform intensity, which builds up as the mode circulates and drops when passing the output coupler. (**B**) Slope efficiency (in arbitrary units) versus disorder strength for three cases differing only in the Haldane phase (of the next-to-nearest neighbor coupling): a topological laser with the maximum gap (blue; Haldane phase of $\pi/2$), one with a small topological gap (red; Haldane phase of $\pi/8$), and a topologically trivial laser with no gap (black; Haldane phase of 0).

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: msegev@tx.technion.ac.il
Cite this article as G. Harari *et al.*, *Science* **359**, eaar4003 (2018). DOI: 10.1126/science.aar4003

RESEARCH ARTICLE

TOPOLOGICAL PHOTONICS

Topological insulator laser: Theory

Gal Harari,[1]* Miguel A. Bandres,[1]* Yaakov Lumer,[2] Mikael C. Rechtsman,[3] Y. D. Chong,[4] Mercedeh Khajavikhan,[5] Demetrios N. Christodoulides,[5] Mordechai Segev[1]†

Topological insulators are phases of matter characterized by topological edge states that propagate in a unidirectional manner that is robust to imperfections and disorder. These attributes make topological insulator systems ideal candidates for enabling applications in quantum computation and spintronics. We propose a concept that exploits topological effects in a unique way: the topological insulator laser. These are lasers whose lasing mode exhibits topologically protected transport without magnetic fields. The underlying topological properties lead to a highly efficient laser, robust to defects and disorder, with single-mode lasing even at very high gain values. The topological insulator laser alters current understanding of the interplay between disorder and lasing, and at the same time opens exciting possibilities in topological physics, such as topologically protected transport in systems with gain. On the technological side, the topological insulator laser provides a route to arrays of semiconductor lasers that operate as one single-mode high-power laser coupled efficiently into an output port.

Topological insulators emerged in condensed matter physics (*1*–*3*) and constitute a new phase of matter, with insulating bulk and quantized and robust edge conductance. The prospect of observing topological effects in non-electronic systems was first proposed and demonstrated in microwaves in gyro-optic crystals with broken time-reversal symmetry (*4*–*6*). The transition to optical frequencies required another conceptual leap. Theoretical proposals (*7*–*10*) were followed by the first experiments using artificial gauge fields—one based on a honeycomb lattice of helical waveguides (*11*) and the other based on an array of aperiodic coupled silicon microring resonators (*12*). Topological protection is now known to be a ubiquitous phenomenon, occurring in physical settings as varied as electromagnetic waves (*6*, *11*–*16*), cold atoms (*17*, *18*), and acoustic, mechanical, and elastic systems (*19*–*22*). So far, however, most of these experiments have been carried out in entirely passive, linear, and conservative settings.

Nonhermiticity in topological systems

Lasers represent complex nonconservative systems capable of exhibiting very rich dynamics. They are fundamentally non-Hermitian and nonlinear entities that rely on saturable gain. Introducing nonhermiticity to topological systems raises fundamental questions concerning whether non-Hermitian topological systems can exist at all (*23*), and if they do, how to define the topological phases and their stability (*24*). Part of this controversy has recently been resolved, and it is now known that one-dimensional non-Hermitian systems can have stationary zero-dimensional topological edge states (*25*, *26*) and topological defect states (*27*–*29*), and these can even lase (*29*–*32*). However, being one-dimensional systems lasing from a localized defect, none of these systems can support transport via edge states. Very recently, lasing from a topological edge state in a photonic crystal subject to an external magnetic field—a photonic analog of the quantum Hall effect—was reported (*33*). That system used magneto-optic effects, which are very weak at optical frequencies, and accordingly produced only a narrow topological band gap (40 pm) within which it was claimed that lasing occurred (*33*). Clearly, it will be important to pursue new approaches in expanding the topological photonic band gap, and hence the degree of protection endowed to such structures. Equally important, in terms of applications, will be to follow an all-dielectric strategy that is by nature compatible with semiconductor laser technologies.

Notwithstanding recent progress in topological photonics and lasing therein, the fundamental question is still open: Can topological protection of transport be incorporated into non-Hermitian, highly nonlinear, open systems such as lasers? Addressing this issue is at the heart of the topological insulator laser, along with many new questions: Can a topological insulator include gain, which has no equivalent in condensed matter physics? What happens to the stimulated emission under topologically protected transport? Can a topological insulator laser exist without a magnetic field, and if so, might it exhibit new features and improved laser action?

Laser cavities are typically optimized to attain high Q factors, and they are always strongly affected by disorder, which can arise from manufacturing imperfections, operational failure, stresses, etc. A prominent consequence of disorder is mode localization, which has dire implications in photonics (*34*, *35*). To a laser, disorder implies degraded overlap of the lasing mode with the gain profile, lower output coupling, and multimode lasing, which together result in an overall reduced

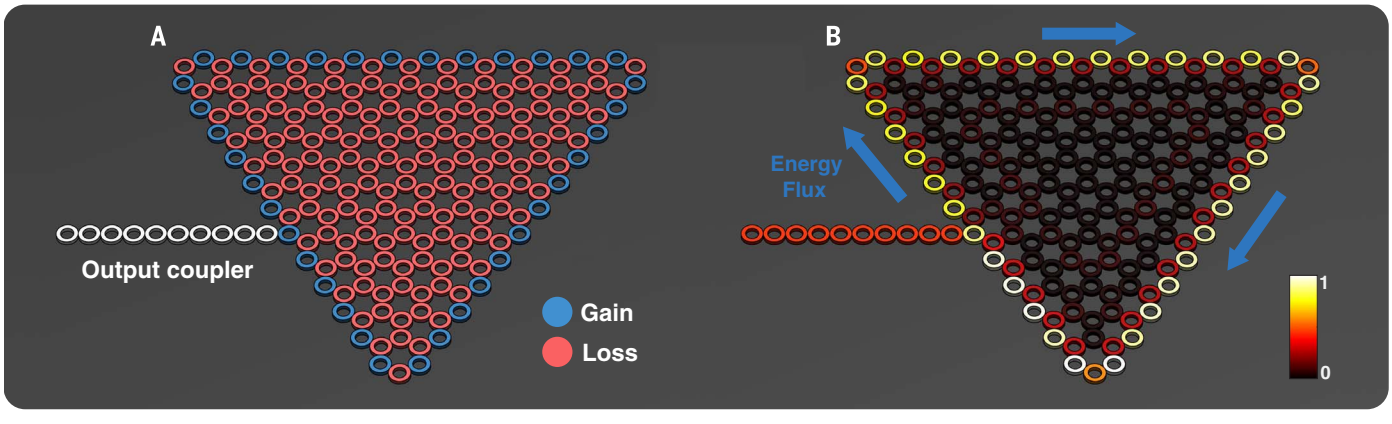


Fig. 1. Geometry and lasing mode in a topological insulator laser based on the Haldane model. (**A**) Cavity geometry (same for topological and trivial): a planar honeycomb lattice of coupled microring resonators. The cavity has unpumped (lossy) resonators (red), pumped resonators (blue), and an output channel (white). (**B**) The steady-state topological lasing mode of the topological cavity. The lasing mode is extended all around the perimeter of the cavity with almost uniform intensity, and its energy flux is unidirectional. The unidirectional energy flux can be detected by the intensity buildup as the mode circulates (clockwise) and by the sudden drop in intensity when passing (clockwise) the output coupler.

efficiency. These problems are further exacerbated in more involved arrangements such as laser arrays. Such laser array structures tend to lase with many modes simultaneously, with their modal structure (near-field and spectrum) varying with the pumping strength. Despite many methods suggested to control the emission pattern of laser diode arrays, current technology is still not able to make the lasers operate as a single high-power coherent laser source. Consequently, the most common application of laser arrays is their usage as pumps (instead of flashtubes) for solid-state lasers but not as coherent high-power laser sources.

Topological insulator lasers

Topological insulator lasers are lasers whose lasing mode exhibits topologically protected transport, such that the light propagates along the edges of the cavity in a unidirectional fashion, immune to scattering and disorder, unaffected by the shape of the edges. We show that the underlying topological properties lead to a highly efficient laser, robust to fabrication and to operational disorder and defects, that exhibits single-mode lasing even at gain values high above the laser threshold.

We consider two possible configurations involving planar arrays of coupled, active resonators. The first is based on the Haldane model (*36*), an archetypical model for time reversal–broken topological insulators. The second relies on an aperiodic array architecture that creates an artificial magnetic field, as was demonstrated experimentally in passive systems (silicon) (*8*, *12*, *13*, *37*). This is an all-dielectric system that has been realized with current semiconductor laser technology, as described in (*38*).

Design of the topological laser cavity

In the Haldane design (*36*), the resonators of the topological insulator laser are arranged in a honeycomb lattice (Fig. 1A). Each resonator is coupled to its nearest neighbors by a real hopping parameter t_1, and to its next-nearest neighbors by a complex parameter $t_2 \exp(i\phi)$, where ϕ is the Haldane flux parameter (*39*). The two sublattices of the honeycomb structure have identical on-site potentials. The passive Haldane model (no gain or loss, but $t_2 \neq 0$) exhibits two phases: the trivial phase when ϕ is equal to 0 or π, and the topological phase when $\phi \neq 0, \pi$. In the topological phase, edge states emerge with energies extending across the topological gap that is proportional to $t_2 \sin\phi$, reaching a maximum at $\phi = \pi/2$. To promote the lasing of the topological edge mode, we specifically design the honeycomb array to have zig-zag edges that have small penetration depth into the bulk. The evolution of the field of this laser system is governed by

$$i\frac{\partial \boldsymbol{\psi}}{\partial t} = H_{\mathrm{Haldane}}\boldsymbol{\psi} - i\gamma\boldsymbol{\psi} + \frac{ig\mathbb{P}}{1+|\boldsymbol{\psi}|^2/I_{\mathrm{sat}}}\boldsymbol{\psi} + H_{\mathrm{output}}\boldsymbol{\psi} \quad (1)$$

where $\boldsymbol{\psi}$ is a column vector encompassing the modal amplitudes of the array elements, and H_{Haldane} is the standard Haldane Hamiltonian (*39*), which depends on the resonance frequency of a single resonator ω_0, the hopping constants $t_{1,2}$, and the Haldane flux ϕ. Here, γ represents the loss in each resonator; this is assumed to be linear loss [as in all constant-wave lasers], although we simulated saturable loss as well, and our results were effectively the same. The third term in Eq. 1 represents optical gain via stimulated emission that is inherently saturable (I_{sat}). Here, $\mathbb{P}$ stands for the spatial profile of the pump, and H_{output} describes the output coupler

[1]Physics Department and Solid State Institute, Technion–Israel Institute of Technology, Haifa 32000, Israel. [2]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]Department of Physics, Pennsylvania State University, University Park, PA 16802, USA. [4]Division of Physics and Applied Physics, Nanyang Technological University, 637371 Singapore. [5]CREOL, College of Optics and Photonics, University of Central Florida, Orlando, FL 32816, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: msegev@tx.technion.ac.il

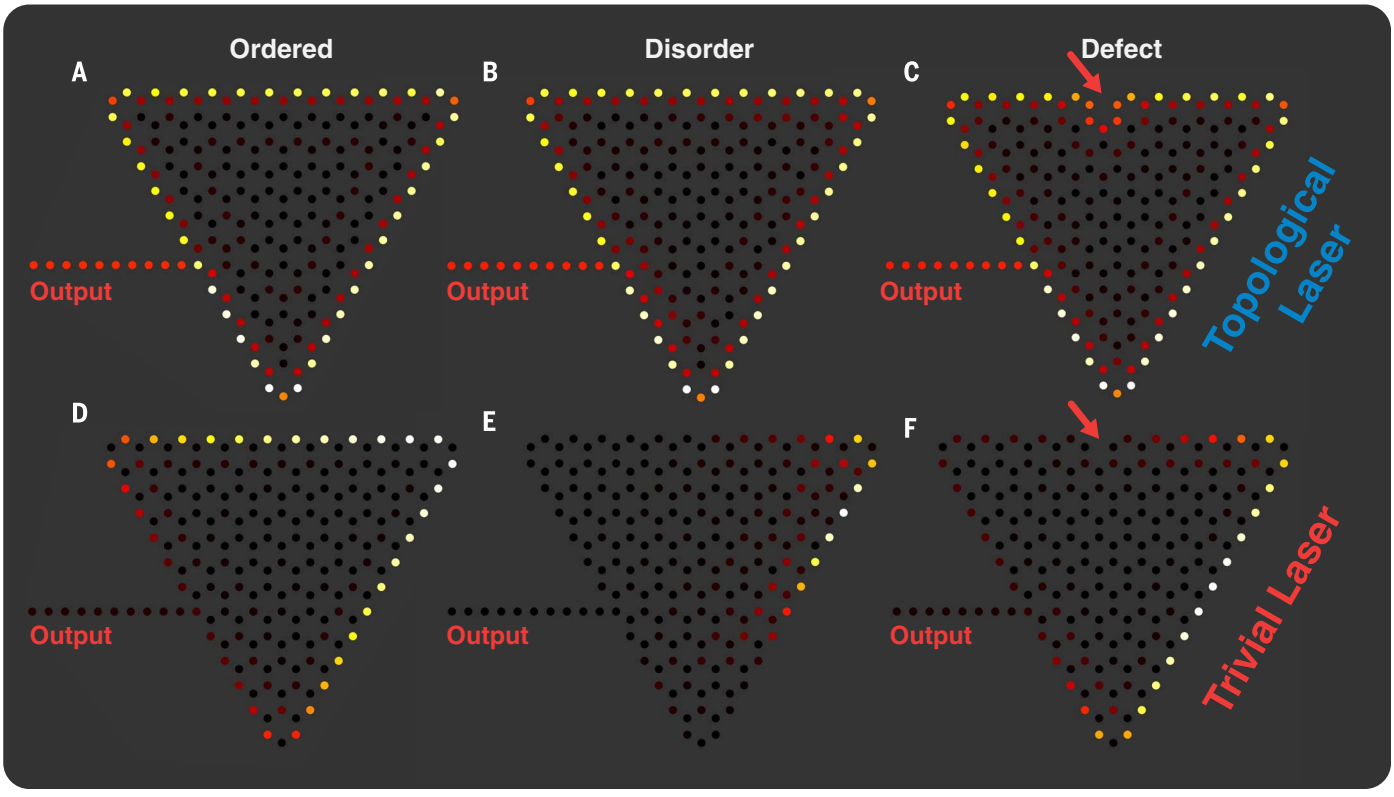


Fig. 2. Lasing modes in the topological and trivial cavities (Haldane model). (**A** to **F**) Steady-state lasing modes (colors indicate amplitude) for the topological and trivial cavities with no disorder [(A) and (D)], disorder [(B) and (E)], and a defect (a missing gain resonator at the perimeter) [(C) and (F)]. Being unidirectional and extended, with almost uniform intensity, the lasing mode in the topological cavity [(A) to (C)] exhausts all the pumped sites and couples strongly to the output coupler in all cases. The trivial cavity, lacking unidirectional extended lasing modes, couples only weakly to the output coupler even without disorder (D). The presence of disorder (E) or defects (F) causes the lasing mode of the trivial cavity to be highly localized, further reducing the efficiency and the coupling to the output port.

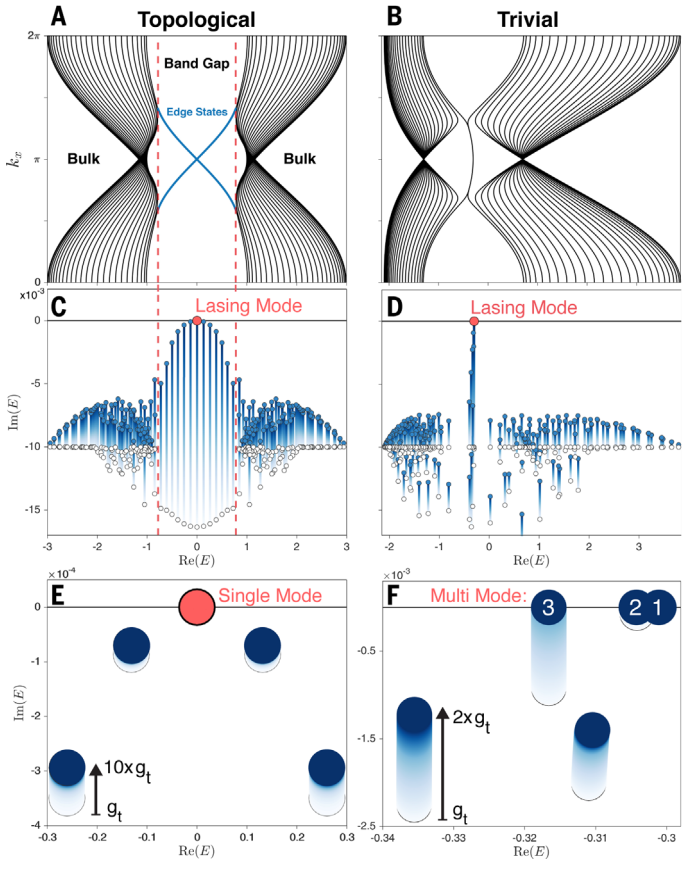


Fig. 3. The lasing process in the topological and trivial laser cavities. (**A** and **B**) Band diagram of the passive (no gain or loss) topological (A) and trivial (B) strip of resonators. The topological cavity has a band gap with topologically protected unidirectional edge states crossing the gap (blue lines). By contrast, the trivial cavity has no protected transport, as there is no gap and the edge states are degenerate. (**C** and **D**) The evolution of the real and imaginary parts of the spectrum of the topological (C) and trivial (D) cavities, from zero gain (white dots) up to the threshold gain (blue dots), which is where the first lasing mode lases. (**E** and **F**) Evolution of the spectrum of the laser cavity modes as the gain is increased further above the threshold gain (g_t). The topological laser (E) maintains single-mode lasing, at least up to gain values of 10 times the threshold gain. By contrast, the trivial laser (F) becomes multimode, with three lasing modes, at a gain level just twice the threshold gain.

(a semi-infinite chain of resonators, which is what makes this an open system) and its coupling to the cavity. To promote lasing of the edge modes, the gain (g) is provided only to the resonators on the perimeter (Fig. 1A). In general, g is a function of the frequency; however, in semiconductors the gain is broadband (relative to the inter-ring coupling constant), and hence it is assumed to be frequency-independent.

Topological insulator laser based on the Haldane model

We study the topological insulator laser by simulating the full dynamics and by directly solving for the nonlinear lasing modes of Eq. 1 (*39*). We consider two cases: (i) a trivial arrangement where $\phi = 0$, and (ii) a topological array when the Haldane flux is $\phi = \pi/2$. In the trivial case ($\phi = 0$), even without disorder, the first lasing mode is localized away from the output coupler (Fig. 2D) so as to minimize the power loss through the coupler. This strongly affects the lasing efficiency because the localized lasing mode does not use all the gain available around the perimeter and it is coupled only very weakly to the output. This adversarial effect is further enhanced in the presence of disorder, given that it tends to highly localize modes. For example (Fig. 2E), in the presence of disorder, the lasing mode of the trivial cavity becomes further confined within just a few resonators. Finally, because the trivial lasing mode does not spatially deplete the gain, if pumped harder, multiple modes can reach threshold, giving rise to multimode operation (Fig. 3F).

In the topological case ($\phi = \pi/2$), the lasing mode possesses all the distinct characteristics of topological chiral edge states: It is extended all around the perimeter of the cavity with almost-uniform intensity, and its energy flux is unidirectional. The unidirectionality of the energy flux can be observed by noticing how the energy starts accumulating while moving clockwise around the perimeter of the array, and then abruptly drops after the output coupler (Fig. 1B). Even though the notion of topological invariants in non-Hermitian, nonlinear open systems (such as lasers) is still largely unexplored, the topological nature of this array can be recognized by observing its dynamics. In this regard, it is instructive to monitor the evolution in the system, from below transparency to lasing. Figure 3, C and D, shows the behavior of the complex frequency of the cavity modes as we increase the gain from zero to the threshold level. During this process, the frequency (real part) of each mode is virtually unchanged, equal to the eigenfrequencies of the closed Hermitian system (the topological insulator lattice), and the band gap of the topological system (blue region in Fig. 3A) remains well defined and constant. Clearly, the topological band gap does not close in the presence of gain (even high above the threshold level), and no topological phase transition occurs. This is a clear indication that the system retains its topological features even when it lases. More important, as shown in Fig. 3C, one of the topological edge modes is the first in line to lase. Moreover, even when we increase the gain (pumping) high above the threshold level, the lasing remains in a single mode—the topological edge mode. This is because the topological lasing mode has almost uniform intensity all around the perimeter of the array, saturating the whole gain medium, hence suppressing all the other cavity modes and preventing them from lasing (Fig. 3E).

To further examine the topological properties of the lasing mode, we study the system in the presence of disorder and defects. We find that even in the presence of disorder, the topological insulator laser still operates in a topological edge mode (Fig. 2B). Even when a defect is introduced—for example, a malfunctioning microring at the perimeter—the topological edge mode is able to bypass the defect with minimal penetration into the bulk (Fig. 2C).

It is instructive to highlight the role of topology in this topological insulator laser by comparing its efficiency to its trivial counterpart. The efficiency of lasers is universally defined by the so-called "slope efficiency": the slope of the function describing the output power as a function of gain above the threshold level. The calculated slope efficiencies (in steady state) as a function of the strength of the disorder are shown in Fig. 4A. When the cavity is in the trivial phase, $\phi = 0$, the slope efficiency is poor and further decreases in the presence of disorder. In sharp contrast, when the laser is topological, $\phi = \pi/2$, the lasing is very efficient and robust; it has a high slope efficiency (with a small variance) that remains high even at strong disorder levels. The topological protection ceases only when the strength of disorder is comparable to the size of the topological band gap. To reaffirm that the high efficiency of this laser is due to its topological protection, we evaluate the slope efficiency for structures with a smaller band gap (e.g., for $\phi = \pi/8$). As seen in Fig. 4A, the slope efficiency of this "small topological gap implementation" is initially high, but when the level of disorder is increased, it deteriorates much faster than the $\phi = \pi/2$ case, because the disorder closes the smaller topological gap more easily. To further highlight the robustness of our topological insulator laser, we calculate the mean value of the slope efficiency as a function of disorder strength and the two main parameters that control the size of the topological band gap: the Haldane flux (Fig. 4B) and the next-nearest neighbor coupling t_2 (Fig. 4C). When the strength of randomness is less than the size of the Haldane topological band gap ($6\sqrt{3}t_2 \sin\phi$), the slope efficiency stays high, and drops only when the disorder level exceeds the size of the topological band gap (Fig. 4, B and C). This is a clear indication that the efficiency and robustness of the

topological insulator laser arise from its topological properties (*40*).

Topological insulator laser based on an aperiodic array of resonators

Next, we study lasing in an aperiodic topological array of microring resonators, which was one of the two platforms first explored for realizing photonic topological insulators in optics (*8*, *12*). This arrangement involves a lattice of coupled resonators with aperiodic couplers (*39*)—an architecture that can be implemented using standard semiconductor technologies (*41*, *42*). Note that this system does not use magnetic fields, nor does it use any exotic materials (such as yttrium iron garnet). It relies only on present semiconductor laser technology. The aperiodic couplers establish an artificial gauge field, thus leading to behavior analogous to the quantum Hall effect (*12*). Because in the linear regime the system is reciprocal, both the clockwise (CW) and counterclockwise (CCW) modes in each microring resonator experience gauge fields with opposite signs. This, in turn, makes the overall cavity degenerate; that is, for any frequency supported by the CW modes, there is a corresponding CCW mode. In passive settings (no gain or loss), for reasonable experimental parameters, one can consider the CW and CCW as decoupled (*8*, *12*). For a topological insulator laser based on this system, one must take into account that the CW and CCW modes inevitably interact with one another through the nonlinear effect of gain saturation and backscattering, both of which naturally occur in active media (*43*).

This topological aperiodic laser array is simulated using an extended version of Eq. 1 (*39*). To account for a realistic setup in current semiconductor laser technology, we implement saturable gain and explicit coupling between the CW and CCW modes through the saturable medium and backscattering. We use the parameters of recent experiments (*42*) and of the accompanying experimental paper (*38*), which yield a topological band gap of ~1 nm, in an aperiodic ring array made of all-dielectric semiconductor laser materials. A topological gap of this size ensures very high topological protection of transport in the laser cavity. We essentially observe the same features as in the Haldane model: The topological lasing mode is extended, uniform, and couples strongly to the output coupler, even in the presence of disorder. In contrast, the topologically trivial aperiodic array, having no edge modes, suffers from localization of its lasing modes and displays strong multimode lasing and low output coupling (Fig. 5). It is important to stress that despite the reciprocity of the system and the inherent degeneracy of every ring and the presence of scattering and saturable gain (which cause lasing in both the CW and CCW directions), the topologically protected features of the lasing modes prevail, and the efficiency of the topological array is much higher than that of the trivial case. The underlying mechanism can be clearly seen from the mode shapes in Fig. 5, B to E, where the mutual contribution of the two counterpropagating (CW and CCW) edge modes with the same temporal frequency depletes the gain and enforces efficient single-frequency lasing. In a similar vein, the topological lasing modes exhibit high robustness against defects and disorder, despite the backscattering and the nonlinear coupling between the CW and CCW modes.

Concluding remarks

As a result of its topological properties, the topological insulator laser exhibits high efficiency, extreme robustness to defects and disorder, and high-efficiency single-mode lasing even at gain levels high above the threshold level. The interplay between topology and nonhermiticity, especially in nonlinear open systems such as lasers, raises many fundamental questions. We have shown that the laser system based on the archetypal Haldane model exhibits topologically protected transport, with features similar to its passive counterpart. This behavior means that there must be associated topological invariants, even though this system is non-Hermitian. Topologically protected transport in lasing systems can change existing paradigms by harnessing the topological features of the lasing mode to yield high-efficiency single-mode lasing even high above threshold.

The systems proposed here are a proof of concept, with no attempt to optimize the integration of topological properties into a laser. The topological design of the laser cavity can take on many different concrete designs, which may lead to new ideas and innovative applications. For example, the topological system can be based on topological network models of strongly coupled resonators (*44*), which can be realized with very small units. More generally, new geometries, wherein a medium with narrow-linewidth gain is matched to the topological gap, can be considered. In a similar vein, topological insulator lasers could be designed from flexible gain media (organics), and operate under distortions, manipulations, and perhaps also extreme conditions. It may even be possible to integrate topological insulator lasers with sensors, antennas, and other photonic devices. The accompanying paper (*38*) describes the realization of a topological insulator laser without magnetic field. It is an all-dielectric system, based on the aperiodic array of ring resonators described here, fabricated with

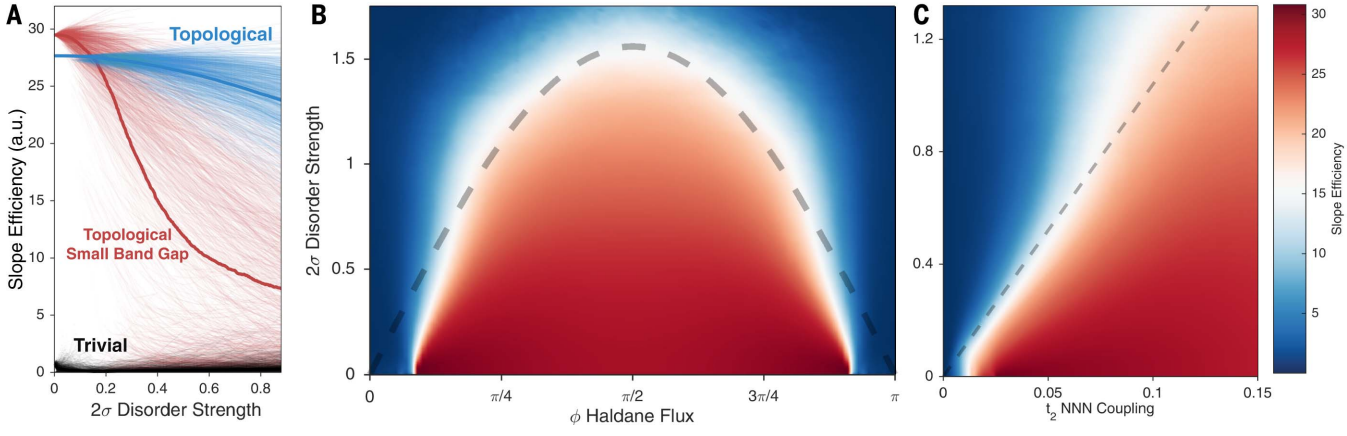


Fig. 4. Slope efficiency versus disorder in the topological and trivial laser systems (Haldane model). (**A**) Slope efficiency (in arbitrary units) versus disorder strength (measured in terms of its standard deviation) for the topological laser with the maximum gap (blue; $\phi = \pi/2$) and with a small topological gap (red; $\phi = \pi/8$), and for the trivial laser with no gap (black; $\phi = 0$). Every point corresponds to the slope efficiency of a different realization of the disorder (1000 in total). Solid lines mark the mean values for each case. The topological cavities exhibit higher slope efficiency (blue line) than the trivial cavity, even under high levels of disorder. For the topological insulator laser with a small band gap, when the disorder level is increased to above the band gap size, the topological protection of this topological laser starts to break and its efficiency is markedly deteriorated (red line). (**B** and **C**) Mean value of slope efficiency (average over 1000 realizations) as a function of disorder and Haldane flux ϕ (B) and as a function of disorder and next-nearest neighbor coupling t_2 (C). The dashed black lines depict the size of the band gap of the Haldane model at zero disorder, given by $6\sqrt{3}t_2 \sin\phi$. Clearly, the slope efficiency of the topological cavity stays high as long as the disorder strength is lower than the size of the topological gap.

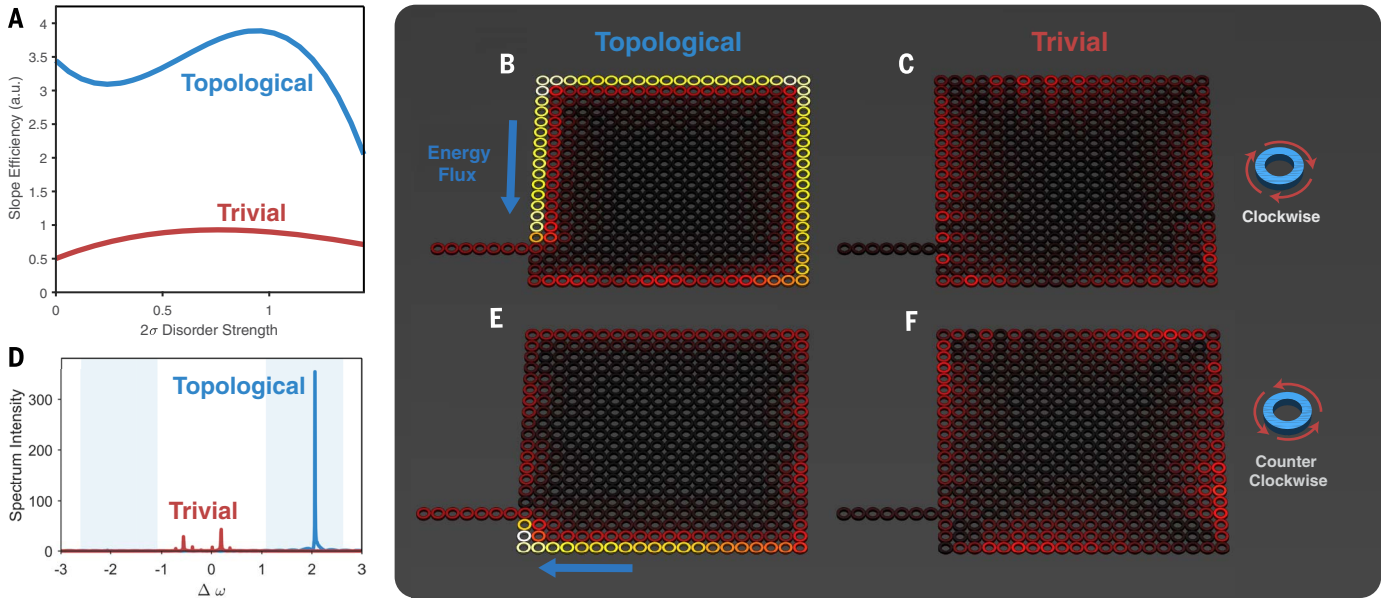


Fig. 5. Lasing modes in the topological and trivial cavities (coupled-resonators model). (**A**) Mean slope efficiency as a function of disorder strength (measured in terms of the standard deviation of the on-site energy) for the topological laser (blue) and the trivial laser (red). The simulation includes 5% backscattering on each site. The slope efficiency of the topological laser is higher and more robust up to disorder levels on the order of the topological band gap. (**B**, **C**, **E**, and **F**) Typical lasing modes (colors indicate amplitude as indicated in Fig. 1) for the topological [(B) and (E)] and the trivial [(C) and (F)] cavities. The first and second rows show the CW and CCW components, respectively. In the topological array, the CW (B) and CCW (E) components travel in opposite directions around the lattice but they both lase at the same frequency, synergistically transporting energy from the pumped sites to the output coupler. By contrast, in the trivial lattice [(C) and (F)], the modes become localized in the presence of disorder, penetrate into the lossy bulk, and couple only weakly to the output coupler. (**D**) Spectra of the topological (blue) and the trivial (red) lasers for the same level of pumping. Even in the presence of disorder and backscattering, the topological array lases in a single mode inside one of the topological band gaps (shaded blue regions), whereas the trivial array, for the same gain and disorder levels, lases weakly and at multiple frequencies.

ordinary technology for making semiconductor lasers.

REFERENCES AND NOTES

1. C. L. Kane, E. J. Mele, Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95**, 226801 (2005). pmid: 16384250
2. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006). doi: 10.1126/science.1133734; pmid: 17170299
3. M. König *et al*., Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007). doi: 10.1126/science.1148047; pmid: 17885096
4. S. Raghu, F. Haldane, Analogs of quantum-Hall-effect edge states in photonic crystals. *Phys. Rev. A* **78**, 033834 (2008). doi: 10.1103/PhysRevA.78.033834
5. Z. Wang, Y. D. Chong, J. D. Joannopoulos, M. Soljačić, Reflection-free one-way edge modes in a gyromagnetic photonic crystal. *Phys. Rev. Lett.* **100**, 013905 (2008). doi: 10.1103/PhysRevLett.100.013905; pmid: 18232767
6. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature* **461**, 772–775 (2009). doi: 10.1038/nature08293; pmid: 19812669
7. R. O. Umucalılar, I. Carusotto, Artificial gauge field for photons in coupled cavity arrays. *Phys. Rev. A* **84**, 043804 (2011). doi: 10.1103/PhysRevA.84.043804
8. M. Hafezi, E. Demler, M. Lukin, J. Taylor, Robust optical delay lines via topological protection. *Nat. Phys.* **7**, 907–912 (2011). doi: 10.1038/nphys2063
9. K. Fang, Z. Yu, S. Fan, Realizing effective magnetic field for photons by controlling the phase of dynamic modulation. *Nat. Photonics* **6**, 782–787 (2012). doi: 10.1038/nphoton.2012.236
10. A. B. Khanikaev *et al*., Photonic topological insulators. *Nat. Mater.* **12**, 233–239 (2013). pmid: 23241532
11. M. C. Rechtsman *et al*., Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013). doi: 10.1038/nature12066; pmid: 23579677
12. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, Imaging topological edge states in silicon photonics. *Nat. Photonics* **7**, 1001–1005 (2013). doi: 10.1038/nphoton.2013.274
13. S. Mittal *et al*., Topologically robust transport of photons in a synthetic gauge field. *Phys. Rev. Lett.* **113**, 087403 (2014). doi: 10.1103/PhysRevLett.113.087403; pmid: 25192126
14. X. Cheng *et al*., Robust reconfigurable electromagnetic pathways within a photonic topological insulator. *Nat. Mater.* **15**, 542–548 (2016). doi: 10.1038/nmat4573; pmid: 26901513
15. V. Peano, M. Houde, F. Marquardt, A. A. Clerk, Topological quantum fluctuations and traveling wave amplifiers. *Phys. Rev. X* **6**, 041026 (2016). doi: 10.1103/PhysRevX.6.041026
16. A. P. Slobozhanyuk *et al*., Experimental demonstration of topological effects in bianisotropic metamaterials. *Sci. Rep.* **6**, 22270 (2016). doi: 10.1038/srep22270; pmid: 26936219
17. M. Aidelsburger *et al*., Measuring the Chern number of Hofstadter bands with ultracold bosonic atoms. *Nat. Phys.* **11**, 162–166 (2015). doi: 10.1038/nphys3171
18. G. Jotzu *et al*., Experimental realization of the topological Haldane model with ultracold fermions. *Nature* **515**, 237–240 (2014). doi: 10.1038/nature13915; pmid: 25391960
19. L. M. Nash *et al*., Topological mechanics of gyroscopic metamaterials. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14495–14500 (2015). doi: 10.1073/pnas.1507413112; pmid: 26561580
20. Z. Yang *et al*., Topological acoustics. *Phys. Rev. Lett.* **114**, 114301 (2015). doi: 10.1103/PhysRevLett.114.114301; pmid: 25839273
21. P. Wang, L. Lu, K. Bertoldi, Topological phononic crystals with one-way elastic edge waves. *Phys. Rev. Lett.* **115**, 104302 (2015). doi: 10.1103/PhysRevLett.115.104302; pmid: 26382680
22. C. L. Kane, T. C. Lubensky, Topological boundary modes in isostatic lattices. *Nat. Phys.* **10**, 39–45 (2013). doi: 10.1038/nphys2835
23. Y. C. Hu, T. L. Hughes, Absence of topological insulator phases in non-Hermitian PT-symmetric Hamiltonians. *Phys. Rev. B* **84**, 153101 (2011). doi: 10.1103/PhysRevB.84.153101
24. K. Esaki, M. Sato, K. Hasebe, M. Kohmoto, Edge states and topological phases in non-Hermitian systems. *Phys. Rev. B* **84**, 205128 (2011). doi: 10.1103/PhysRevB.84.205128
25. M. S. Rudner, L. S. Levitov, Topological transition in a non-Hermitian quantum walk. *Phys. Rev. Lett.* **102**, 065703 (2009). doi: 10.1103/PhysRevLett.102.065703; pmid: 19257606
26. J. M. Zeuner *et al*., Observation of a topological transition in the bulk of a non-Hermitian system. *Phys. Rev. Lett.* **115**, 040402 (2015). doi: 10.1103/PhysRevLett.115.040402; pmid: 26252670
27. S. D. Liang, G. Y. Huang, Topological invariance and global Berry phase in non-Hermitian systems. *Phys. Rev. A* **87**, 012118 (2013). doi: 10.1103/PhysRevA.87.012118
28. S. Weimann *et al*., Topologically protected bound states in photonic parity-time-symmetric crystals. *Nat. Mater.* **16**, 433–438 (2017). pmid: 27918567
29. H. Zhao, P. Miao, M. H. Teimourpour, S. Malzard, R. El-Ganainy, H. Schomerus, L. Feng, Topological hybrid silicon microlasers. arXiv:1709.02747 (2017).
30. L. Pilozzi, C. Conti, Topological lasing in resonant photonic structures. *Phys. Rev. B* **93**, 195317 (2016). doi: 10.1103/PhysRevB.93.195317
31. P. St-Jean, V. Goblot, E. Galopin, A. Lemaître, T. Ozawa, L. Le Gratiet, I. Sagnes, J. Bloch, A. Amo, Lasing in topological edge states of a 1D lattice. arXiv:1704.07310 (2017).
32. M. Parto, S. Wittek, H. Hodaei, G. Harari, M. A. Bandres, J. Ren, M. C. Rechtsman, M. Segev, D. N. Christodoulides, M. Khajavikhan, Complex edge-state phase transitions in 1D topological laser arrays. arXiv:1709.00523 (2017).
33. B. Bahari *et al*., Nonreciprocal lasing in topological cavities of arbitrary geometries. *Science* **358**, 636–640 (2017). pmid: 29025992
34. M. Segev, Y. Silberberg, D. Christodoulides, Anderson localization of light. *Nat. Photonics* **7**, 197–204 (2013). doi: 10.1038/nphoton.2013.30
35. J. Liu *et al*., Random nanolasing in the Anderson localized regime. *Nat. Nanotechnol.* **9**, 285–289 (2014). doi: 10.1038/nnano.2014.34; pmid: 24658170
36. F. D. M. Haldane, Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the "parity anomaly". *Phys. Rev. Lett.* **61**, 2015–2018 (1988). doi: 10.1103/PhysRevLett.61.2015; pmid: 10038961

37. S. Mittal, S. Ganeshan, J. Fan, A. Vaezi, M. Hafezi, Measurement of topological invariants in a 2D photonic system. *Nat. Photonics* **10**, 180–183 (2016). doi: 10.1038/nphoton.2016.10
38. M. A. Bandres *et al*., Topological insulator laser: Experiment. *Science* **359**, eaar4005 (2018). doi: 10.1038/nphoton.2016.10
39. See supplementary materials.
40. For completeness, we also studied the operation of the laser when the same system has a trivial gap, with no edge states crossing the gap. To open a trivial gap, we detuned the resonance frequencies of the resonators associated with the two sublattices of the Haldane model. When the detuning was large enough, it opened a topologically trivial band gap even with $\phi = \pi/2$, but this trivial gap does not yield topological protection nor high-efficiency lasing. We found that the efficiency of the laser is high and robust to disorder only when the gap is topological.
41. L. Feng, Z. J. Wong, R. M. Ma, Y. Wang, X. Zhang, Single-mode laser by parity-time symmetry breaking. *Science* **346**, 972–975 (2014). doi: 10.1126/science.1258479; pmid: 25414307
42. H. Hodaei, M. A. Miri, M. Heinrich, D. N. Christodoulides, M. Khajavikhan, Parity-time-symmetric microring lasers. *Science* **346**, 975–978 (2014). doi: 10.1126/science.1258480; pmid: 25414308
43. D. Liu *et al*., Symmetry, stability, and computation of degenerate lasing modes. *Phys. Rev. A* **95**, 023835 (2016). doi: 10.1103/PhysRevA.95.023835
44. G. Q. Liang, Y. D. Chong, Optical resonator analog of a two-dimensional topological insulator. *Phys. Rev. Lett.* **110**, 203904 (2013). doi: 10.1103/PhysRevLett.110.203904; pmid: 25167412

ACKNOWLEDGMENTS

Funding: Supported by Singapore MOE Academic Research Fund Tier 2 Grant MOE2015T2-2-008 and Tier 3 Grant MOE2016-T3-1006 (C.Y.D.); Penn State NSF MRSEC Center for Nanoscale Science (under award NSF DMR-1420620), NSF grant DMS-1620422, and the Packard, Sloan, and Kaufman foundations (M.C.R.); the Israel Science Foundation; Office of Naval Research grant N0001416-1-2640; NSF grants ECCS 1454531, DMR-1420620, and ECCS 1757025; U.S. Air Force Office of Scientific Research grant FA9550-14-1-0037; U.S.-Israel Binational Science Foundation grant 2016381; the German-Israeli Deutsch-Israelische Projektkooperation program; Army Research Office grants W911NF-16-1-0013 and W911NF-17-1-0481; and ERC grant NHQWAVE (MSCA-RISE 691209). M.S. thanks M. Karpovsky and B. Shillman for their support that came at a critical time. **Author contributions:** All authors contributed to all aspects of this work. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/eaar4003/suppl/DC1
Materials and Methods
Figs. S1 and S2
References (*45*, *46*)

3 November 2017; accepted 17 January 2018
Published online 1 February 2018
10.1126/science.aar4003

RESEARCH ARTICLE SUMMARY

TOPOLOGICAL PHOTONICS

Topological insulator laser: Experiments

Miguel A. Bandres,* Steffen Wittek,* Gal Harari,* Midya Parto, Jinhan Ren, Mordechai Segev,† Demetrios N. Christodoulides,† Mercedeh Khajavikhan†

INTRODUCTION: Physical systems that exhibit topological invariants are naturally endowed with robustness against perturbations, as was recently demonstrated in many settings in condensed matter, photonics, cold atoms, acoustics, and more. The most prominent manifestations of topological systems are topological insulators, which exhibit scatter-free edge-state transport, immune to perturbations and disorder. Recent years have witnessed intense efforts toward exploiting these physical phenomena in the optical domain, with new ideas ranging from topology-driven unidirectional devices to topological protection of path entanglement. But perhaps more technologically relevant than all topological photonic settings studied thus far is, as proposed by the accompanying theoretical paper by Harari *et al.*, an all-dielectric magnet-free topological insulator laser, with desirable properties stemming from the topological transport of light in the laser cavity.

RATIONALE: We demonstrate nonmagnetic topological insulator lasers. The topological properties of the laser system give rise to single-mode lasing, robustness against fabrication defects, and notably higher slope efficiencies compared to those of the topologically trivial counterparts. We further exploit the properties of the active topological platform by assembling topological insulator lasers from *S*-chiral microresonators that enforce predetermined unidirectional lasing even in the absence of magnetic fields.

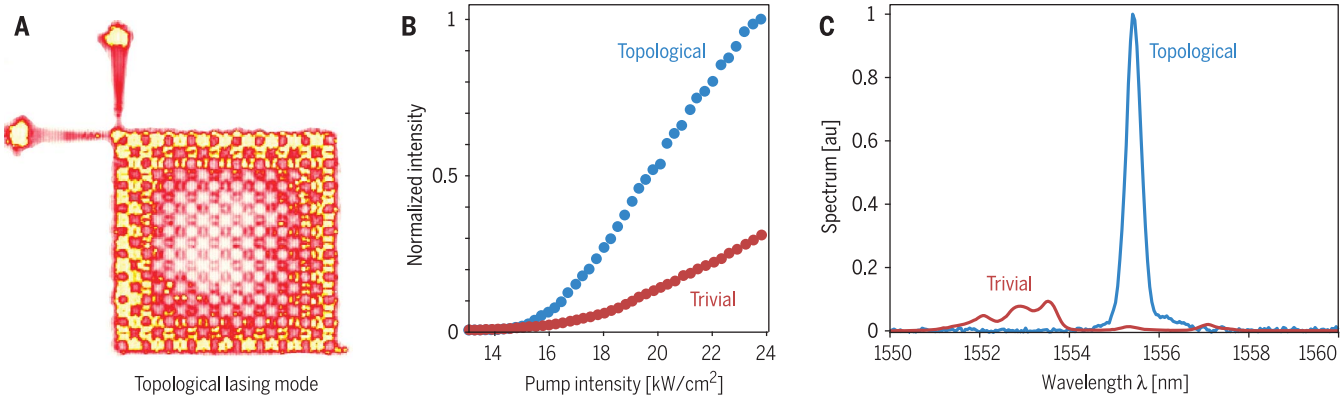
RESULTS: Our topological insulator laser system is an aperiodic array of 10 unit cell–by–10 unit cell coupled ring resonators on an InGaAsP quantum wells platform. The active lattice uses the topological architecture suggested in the accompanying theoretical paper. This two-dimensional setting is composed of a square lattice of ring resonators coupled to each other by means of link rings. The intermediary links are judiciously spatially shifted to introduce a set of hopping phases, establishing a synthetic magnetic field and two topological band gaps. The gain in this laser system is provided by optical pumping. To promote lasing of the topologically protected edge modes, we pump the outer perimeter of the array while leaving the interior lossy. We find that this topological insulator laser operates in single mode even considerably above threshold, whereas the corresponding topologically trivial realizations lase in multiple modes. Moreover, the topological laser displays a slope efficiency that is considerably higher than that in the corresponding trivial realizations. We further demonstrate the topological features of this laser by observing that in the topological array, all sites emit coherently at the same wavelength, whereas in the trivial array, lasing occurs in localized regions, each at a different frequency. Also, by pumping only part of the topological array, we demonstrate that the topological edge mode always travels along the perimeter and emits light through the output coupler. By contrast, when we pump the trivial array far from the output coupler, no light is extracted from the coupler because the lasing occurs at stationary modes. We also observe that, even in the presence of defects, the topological protection always leads to more efficient lasing compared to that of the trivial counterpart. Finally, to show the potential of this active system, we assemble a topological system based on *S*-chiral resonators, which can provide new avenues to control the topological features.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aar4005

CONCLUSION: We have experimentally demonstrated an all-dielectric topological insulator laser and found that the topological features enhance the lasing performance of a two-dimensional array of microresonators, making them lase in unison in an extended topologically protected scatter-free edge mode. The observed single longitudinal-mode operation leads to a considerably higher slope efficiency as compared to that of a corresponding topologically trivial system. Our results pave the way toward a new class of active topological photonic devices, such as laser arrays, that can operate in a coherent fashion with high efficiencies. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: msegev@techunix.technion.ac.il (M.S.); demetri@creol.ucf.edu (D.N.C.); mercedeh@creol.ucf.edu (M.K.)
Cite this article as M. A. Bandres *et al.*, *Science* **359**, eaar4005 (2018). DOI: 10.1126/science.aar4005



Topological insulator laser. (**A**) Top-view image of the lasing pattern (topological edge mode) in a 10 unit cell–by–10 unit cell array of topologically coupled resonators and the output ports. (**B**) Output intensity versus pump intensity for a topological insulator laser and its trivial counterpart. The enhancement of the slope efficiency is about threefold. Comparing the power emitted in the single mode of the topological array to that of the highest power mode in the trivial array, the topological outperforms the trivial by more than a factor of 10. (**C**) Emission spectra from a topological insulator laser and its topologically trivial counterpart. au, arbitrary units.

RESEARCH ARTICLE

TOPOLOGICAL PHOTONICS

Topological insulator laser: Experiments

Miguel A. Bandres,[1]* Steffen Wittek,[2]* Gal Harari,[1]* Midya Parto,[2] Jinhan Ren,[2] Mordechai Segev,[1]† Demetrios N. Christodoulides,[2]† Mercedeh Khajavikhan[2]†

Physical systems exhibiting topological invariants are naturally endowed with robustness against perturbations, as manifested in topological insulators—materials exhibiting robust electron transport, immune from scattering by defects and disorder. Recent years have witnessed intense efforts toward exploiting these phenomena in photonics. Here we demonstrate a nonmagnetic topological insulator laser system exhibiting topologically protected transport in the cavity. Its topological properties give rise to single-mode lasing, robustness against defects, and considerably higher slope efficiencies compared to the topologically trivial counterparts. We further exploit the properties of active topological platforms by assembling the system from *S*-chiral microresonators, enforcing predetermined unidirectional lasing without magnetic fields. This work paves the way toward active topological devices with exciting properties and functionalities.

Topological insulators are a phase of matter that feature an insulating bulk while supporting conducting edge states (*1*–*3*). Notably, the transport of edge states in topological insulators is granted topological protection, a property stemming from the underlying topological invariants (*2*). For example, in two-dimensional (2D) systems, the ensued one-way conduction along the edge of a topological insulator is, by nature, scatter free—a direct outcome of the nontrivial topology of the bulk electronic wave functions (*3*). Although topological protection was initially encountered in the integer quantum Hall effect (*4*), the field of topological physics developed rapidly after it was recognized that topologically protected transport can also be observed even in the absence of a magnetic field (*5*, *6*). This, in turn, spurred a flurry of experimental activities in a number of electronic material systems (*7*). The promise of robust transport inspired studies in many and diverse fields beyond solid-state physics, such as optics, ultracold atomic gases, mechanics, and acoustics (*8*–*25*). Along these lines, unidirectional topological states were observed in microwave settings (*9*) in the presence of a magnetic field (the electromagnetic analog of the quantum Hall effect), and, more recently, topologically protected transport phenomena have been successfully demonstrated in optical passive all-dielectric environments by introducing artificial gauge fields (*14*, *15*).

In photonics, topological concepts could lead to new families of optical structures and devices by exploiting robust, scatter-free light propagation. Lasers, in particular, could directly benefit from such attributes [see the accompanying theoretical paper (*26*)]. In general, laser cavities are prone to disorder, which inevitably arises from fabrication imperfections, operational degradation, and malfunction. Specifically, the presence of disorder in a laser gives rise to spatial light localization within the cavity, ultimately resulting in a degraded overlap of the lasing mode with the gain profile. This implies lower output coupling, multimode lasing, and reduced slope efficiency. These issues become acute in arrays of coupled laser resonators (used to yield higher power), in which a large number of elements is involved. Naturally, it would be of interest to exploit topological features in designing laser systems that are immune to disorder. In this spirit, several groups have recently studied edge-mode lasing in topological 1D Su-Schrieffer-Heeger resonator arrays (*27*–*29*). However, being one-dimensional, they lase in a zero-dimensional defect state, which inherently cannot provide protected transport. Conversely, 2D laser systems can directly benefit from topological protection. Indeed, it was shown theoretically that it is possible to harness the underlying features of topological insulators in 2D laser arrays, when lasing in an extended topological state (*26*, *30*, *31*). As indicated there, such systems can operate in a single-mode fashion with high slope efficiencies, in spite of appreciable disorder. In a following development, unidirectional edge-mode lasing was demonstrated in a topological photonic crystal configuration involving a yttrium iron garnet (YIG) substrate under the action of a magnetic field (*32*). In that system, lasing occurred within a narrow spectral band gap induced by magneto-optic effects. Clearly, it would be of interest to pursue magnet-free approaches that are, by nature, more compatible with fabrication procedures and photonic integration involving low-loss components. In addition, such all-dielectric systems can prove advantageous in substantially expanding the topological band gap and, in doing that, bring the topological protection of photon transport to the level at which lasing is immune to defects and disorder.

Here we report the first observation of topologically protected edge-mode lasing in nonmagnetic,

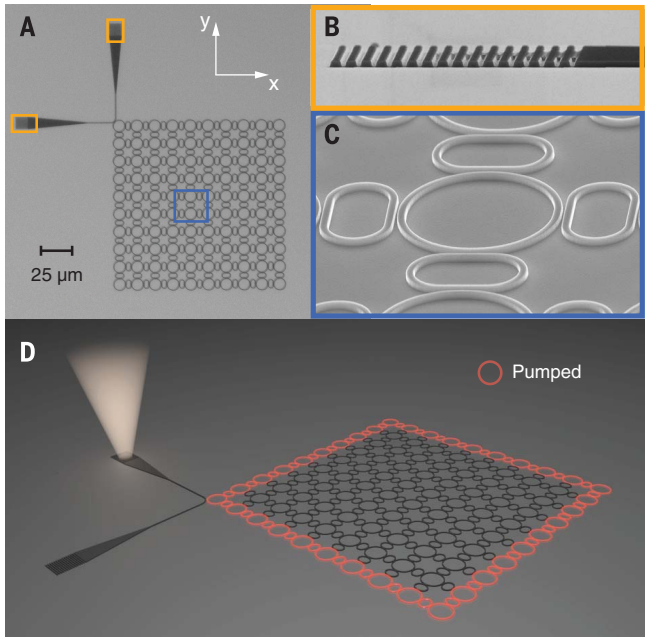


Fig. 1. Topological insulator laser: Lattice geometry. (**A**) Microscope image of an active InGaAsP topological 10 unit cell–by–10 unit cell microresonator array. (**B**) Scanning electron microscopy (SEM) image of the outcoupling grating structures used to probe the array at the orange-outlined locations indicated in (A). (**C**) Magnification of the blue-outlined area indicated in (A), showing a SEM micrograph of a unit cell comprised of a primary ring site surrounded by four identical intermediary racetrack links. (**D**) A schematic of the topological array when pumped along the perimeter to promote lasing of the topological edge mode.

[1]Physics Department and Solid State Institute, Technion, Haifa 32000, Israel. [2]CREOL, College of Optics and Photonics, University of Central Florida, Orlando, FL 32816, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: msegev@techunix.technion.ac.il (M.S.); demetri@creol.ucf.edu (D.N.C.); mercedeh@creol.ucf.edu (M.K.)

2D topological cavity arrays. We show that this topological insulator laser can operate in single mode, even considerably above threshold, with a slope efficiency that is significantly higher than that achieved in their corresponding trivial realizations. Moreover, we observe experimentally that the topological protection leads to more efficient lasing in the 2D array than in the trivial counterpart, even in the presence of defects. Finally, to show the potential of this active system, we assemble a topological system based on *S*-chiral resonators, which can provide new avenues to control the topological features.

Design of the topological insulator laser

We fabricated a 10 unit cell–by–10 unit cell coupled ring-resonator array on an active platform involving vertically stacked 30-nm-thick InGaAsP quantum wells [see (*33*), Fig. 1A]. We coupled the array to a waveguide that acts as an output coupler, which allowed us to interrogate the system using outcoupling gratings (Fig. 1B, corresponding to the yellow framed regions in Fig. 1A). The active lattice investigated here uses a topological architecture suggested in (*26*), which is based on adding gain and loss to the topological passive silicon platform demonstrated in (*15*). This 2D setting comprises a square lattice of ring resonators, which are coupled to each other through link rings (Fig. 1, A and C). The link rings are designed to be antiresonant to the main ring resonators. In this all-dielectric design, the intermediary links are judiciously spatially shifted along the y axis, with respect to the ring resonators, to introduce an asymmetric set of hopping phases. The phase shift is sequentially increased along the y axis in integer multiples of $\pm 2\pi\alpha$, designed here to be $\alpha = 0.25$, where α is proportional to the equivalent magnetic flux quanta passing through a unit cell. In this way, a round trip along any plaquette (consisting of four rings and four links) results in a total accumulated phase of $\pm 2\pi\alpha$, where the sign depends on the direction of the path along this unit cell. This provides the lattice with a synthetic magnetic field and establishes two topologically nontrivial band gaps. The cross section of each ring (500-nm width and 210-nm height) is designed to ensure single transverse-mode conditions at a wavelength of operation of 1550 nm (*33*). The nominal separation between the ring-resonators and off-resonant links is 150 nm, thus leading to two frequency band gaps, each having a width of 80 GHz (0.64 nm). The spectral size of the two band gaps was obtained by experimentally measuring the frequency splitting (0.8 nm) in a binary system of primary resonators, linked by means of an intermediate racetrack ring [see (*33*), part 9]. To promote protected edge-mode lasing, we optically pumped only the outer perimeter of this array at 1064 nm with 10-ns pulses (Fig. 1D). This was achieved using a set of appropriate amplitude masks [see (*33*), part 1]. The intensity structure of the lasing modes was captured by using an InGaAs infrared camera, and their spectral content was then analyzed by using a spectrometer with an array detector (*33*). In what follows, we compare the features of the topological insulator lasers ($\alpha = 0.25$) with those of their trivial counterparts ($\alpha = 0$) under various conditions.

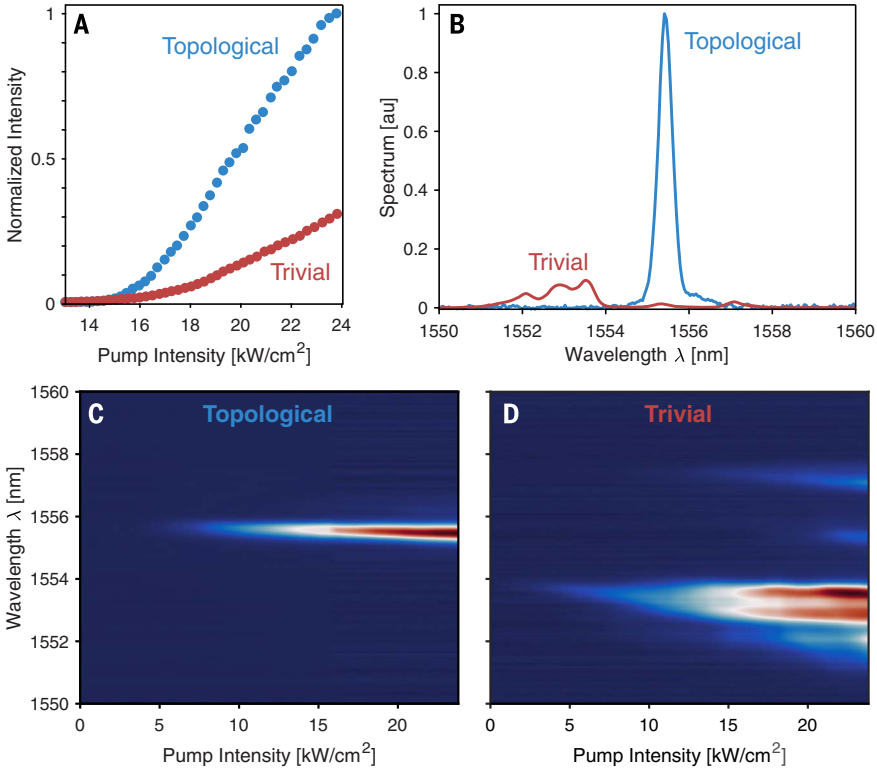


Fig. 2. Slope efficiencies and associated spectra of topological and trivial laser arrays. (**A**) Output intensity versus pump intensity for a 10 unit cell–by–10 unit cell topological array with $\alpha = 0.25$ and its corresponding trivial counterpart ($\alpha = 0$). In this experiment, the enhancement of the slope efficiency is about threefold. (**B**) Emission spectra from a trivial and a topological array when pumped at 23.5 kW/cm^2. au, arbitrary units. (**C** and **D**) Evolution of the spectrum as a function of the pumping intensity for (C) topological and (D) trivial arrays. Single-mode, narrow-linewidth lasing in (C) is clearly evident.

Studying the features of the topological insulator laser

The edge mode can be made to lase by pumping the boundary of the topological array (*26*). In this case, a clear signature of topological lasing would be highly efficient single-mode emission, even at gain values high above the threshold. To observe these features in experiments, we pumped the perimeter of the topological and trivial arrays and measured the lasing output power (integrated over the two outcoupling gratings) and its spectral content. The light-light curves measured for the topological and the trivial arrays (Fig. 2A) clearly show that the topological system lases with a higher efficiency than its trivial counterpart. From their measured spectra (Fig. 2, B and D), we observe that the topological arrays remain single moded over a wide range of pumping densities (Fig. 2C), whereas the trivial arrays (tested over multiple samples) always emit in multiple wavelengths with considerably broader linewidths (Fig. 2D). Importantly, if we only compare the power emitted in the dominant (longitudinal) mode of the topological array to the mode with the highest power in the trivial array, the topological laser outperforms the trivial one by more than an order of magnitude. This difference in performance is attributed to the physical properties of the topological edge modes. The trivial array suffers from several drawbacks. First, the trivial lasing modes extend into the lossy bulk, thus experiencing suppressed emission. Second, the trivial lasing modes try to avoid the output coupler so as to optimize their gain. And finally, because of intrinsic disorder in fabrication, the lasing mode localizes in several different parts of the trivial lattice, each lasing at a different frequency, thereby giving rise to a multimode behavior. Conversely, apart from a weak exponential penetration into the bulk, in topological arrays the edge states are strongly confined to the edge. Moreover, because they are forced to flow around the perimeter, they are always in contact with the output coupler. Finally, because of its inherent topological properties, the lasing edge mode does not suffer from localization, and therefore it uniformly extends around the perimeter (in a single mode), using all the available gain in the system by suppressing any other parasitic mode.

To demonstrate that these active lattices exhibit topological features, we compared their

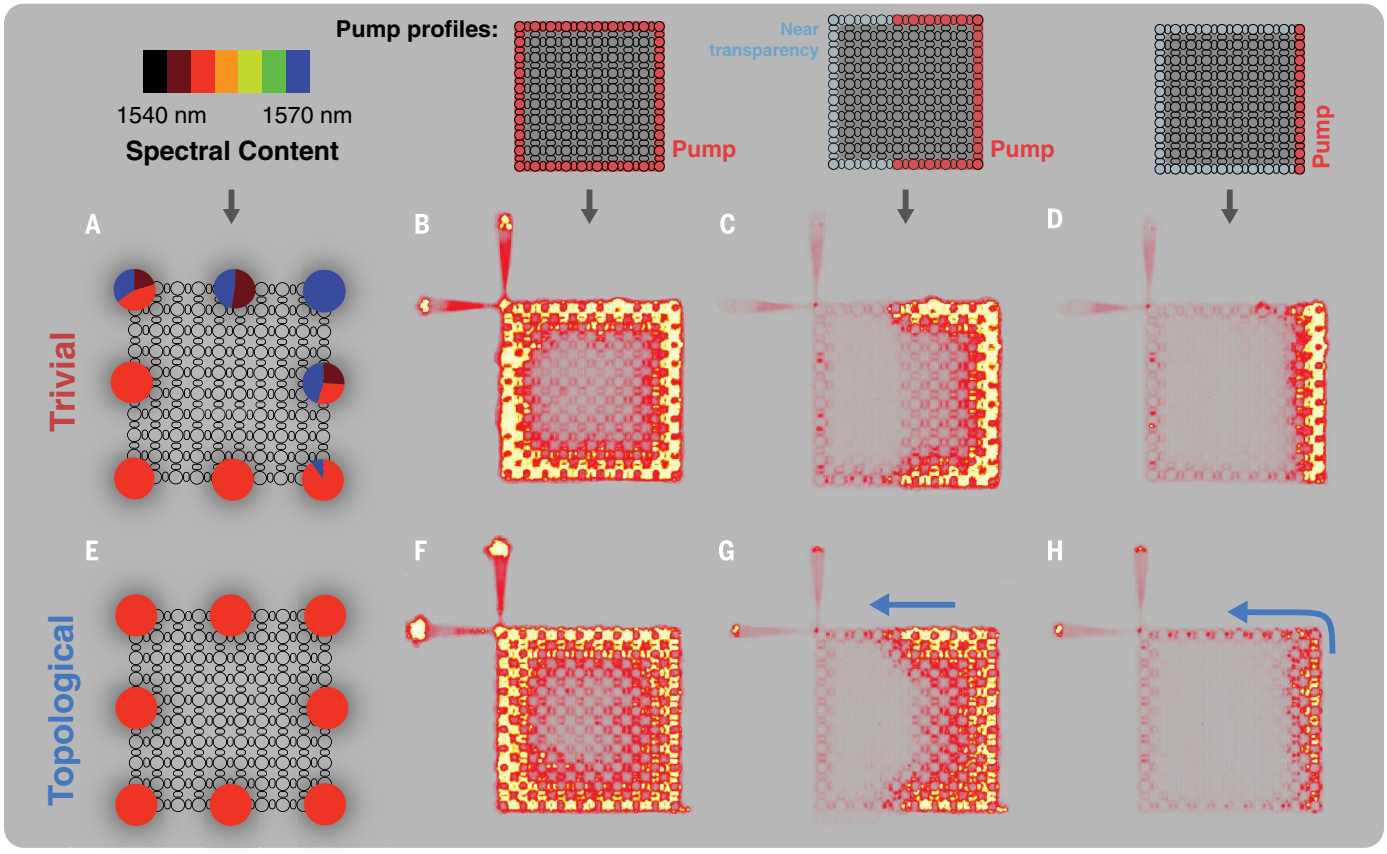


Fig. 3. Lasing characteristics of topological lattices versus those of their corresponding trivial counterparts under several pumping conditions. (**A** to **D**) Trivial and (**E** to **H**) topological lattices. Lasing in a (B) trivial and (F) topological array when their full perimeter is selectively pumped. (A) and (E) represent the spectral content as obtained from specific edge sites of the arrays depicted in (B) and (F), correspondingly. Notice that the topological lattice remains single moded, whereas the trivial one emits in multiple modes. Lasing transport in a (D) trivial array and (H) topological lattice when the right side is pumped. The lasing edge mode in (H) travels (blue arrow) along the unpumped perimeter and exits mostly through the left output grating, whereas the lasing mode in (D) never reaches the output coupler. (C) and (G) present similar results when the pumping region at the right side is further extended into the array. In (G), the lasing edge mode again travels all the way to the output port, whereas the lasing modes of the trivial lattice never reach the extracting ports in (C). This proves that the topological mode travels around the perimeter and always reaches the output port, whereas the lasing modes of the trivial lattice are stationary. The pumping conditions are shown above the labeled panels.

lasing response against that of their trivial counterparts ($\alpha = 0$) when their periphery is pumped. The emission intensity profiles obtained from these two systems are shown in Fig. 3, B and C. To check whether the lasing modes are extended or localized around the perimeter of the lattice, we measured the spectrum of the light emitted from different sites around the arrays (Fig. 3, A and E). For the trivial array, we observed that the spectrum varies around the lattice, with emission occurring over a wide wavelength range, spanning from 1543 to 1570 nm, as shown in Fig. 3A. This is an indication that the trivial array lases in localized domains, each one at a different frequency. By sharp contrast, in the topological array, all sites emit coherently at the same wavelength (Fig. 3E). Such lasing, in a single extended topological edge mode, is a direct manifestation of topologically protected transport. These results are consistent with those presented in Fig. 2. The full spectra of Fig. 3, A and E, are given in (*33*).

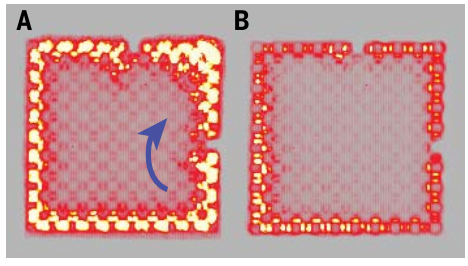


Fig. 4. Robust behavior of the lasing edge mode with respect to defects in a topological array. (**A** and **B**) Lasing response of a (A) topological and (B) trivial array in the presence of two defects intentionally inserted on the periphery, under the same pumping intensity. Note that the edge mode transport in (A) (blue arrow) bypasses the defects, whereas in its trivial counterpart, the lasing occurs from the three separate sections.

Topological transport in these structures was further investigated by selectively pumping the lattice. First, we pumped only one edge of the 2D array, as depicted in Fig. 3D (inset). Under these conditions, the lasing mode in the trivial system is confined to the pumped region (Fig. 3D). In this arrangement, the emission is heavily suppressed both in the bulk as well as along the perimeter, and, consequently, no light is extracted from any of the output grating couplers. By contrast, for the topological array, even when only one side is pumped, the edge mode flows along the periphery, finally reaching the output coupler, as shown in Fig. 3H. In this case, only one output coupler grating emits strongly. This indicates that the lasing mode that reaches the output coupler has a definite chirality in each ring. Given that the emission is in a single mode, one can conclude that lasing takes place in only one topological mode. To show that in the topological case, it is the edge mode that lases,

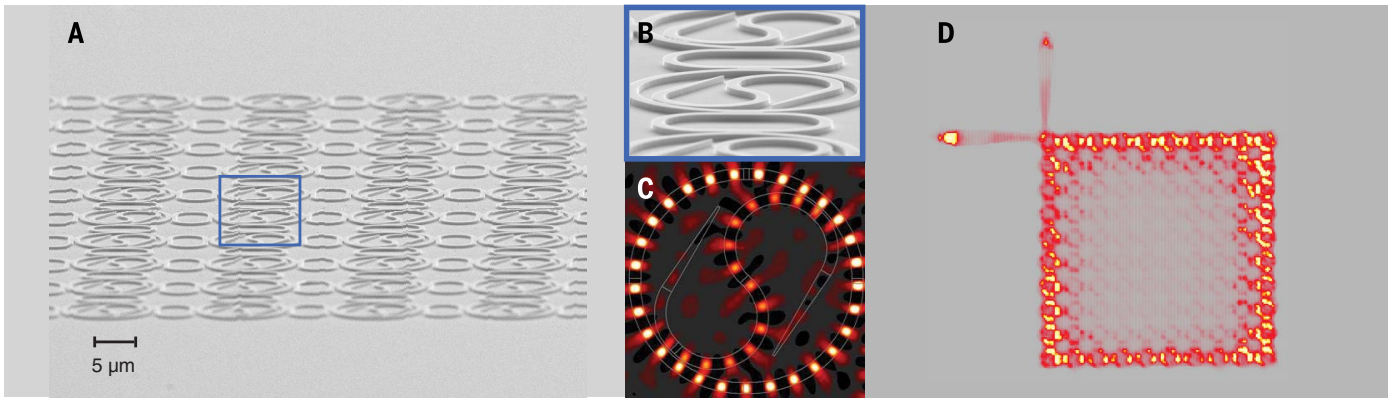


Fig. 5. Topological active array involving chiral *S*-microresonator elements. (**A**) SEM image of a 10 unit cell–by–10 unit cell topological array. The primary resonators feature an internal *S*-bend for enforcing, in this case, a right spin, whereas the intermediate link design is the same as in Fig. 1C. (**B**) A magnification of the SEM micrograph in (A) of the basic elements involved. (**C**) Field distribution in an individual *S*-element, as obtained from finite-difference time-domain simulations. (**D**) Measured intensity profile associated with the lasing edge mode in a topological array with $\alpha = 0.25$. In this system, the perimeter is selectively pumped, and the energy in each ring circulates in a counterclockwise manner, as indicated by the radiation emerging from the extracting ports.

whereas in the trivial case the bulk states lase, we expanded the pumping region at the right edge (Fig. 3C, inset). In the trivial case, even though pumping over a larger area is now provided, still no laser light reaches the output ports (Fig. 3C). This means that there are no traveling edge modes that could reach the output ports. Conversely, for the topological array, the lasing edge mode reaches the output coupler with a fixed chirality within each ring (Fig. 3G). This shows that the topological lasing mode extends around the perimeter and travels all the way to the output ports, whereas the lasing in the trivial case occurs in stationary localized modes. In this vein, we tested multiple samples and found that the same features consistently emerged in a number of different designs (different resonance frequencies, couplings, etc.) in a universal manner. A video showing the behaviors of partially pumped topological and trivial arrays are provided as supplementary materials [see (*33*), part 10] and more features associated with these arrays can be found in [(*33*), parts 3 and 5].

Introducing defects

Next, we studied lasing in a topological structure and in its trivial counterpart, in the presence of defects, which are intentionally introduced into the structures. We removed specific microrings along the perimeter, where pumping is provided (i.e., we remove two gain elements). Figure 4 shows the light emission from these two types of structures. These results demonstrate that in a topological system (Fig. 4A), light is capable of bypassing the defects by penetrating into the bulk and displaying lasing in an extended edge mode of almost uniform intensity. Conversely, the intensity of the emitted light in the trivial structure is considerably suppressed (Fig. 4B), and the defects effectively subdivide the perimeter into separate regions that lase independently (both measurements were performed at identical pump power levels). Hence, the topological insulator laser is robust against defects, even when introduced into the gain regions. [Further evidence of this robustness in the presence of disorder is given in (*33*), part 8.]

Laser based on an array of chiral *S*-bend elements

Having established the underlying concepts of the topological insulator laser and the promise it holds for exploring new aspects of topological physics, specifically in active media, it is interesting to discuss some of the directions to which these ideas can lead. As an example, embarking on fundamental aspects such as Maxwell's reciprocity in the presence of nonlinearity, as well as on potential applications, we modified the individual resonators in the topological insulator laser so as to break the symmetry between the clockwise (CW) and counterclockwise (CCW) modes in each ring. Instead of using conventional rings, we used a special *S*-bend design (*34*) for each primary cavity element in the topological lattice (Fig. 5, A and B). The intermediary links remained the same as in the previous designs (Fig. 1A). In this system, each laser microresonator selectively operates in a single spinlike manner, i.e., in either the CW or the CCW direction, by exploiting gain saturation and energy recirculation among these modes. The *S*-chiral elements involved, in the presence of nonlinearity (gain saturation) and the spatial asymmetry of the *S*-bends, add unidirectionality to the topological protection of transport. In the experiments described in Fig. 5, we observe suppression of more than 12 dB between the right- and left-hand spins in each resonator [see (*33*), part 6]. Finite-difference time-domain simulations also indicate that the differential photon lifetime between the right and left spins in these *S*-bend cavities is about 3 ps, corresponding to an equivalent loss coefficient of 10 cm^{-1} [more details are outlined in (*33*), part 7]. The field distribution in the prevalent spinning mode in these active *S*-resonators is shown in Fig. 5C, featuring a high degree of power recirculation through the *S*-structure that is responsible for the spinlike mode discrimination. The corresponding intensity profile associated with this unidirectional edge-mode energy transport is shown in Fig. 5D. In this case, energy is predominantly extracted from only one of the two outcoupling gratings (with a ~10-dB rejection ratio), which is an indication of unidirectional energy flow in the rings of this topological array.

Discussion

Our all-dielectric topological insulator laser exploits its topological features to enhance the lasing performance of a 2D array of microresonators, making them lase in unison in an extended topologically protected scatter-free edge mode. The observed single longitudinal-mode operation leads to a considerably higher slope efficiency as compared to a corresponding topologically trivial system. The systems described here are based on contemporary fabrication technologies of semiconductor lasers, without need for magnetic units of exotic materials. Our results provide a route for developing a new class of active topological photonic devices, especially arrays of semiconductor lasers, that can operate in a coherent fashion with high slope efficiencies.

REFERENCES AND NOTES

1. D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. Den Nijs, Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982). doi: 10.1103/PhysRevLett.49.405
2. M. Hasan, C. Kane, Colloquium: Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010). doi: 10.1103/RevModPhys.82.3045
3. X.-L. Qi, S.-C. Zhang, Topological insulators and superconductors. *Rev. Mod. Phys.* **83**, 1057–1110 (2011). doi: 10.1103/RevModPhys.83.1057
4. K. V. Klitzing, G. Dorda, M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980). doi: 10.1103/PhysRevLett.45.494
5. C. L. Kane, E. J. Mele, Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95**, 226801 (2005). pmid: 16384250

6. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006). doi: 10.1126/science.1133734; pmid: 17170299
7. M. König *et al*., Quantum spin hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007). doi: 10.1126/science.1148047; pmid: 17885096
8. F. D. Haldane, S. Raghu, Possible realization of directional optical waveguides in photonic crystals with broken time-reversal symmetry. *Phys. Rev. Lett.* **100**, 013904 (2008). doi: 10.1103/PhysRevLett.100.013904; pmid: 18232766
9. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature* **461**, 772–775 (2009). doi: 10.1038/nature08293; pmid: 19812669
10. L. Lu, J. D. Joannopoulos, M. Soljačić, Topological photonics. *Nat. Photonics* **8**, 821–829 (2014). doi: 10.1038/nphoton.2014.248
11. M. Hafezi, E. Demler, M. Lukin, J. Taylor, Robust optical delay lines via topological protection. *Nat. Phys.* **7**, 907–912 (2011). doi: 10.1038/nphys2063
12. K. Fang, Z. Yu, S. Fan, Realizing effective magnetic field for photons by controlling the phase of dynamic modulation. *Nat. Photonics* **6**, 782–787 (2012). doi: 10.1038/nphoton.2012.236
13. A. B. Khanikaev *et al*., Photonic topological insulators. *Nat. Mater.* **12**, 233–239 (2013). doi: 10.1038/nmat3520; pmid: 23241532
14. M. C. Rechtsman *et al*., Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013). doi: 10.1038/nature12066; pmid: 23579677
15. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, Imaging topological edge states in silicon photonics. *Nat. Photonics* **7**, 1001–1005 (2013). doi: 10.1038/nphoton.2013.274
16. L. Tarruell, D. Greif, T. Uehlinger, G. Jotzu, T. Esslinger, Creating, moving and merging Dirac points with a Fermi gas in a tunable honeycomb lattice. *Nature* **483**, 302–305 (2012). doi: 10.1038/nature10871; pmid: 22422263
17. M. Atala *et al*., Direct measurement of the Zak phase in topological Bloch bands. *Nat. Phys.* **9**, 795–800 (2013). doi: 10.1038/nphys2790
18. C. L. Kane, T. C. Lubensky, Topological boundary modes in isostatic lattices. *Nat. Phys.* **10**, 39–45 (2013). doi: 10.1038/nphys2835
19. Z. Yang *et al*., Topological acoustics. *Phys. Rev. Lett.* **114**, 114301 (2015). doi: 10.1103/PhysRevLett.114.114301; pmid: 25839273
20. R. Fleury, A. B. Khanikaev, A. Alù, Floquet topological insulators for sound. *Nat. Commun.* **7**, 11744 (2016). doi: 10.1038/ncomms11744; pmid: 27312175
21. Z. Yu, G. Veronis, Z. Wang, S. Fan, One-way electromagnetic waveguide formed at the interface between a plasmonic metal under a static magnetic field and a photonic crystal. *Phys. Rev. Lett.* **100**, 023902 (2008). doi: 10.1103/PhysRevLett.100.023902; pmid: 18232868
22. X. Cheng *et al*., Robust reconfigurable electromagnetic pathways within a photonic topological insulator. *Nat. Mater.* **15**, 542–548 (2016). doi: 10.1038/nmat4573; pmid: 26901513
23. G. Jotzu *et al*., Experimental realization of the topological Haldane model with ultracold fermions. *Nature* **515**, 237–240 (2014). doi: 10.1038/nature13915; pmid: 25391960
24. M. Aidelsburger *et al*., Measuring the Chern number of Hofstadter bands with ultracold bosonic atoms. *Nat. Phys.* **11**, 162–166 (2015). doi: 10.1038/nphys3171
25. A. P. Slobozhanyuk *et al*., Experimental demonstration of topological effects in bianisotropic metamaterials. *Sci. Rep.* **6**, 22270 (2016). doi: 10.1038/srep22270; pmid: 26936219
26. G. Harari *et al*., Topological insulator laser: Theory. *Science* **359**, eaar4003 (2018). doi: 10.1126/science.aar4003
27. P. St-Jean, V. Goblot, E. Galopin, A. Lemaître, T. Ozawa, L. Le Gratiet, I. Sagnes, J. Bloch, A. Amo, Lasing in topological edge states of a 1D lattice. arXiv:1704.07310 [cond-mat.mes-hall] (24 April 2017).
28. M. Parto, S. Wittek, H. Hodaei, G. Harari, M. A. Bandres, J. Ren, M. C. Rechtsman, M. Segev, D. N. Christodoulides, M. Khajavikhan, Complex edge-state phase transitions in 1D topological laser arrays. arXiv:1709.00523 [physics.optics] (2 September 2017).
29. H. Zhao, P. Miao, M. H. Teimourpour, S. Malzard, R. El-Ganainy, H. Schomerus, L. Feng, Topological hybrid silicon microlasers. arXiv:1709.02747 [physics.optics] (8 September 2016).
30. G. Harari *et al*., in *Conference on Lasers and Electro-Optics* (OSA Technical Digest, Optical Society of America, paper FM3A.3, 2016).
31. S. Wittek *et al*., in *Conference on Lasers and Electro-Optics* (OSA Technical Digest, Optical Society of America, paper FTh1D.3, 2017).
32. B. Bahari *et al*., Nonreciprocal lasing in topological cavities of arbitrary geometries. *Science* **358**, 636–640 (2017). doi: 10.1126/science.aao4551; pmid: 29025992
33. Materials and methods are available as supplementary materials.
34. J. P. Hohimer, G. A. Vawter, D. C. Craft, Unidirectional operation in a semiconductor ring diode laser. *Appl. Phys. Lett.* **62**, 1185–1187 (1993). doi: 10.1063/1.108728

ACKNOWLEDGMENTS

Funding: The authors gratefully acknowledge financial support from the Israel Science Foundation, Office of Naval Research (N0001416-1-2640), NSF (ECCS1454531, DMR-1420620, ECCS 1757025), European Commission Non-Hermitian Quantum Wave Engineering (NHQWAVE) project (MSCA-RISE 691209), Air Force Office of Scientific Research (FA9550-14-1-0037), United States–Israel Binational Science Foundation (2016381), German-Israeli Project Cooperation (Deutsch-Israelische Projektkooperation) program, and Army Research Office (W911NF-16-1-0013, W911NF-17-1-0481). M.S. thanks M. Karpovsky and B. Shillman for their support that came at a critical time. **Author contributions:** All authors contributed to all aspects of this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in this paper are available in the manuscript and in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/eaar4005/suppl/DC1
Materials and Methods
Figs. S1 to S11
Reference (*35*)
Movies S1 and S2

3 November 2017; accepted 17 January 2018
Published online 1 February 2018
10.1126/science.aar4005

RESEARCH ARTICLE SUMMARY

CIRCADIAN RHYTHMS

Diurnal transcriptome atlas of a primate across major neural and peripheral tissues

Ludovic S. Mure, Hiep D. Le, Giorgia Benegiamo, Max W. Chang, Luis Rios, Ngalla Jillani, Maina Ngotho, Thomas Kariuki, Ouria Dkhissi-Benyahya, Howard M. Cooper,* Satchidananda Panda*

INTRODUCTION: The interaction among cell-autonomous circadian oscillators—daily cycles of activity–rest and feeding–fasting—produces diurnal rhythms in gene expression in almost all animal tissues. These rhythms control the timing of a wide range of functions across different organs and brain regions, affording optimal fitness. Chronic disruption of these rhythms predisposes to and are hallmarks of numerous diseases and affective disorders.

RATIONALE: Time-series gene expression studies in a limited number of tissues from rodents have shown that 10 to 40% of the genome exhibits a ~24-hour rhythm in expression in a tissue-specific manner. However, rhythmic expression data from diverse tissues and brain regions from humans or our closest primate relatives is rare. Such multitissue diurnal gene expression data are necessary for gaining mechanistic understanding of how spatiotemporal orchestration of gene expression maintains normal physiology and behavior. We used a RNA sequencing technique to assess gene expression in major tissues and brain regions from baboons (a primate closely related to humans) housed under a defined 24-hour light–dark and feeding–fasting schedule.

RESULTS: We assessed gene expression in 64 different tissues and brain regions of male baboons, collected every 2 hours over the 24-hour day. Tissue-specific transcriptomes in baboon were comparable with that from humans (Human GTEx data set). We detected >25,000 expressed transcripts, including protein-coding and -noncoding RNAs. Nearly 11,000 genes were commonly expressed in all tissues. These universally expressed genes (UEGs) encoded for basic cellular functions such as transcription, RNA processing, DNA repair, protein homeostasis, and cellular metabolism. The remainders were expressed in distinct sets of tissues, with ~1500 genes expressed exclusively in a single tissue.

Rhythmic transcripts were found in all tissues, but the number of cycling transcripts varied from ~200 to >3000 in a given tissue, with only limited overlap in the repertoire of rhythmic transcripts between tissues. Of the 11,000 UEGs, the vast majority (96.6%) showed 24-hour rhythmicity in at least one tissue. A majority (>80%) of the 18,000 protein-coding genes detected also exhibited 24-hour rhythms in expression. The most enriched rhythmic transcripts across tissues were core clock components and their immediate output targets. However, their relative abundance and robustness of daily rhythms varied across tissues. Considered at the organismal level, global rhythmic transcription in 64 tissues organized into bursts of peak transcription, during early morning and late afternoon (when 11,000 transcripts reach their peak level). By contrast, during a relative "quiescent phase" in early evening that coincides with the onset of sleep and no food intake, only 700 rhythmic transcripts reach their peak expression level.

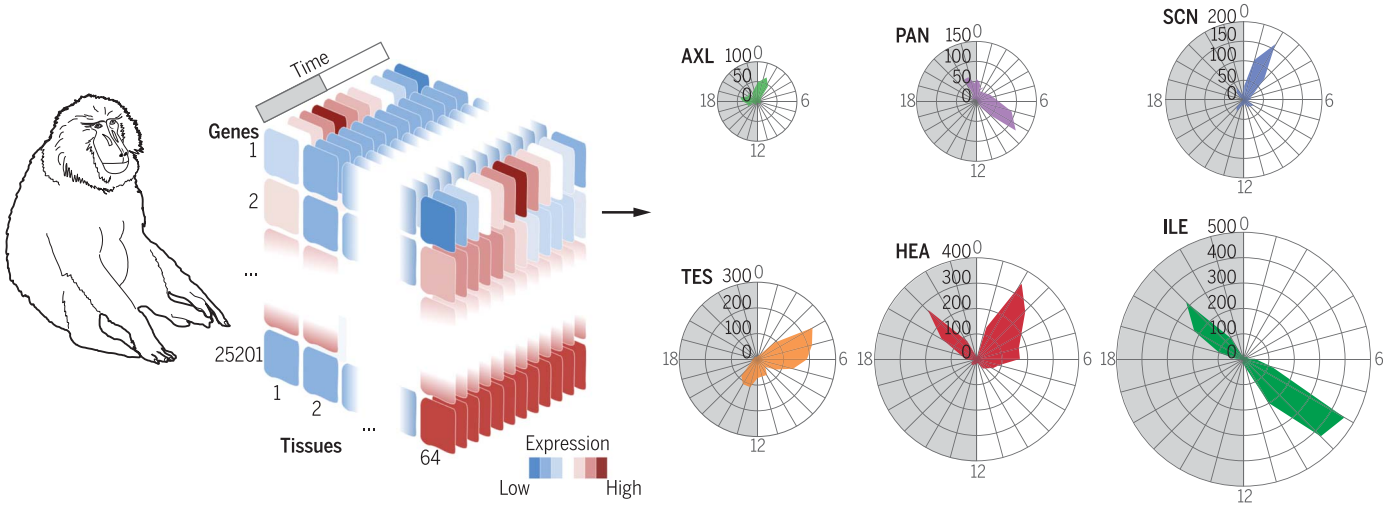
ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aao0318

CONCLUSION: The daily expression rhythms in >80% of protein-coding genes, encoding diverse biochemical and cellular functions, constitutes by far the largest regulatory mechanism that integrates diverse biochemical functions within and across cell types. From a translational point of view, rhythmicity may have a major impact in health because 82.2% of genes coding for proteins that are identified as druggable targets by the U.S. Food and Drug Administration show cyclic changes in transcription. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: howard.cooper@inserm.fr (H.M.C.); satchin@salk.edu (S.P.)
Cite this article as L. S. Mure *et al.*, *Science* **359**, eaao0318 (2018). DOI: 10.1126/science.aao0318

Spatiotemporal gene expression atlas of a primate. (**Left**) Gene expression analysis across 64 tissues of a diurnal primate sampled over the 24-hour day shows that 82% of protein-coding genes are rhythmic in at least one tissue. (**Right**) Rhythmic expression is tissue-specific and confers an additional layer of regulation and identity to the transcriptome of a given tissue.



RESEARCH ARTICLE

CIRCADIAN RHYTHMS

Diurnal transcriptome atlas of a primate across major neural and peripheral tissues

Ludovic S. Mure,[1] Hiep D. Le,[1] Giorgia Benegiamo,[1] Max W. Chang,[1,2] Luis Rios,[1] Ngalla Jillani,[3] Maina Ngotho,[3] Thomas Kariuki,[3] Ouria Dkhissi-Benyahya,[4] Howard M. Cooper,[4]* Satchidananda Panda[1]*

Diurnal gene expression patterns underlie time-of-the-day–specific functional specialization of tissues. However, available circadian gene expression atlases of a few organs are largely from nocturnal vertebrates. We report the diurnal transcriptome of 64 tissues, including 22 brain regions, sampled every 2 hours over 24 hours, from the primate *Papio anubis* (baboon). Genomic transcription was highly rhythmic, with up to 81.7% of protein-coding genes showing daily rhythms in expression. In addition to tissue-specific gene expression, the rhythmic transcriptome imparts another layer of functional specialization. Most ubiquitously expressed genes that participate in essential cellular functions exhibit rhythmic expression in a tissue-specific manner. The peak phases of rhythmic gene expression clustered around dawn and dusk, with a "quiescent period" during early night. Our findings also unveil a different temporal organization of central and peripheral tissues between diurnal and nocturnal animals.

Daily cycles in behavior, physiology, and metabolism are presumed to arise from coordinated gene expression rhythms in multiple tissues and brain regions. Gene expression rhythms have been cataloged in a handful of tissues in model organisms (*1*–*3*). Among these, the mouse, which is nocturnal and evolutionarily diverged from humans 75 million years ago, is considered as a representative model for mammalian gene expression rhythms. However, these data sets are primarily collected from the laboratory mouse (*Mus musculus*), which is active mainly at night, generally fed ad libitum, and has fragmented daytime sleep. Although standard laboratory mouse strains do not produce melatonin, wild mice secrete both melatonin and cortisol during the night time. By contrast, diurnal primates, including humans, are active and eat discrete meals during the daytime, produce melatonin at night during their phase of consolidated sleep, and show a daytime peak in corticosteroids (*4*). Despite these differences in timing of behavior, a study of deoxy-glucose uptake has led to the view that the phase of the master circadian oscillator in the suprachiasmatic nucleus (SCN) is similar in nocturnal rodents and diurnal primates (*5*). However, limited gene expression data from human blood (*6*) and post mortem cortex samples (*7*) have indicated that expression, phase, and amplitude of daily gene expression rhythms in primates differ from that in mice. We therefore analyzed diurnal gene expression profiles of 768 samples across 64 different central and peripheral tissues (Fig. 1 and table S1) from young male baboons (*Papio anubis*), a close primate relative of humans (diverged 24 million years ago).

[1]Regulatory Biology Laboratory, Salk Institute for Biological Studies, 10010, North Torrey Pines Road, La Jolla, CA 92037, USA. [2]Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. [3]Institute of Primate Research (IPR), National Museums of Kenya, Nairobi, Kenya. [4]Université Lyon, Universite Claude Bernard Lyon 1, Inserm, Stem Cell and Brain Research Institute U1208, 69500 Bron, France.
***Corresponding author. Email: howard.cooper@inserm.fr (H.M.C.); satchin@salk.edu (S.P.)**

Transcriptional complexity across tissues

Animals housed individually in seminatural conditions under a 12 hours light–12 hours dark cycle showed highly stable and synchronous phases of activity, body temperature, and cortisol concentration in the blood across the 24-hour cycle (Fig. 2A and fig. S1, A to E). Tissues were collected every 2 hours across the 24-hour day (ZT0, -2, -4, -6, -8, -10, -12, -14, -16, -18, -20, and -22, where ZT0 is the time when light is ON and ZT12 is when light is OFF) and were flash frozen within 2 hours of collection. Poly A^+ RNA^- sequencing performed on an Illumina platform (*8*) produced an average of 15.4 million 50–base pair single-end (SE50) mapped reads per sample (table S2). The sequence reads were mapped to the *P. anubis* genome data set (PapAnu2.0, International Nucleotide Sequence Database Collaboration Assembly, GenBank Assembly ID GCA_000264685, version 88) and annotated with Ensembl (Full genebuild, released July 2014). This ensemble annotation comprises 19,210 protein-coding genes (PCGs), 9272 noncoding RNAs [including 5888 small nuclear RNA (snRNA), 1669 long noncoding RNA (lncRNA), and 1715 miscellaneous RNA (miscRNA)] and 720 pseudogenes. At a threshold of 0.1 FPKM (fragments per kilobase of transcript per million mapped reads) for gene expression (*9*), 97.3% of PCGs (18,684) and 89.6% of lncRNAs (1495) were detected in at least one organ (Fig. 2B). Principal component analyses of transcriptome from all 768 samples revealed multidimensional clustering or hierarchical clustering of samples from similar tissues and tissue types (Fig. 2E and fig. S2, A and B). Overall, the relative distance between tissue clusters was similar to that observed for comparable samples in the Nonhuman Primate Reference Transcriptome Resource database (13 tissues from one female baboon) (fig. S2C) (*10*) and in 41 human tissue transcriptomes (*9*, *11*). Although brain regions showed little separation in multidimensional clustering, hierarchical clustering of these tissues revealed grouping of distinct clusters of eye, basal ganglia, hypothalamus, thalamus (THA), and cortices (fig. S2B). The unbiased clustering of ontologically related organs and brain nuclei validated dissection, sampling, and analyses protocols.

Out of the 25,098 transcripts detected in at least one organ, 10,989 transcripts were common to all 64 tissues. These ubiquitously expressed genes (UEGs) encode gene products necessary for basic cellular functions, including basic transcription, RNA processing, DNA repair, protein homeostasis, and secretory function (table S3). The remaining 14,109 genes were expressed in distinct sets of tissues, with 1473 genes expressed exclusively in one tissue (table S4 and fig. S2D). Tissues differed in the number of expressed genes; the top quartile included highly heterogeneous tissues such as the arcuate nucleus (ARC), supra optic nucleus (SON), ventromedial regions of the hypothalamus (VMH), and the testis (TES). Each of these tissues expressed >17,000 genes, whereas the bottom quartile included relatively fewer heterogeneous tissues such as the abdominal and gastrocnemian muscles (MUA and MUG, respectively) and expressed 14,000 to 16,000 genes (table S1). The number of expressed genes does not entirely represent the complexity of expression. Indeed, transcriptional complexity, expressed as the fraction of total RNAs accounted for by the 10 or 100 most frequently expressed genes (table S5) (*9*), showed wide distribution. The pancreas (PAN) was the least complex tissue (top 10 expressed genes, accounting for >60% of all sequenced tags), whereas the TES transcriptional output was the most complex (Fig. 2D). Although several brain regions are among the top quartile in terms of number of expressed genes, they showed a relatively low complexity of gene expression. In 16 out of 24 brain regions, the 10 most highly expressed genes accounted for >50% of all genes expressed (Fig. 2D). Thus, transcriptome signature confers tissue identity in this primate as it does in humans (*9*, *11*).

Diurnal gene expression across tissues

Diurnal rhythms of gene expression from the 12 time points for all tissues were analyzed with

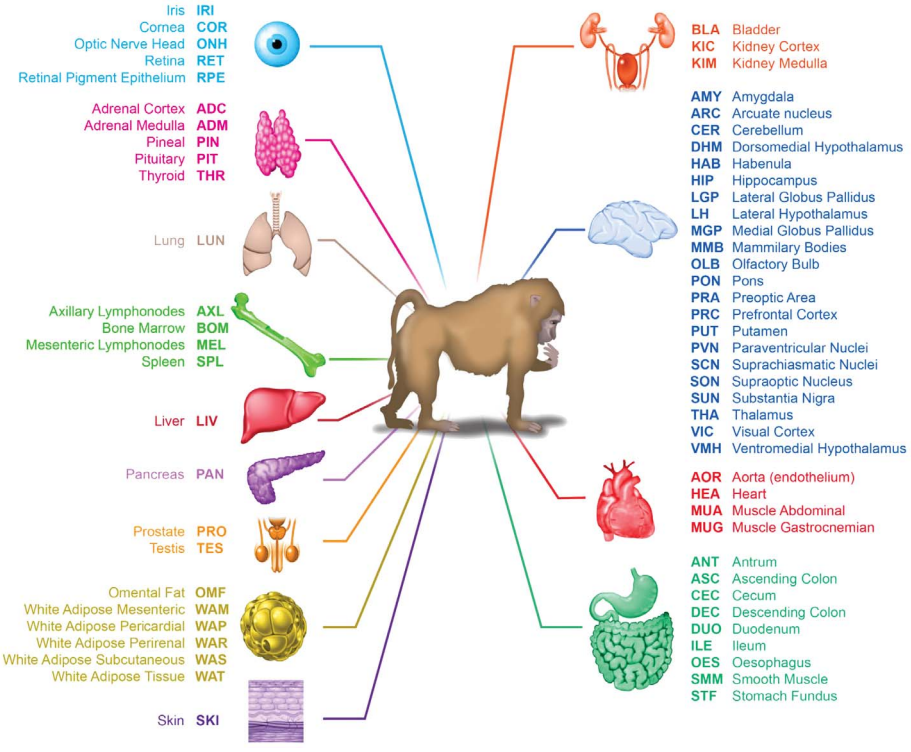


Fig. 1. Tissue collection. List of the 64 tissues collected, according to system and functional types (including 22 brain regions and nuclei).

MetaCycle (*12*), an R package that integrates multiple algorithms to determine periodicity and rhythmicity. Transcripts with ~24-hour rhythms are discussed further. All 64 tissues examined showed rhythmic genes (Fig. 3A and table S6), but there was a clear difference between tissue-specific gene expression and tissue-specific rhythmic gene expression. In general, the number of rhythmic transcripts in a given tissue was not correlated to the number of genes expressed in the same tissue (fig. S3A). Furthermore, the fraction of the transcriptome detected as rhythmic and the extent of rhythmic genes shared among tissues were more complex than spatial gene expression signatures of tissues (Fig. 3C and table S7). In thyroid (THR), stomach fundus (STF), gastrocnemian muscle (MUG), paraventricular nucleus (PVN), and prefrontal cortex (PRC), more than 3000 genes were detected to cycle, whereas in the pineal (PIN), mesenteric lymphonodes (MEL), supraoptic nucleus (SON), lateral hypothalamus (LH), and bone marrow (BOM), less than 200 genes had 24-hour rhythms of expression (Fig. 3, A and B). More than 50% of genes detected in a given tissue were also expressed in all 64 tissues. However, even among tissues with >1000 rhythmic genes, less than 1% (<10) rhythmic genes were shared. Becuse of this limited overlap of rhythmic genes between tissues, and even though <4000 genes cycle in most tissues, 16,442 genes (65.5% of all expressed genes) (table S8) were rhythmic, with a ~24-hour period in at least one tissue (Fig. 3, A and C). Rhythmic transcripts were particularly enriched in PCGs (fig. S3, B and C), accounting for 81.7% of expressed PCGs (15,269 of 18,684). Thus, as seen in rodents (*1*, *13*, *14*) and insects (*2*), rhythmic transcription is tissue-specific. More importantly, more than two-thirds of the transcriptome showed ~24-hour rhythms in expression.

Tissue-specific rhythmic expression is not a function of tissue-specific gene expression. Surprisingly, in nearly all tissues, the majority of cycling genes were UEGs (76.7% on average) (Fig. 3A). Of the total 10,989 UEGs, the vast majority (10,607 genes, 97%) cycled in at least one organ (Fig. 3D), leaving only 382 UEGs not detected as rhythmic in any of the 64 tissues (table S9). Comparison of the number of cycling and noncycling UEGs as a function of the number of tissues analyzed indicated that with an increased sampling of cell and tissue types, all UEGs may indeed cycle in at least one organ (Fig. 3D). This implies that although the rhythmic transcriptome is tissue-specific, the majority of rhythmic genes is still expressed in each tissue but may not cycle in other tissues. This suggests that rhythmic genes undergo tissue-specific transcriptional regulation. Because UEGs participate in fundamental cell biological processes, circadian regulation of subsets of these genes in different organs may reflect a mechanism for organ-specific modulation of basic cellular function. For example, in endocrine organs, the diurnal regulation of exocytosis specifically alters basic cellular functions to produce rhythmic release of endocrine factors.

Because MetaCycle detects rhythmic genes regardless of transcript abundancy (*12*), we also examined the relation between gene rhythmicity and amount of expression. By calculating the fraction of rhythmic genes in each decile of average expression values, we found that in 60% of the tissues, the most rhythmic genes are among the genes that account for the top three deciles of transcriptional output (the 30% most expressed genes), whereas less than 1% of rhythmic genes were found in the decile of the least expressed genes (Fig. 3E). Such rhythmic expression of highly expressed genes supports the hypothesis that rhythmic transcription reduces or optimizes energy expenditure in gene expression (*15*).

We analyzed the times of peak expression (phase) of rhythmic genes. The phases in each tissue were not randomly distributed throughout the 24-hour day. The peak phases of expression of rhythmic genes for nearly all organs were largely clustered in one or two narrow temporal windows (of <6 hours) (Fig. 4). In tissues such as amygdala (AMY), habenula (HAB), optic nerve head (ONH), PVN, SCN, VMH, LH, lateral globus pallidus (LGP), substantia nigra (SUN), PAN, PRC, prostate (PRO), and spleen (SPL), more than two-thirds of rhythmic genes peaked within a narrow (<6 hour) interval, whereas some tissues—including adrenal gland cortex (ADC), aorta (AOR), cerebellum (CER), dorsomedial hypothalamus (DMH), THA, and putamen (PUT)—had genes that peaked within two distinct time intervals. However, even for anatomically adjacent tissues these phase clusters were temporally distinct. For example, in the kidney cortex (KIC) the main phase cluster was at ZT2, whereas in the medulla (KIM), it appeared at ZT7. Similarly, along the digestive track the phase clusters of esophagus (OES), STF, antrum (ANT), duodenum (DUO), ileum (ILE), ascending colon (ASC), and descending colon (DSC) were distinct (Fig. 4). A compilation of peak phases of expression across all rhythmic genes in all tissues revealed two major peaks, early afternoon (11,146 genes peak at ZT8) and late night or dawn (5127 genes peak at ZT22), with a distinct trough around midnight (696 genes peak at ZT16) (Fig. 3F and fig. S3D). This relative "quiescent zone" in the first half of the night (ZT14 to -19) is thus a distinct feature of rhythmic transcriptional output in the diurnal primate.

Tissue-specific features of the molecular clock

Because the number of common rhythmic genes between tissues is rather limited, we examined which genes were most widely detected as rhythmic in multiple tissues. No single transcript was scored significantly rhythmic in all 64 tissues. Genes encoding for the core circadian clock components and their immediate output targets were among the genes detected as rhythmic in most (but not all) tissues (Fig. 5A and table S10). The recently described circadian repressor gene *Ciart* [Circadian-associated transcriptional repressor (or *Chrono*)] was detected cycling in the most number of tissues (52 tissues) (Fig. 5A and fig S4A). For the 13 transcriptional regulators considered canonical components of the circadian

clock (*Per1-3*, *Cry1-2*, *Clock*, *Bmal1*, *Npas2*, *Rora*, *Rorb*, *Rorc* and *Nr1d1-2*), their relative abundance and robustness of daily rhythms varied across tissues. Of the 13 components, 11 were detected in all 64 tissues, whereas *Rorb* and *Rorc* were expressed in 60 and 62 tissues, respectively. The median expression of clock components also varied across tissues, with paralog pairs (*Cry1* and *Cry2*; *Nr1d1* and *Nr1d2*) exhibiting mutually complementary expression in different tissues (Fig. 5B and fig. S4B). Within the hypothalamus, the SCN, PVN, preoptic area (PRA), and DMH showed a distinctive feature: Both *Cry* and *Nr1d* paralogous pairs were expressed in equivalent amounts (Fig. 5B), whereas in most other brain regions and organs, one of the members was expressed in higher amounts. Although core clock genes did not oscillate in all tissues, nine of the clock components showed daily rhythms in more than 20 tissues (Fig. 5A). Furthermore, the amplitude of oscillation of core clock components varied between tissues (Fig. 5B). The heterogeneity in the overall transcript abundance and rhythmicity of expression of circadian clock components implies different composition of core activators, repressors, and modulators in different tissues, generating tissue-specific characteristics of core oscillators and input and output mechanisms.

Although the transcript abundance and rhythmic properties of clock gene expression varied across tissues, the peak phase of expression of each clock component was mostly restricted to a roughly 6-hour window (Fig. 5C, fig. S4C, and table S8). The mean peak times of expression for *Bmal1* and *Npas2* were similar (Bmal1, ZT13.3 ± 0.2; NPAS2, ZT14.3 ± 0.5). Although *Bmal1* (heterodimerized with CLOCK or NPAS2) is considered to activate *Per*, *Cry*, *Nr1d*, and *Ror* expression, the peak phases of expression of these targets were not identical. *Per1*, *Per2*, and *Cry2* shared similar peak phases of expression (ZT2.4, -2.9, and -2.5, respectively), whereas *Cry1* peaked around ZT11. *Nr1d1* and *Nr1d2* peaked around ZT22 to ZT23, and *Rorc* peaked around ZT8 (fig. S4D and table S11). Although different combinations of paralogous clock components and their abundance provide evidence for tissue-specific characteristics of the core oscillator, the relative phases of expression of these components are remarkably conserved across tissues.

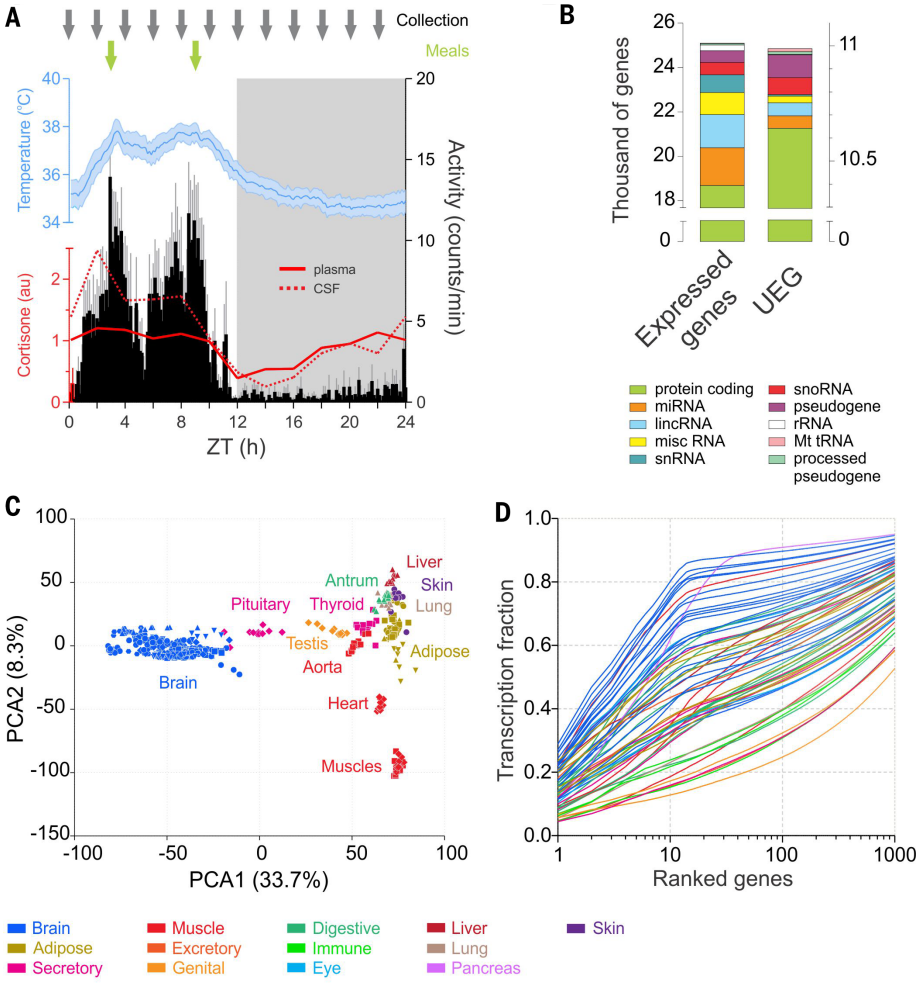


Fig. 2. Transcriptome complexity. (**A**) Timing of sampling (gray arrows) across the 24-hour cycle (dark phase in gray shading) and timing of meals (green arrows) superimposed on the daily cycles of locomotor activity (black), body temperature (blue) (averages for 12 animals, ± 95% confidence interval, over the 14 days preceding the sample collection, average activity and temperature periods respectively 23.98 ± 0.02 and 23.97 ± 0.07 hours), and cortisone (plasma, solid red line; cerebrospinal fluid, dotted red line). (**B**) Number of genes expressed (average FPKM of the 12 time points >0.1) in at least one of the 64 tissues (25,098, left) or present in all the tissues sampled (UEG, 10,989, right). Different colors indicate the gene types. (**C**) Principal component analysis performed on the 12 time points of 36 representative tissues shows clustering of similar tissues groups based on gene expression profiles. The 22 tissues constituting the brain group clustered very tightly apart from the peripheral organs. (**D**) Cumulative distribution of the average fraction of total transcription contributed by genes when sorted from most to least expressed in each tissue (*x* axis). In all figures, each tissue is color-coded according to its system or functional group presented in Fig. 1.

Basic cellular functions are rhythmic

Because tissues differed in the number of rhythmic genes and shared only a limited number of them (Fig. 3, B and C), even in terms of the circadian clock components (Fig. 5A), it is likely that different components of a given metabolic or signaling pathway were rhythmic in different tissues. To examine this, we used Gene Ontology (GO)–based annotation of rhythmic genes in different tissue types (*16*) and compared the ontology terms that were significantly overrepresented in at least 10 organs. As expected, among the 32 Kyoto Encyclopedia of Genes and Genomes (KEGG) biochemical pathways shared by at least 10 organs, circadian rhythmicity was the most frequently overrepresented. Other highly overrepresented pathways were those implicated in DNA replication, DNA repair, protein ubiquitination, the mTOR (mechanistic target of rapamycin) pathway, lysosomal function, amino-sugar metabolism, pyruvate metabolism, and oxidative phosphorylation (Fig. 6A and fig. S5). Overall, basic energy metabolism pathways and pathways that ensure the integrity of DNA, protein, and other macromolecules showed daily rhythms in multiple tissues. Phase set enrichment analysis (PSEA) (*17*) allowed us to calculate the peak phase of various ontology groups. Although DNA damage repair and lysosomal function groups showed peak expression during the afternoon (Fig. 6B), other functional groups were relatively more dispersed. Genes in the mTOR pathway peaked mostly when animals ate, whereas the protein ubiquitination pathways components were expressed before those of the mTOR pathway and were mainly active between midnight and noon. Annotation along cellular compartments also complemented the pathway-centric annotation. Specifically, cellular transcripts that encode components participating in protein synthesis,

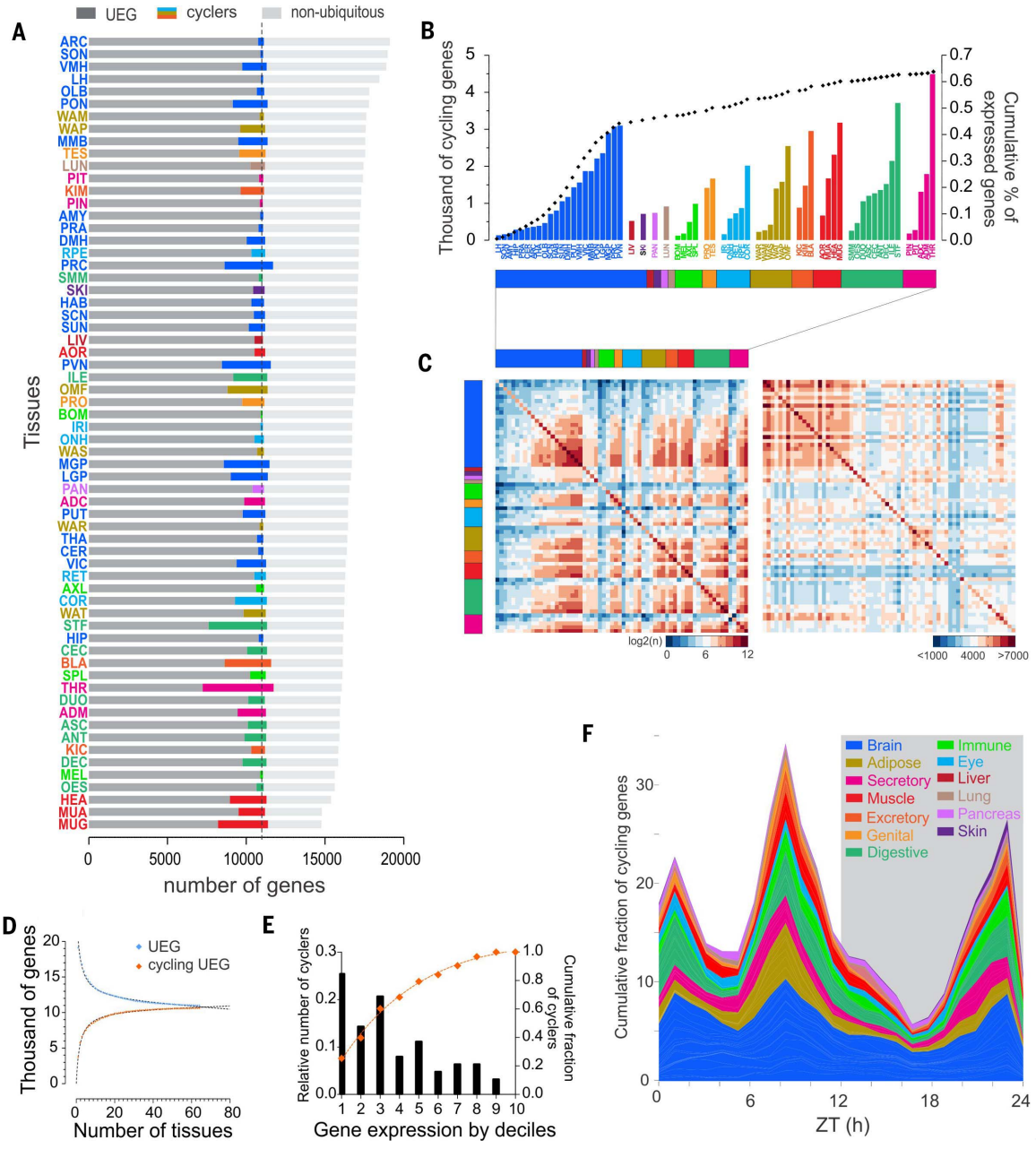


Fig. 3. Rhythmic gene expression across tissues. (**A**) Number of rhythmic genes (colored bars) and their distribution in each tissue between the pool of genes expressed in all tissues (UEGs, dark gray) and the pool of genes expressed more specifically (light gray). (**B**) Rhythmic identity. Shown are the number of cycling genes per tissue (grouped according to type) and their relative contribution to the total pool of genes (dotted black line). (**C**) Tissue by tissue of the overlap of the cycling genes (left) and of expressed genes (right). The color coding represents the degree of the overlap: from blue, few common genes, to red, high number of shared genes [on the left, log2 (common cycling genes), and on the right <1000 to >7000 commonly expressed genes (excluding UEGs)]. Raw data is provided in table S6. (**D**) Number of UEGs and rhythmic UEGs detected as a function of the number of tissues sampled. Best-fit function is overlaid (R^2 = 0.9909 and 0.9944, respectively) (**E**) Weights of rhythmic transcription. Shown is distribution according to deciles of expressed genes (from most expressed to least expressed) indicating where the highest proportion of rhythmic genes are located. The curve indicates the cumulative fraction of cycling genes according to their level of expression. (**F**) Cumulative distribution of the peak phases of gene expression in the different tissues (grouped by systems and functions) throughout the day–night cycle. Peak phases of gene expression were normalized to the maximum in each tissue.

protein degradation, autophagy, and mitochondrial function commonly showed daily variation in abundance (Fig. 6C). Golgi-related GO terms were enriched within the cellular compartments rhythmic in the most number of tissues. In particular, components of the cis side of the Golgi apparatus, such as the endoplasmic reticulum–Golgi intermediate compartment, were the most represented among the most widely regulated (Fig. 6D). By contrast, tissues showing daily rhythms in components of the distal Golgi complexes were relatively few in number (Fig. 6D), which possibly reflects the fact that rhythmic regulation of the secretory pathway is more important for earlier stages of secretory machinery.

Comparative rhythmic transcriptomes between mouse and baboon

The data set allowed comparison between diurnal baboon and nocturnal mouse in the phases of circadian clock components and rhythmic transcriptome in the SCN and other tissues. We selected 11 tissue data sets from rodents (*1*, *18*) for which we obtained comparable data sets in the baboon [including liver (LIV), lungs (LUN), heart (HEA), kidney (KID), adrenal gland (ADR), muscle (MUS), AOR, white adipose tissue (WAT), CER, brainstem (BST), pituitary (PIT), and SCN]. These data sets are from a young male C57BL/6 mouse strain fed a standard laboratory diet ad libitum. Although these data were collected from mice that were released into constant darkness for 1 to 2 days before tissue collection, the phase of expression of clock genes under such conditions deviate <1 hour from that observed under a normal light–dark cycle (*13*). We compared the phase of the core clock components in baboons and mice and found that the peak phase of expression of each clock component was shifted

by ~12 hours between mouse and baboon tissues (with the notable exception of the SCN). This is evident for *Bmal1* and *Per1* (Fig. 7, A and B), which showed peaks of expression in the baboon in evening and morning, respectively, whereas in mice, the peak phases occurred in morning and evening, respectively. The peak phase of expression of other clock components similarly showed opposite phases in most organs. However, the expression of Cry1 was distinct. In the baboon, the median phase for Cry1 expression was during late afternoon, whereas in mice, its expression is delayed by only 7 hours, to peak after midnight (Circadian Time 18). In the SCN, unlike other tissues, the phase of expression of circadian clock components was not systematically opposite to that of their baboon orthologs. The peak phase of *Bmal1* was similar between baboon and mouse

Fig. 4. Tissue-specific rhythmic gene expression. Radial plot of the distribution of the peak phase of expression of the cycling genes in each of the 64 tissues of the present atlas. On the first plot, gray indicates ZT, and number of gene peaks of expression are listed in black.

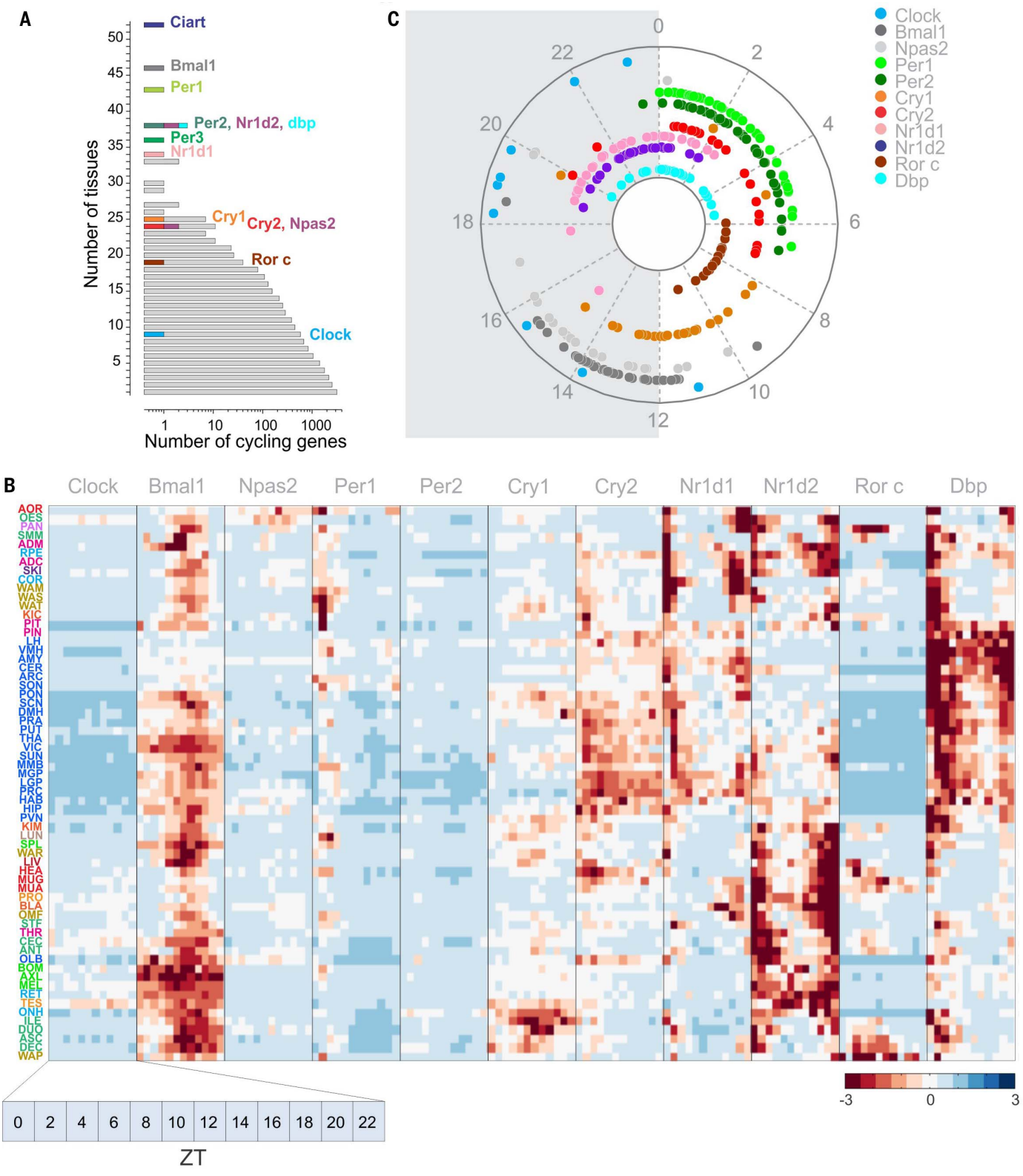


Fig. 5. The molecular clock across tissues. (**A**) Distribution of the cycling genes ranked according to the number of tissues in which they cycle. The most frequent cycling clock component is *Ciart*, cycling in 52 tissues, followed by *Bmal1* and *Per*. The contributions of the core clock components and output (Dbp) to this distribution are highlighted. (**B**) Detailed transcript abundance time course of the molecular clock components in the 64 tissues examined [log2(1 + FPKM)]. (**C**) Distribution of the peak phases of core clock components in the tissues where they are detected as cycling.

SCN—that is, peaking in early evening—whereas *Per1* and *Cry2* phases were 6 and 12 hours apart, respectively (Fig. 7C). This implies a complex diurnal-nocturnal switch downstream of circadian clock components of the SCN.

Because the peak phases of expression of core clock components are ~12 hours apart between mouse and baboon, we tested whether other rhythmic genes in the same tissues shared similar phase differences. Given the differences in sampling (day–night in baboon versus constant dark in mice, timed eating in baboon versus ad lib eating in rodents, and differences in sampling frequency), we anticipated a limited number of common cyclers between baboon and

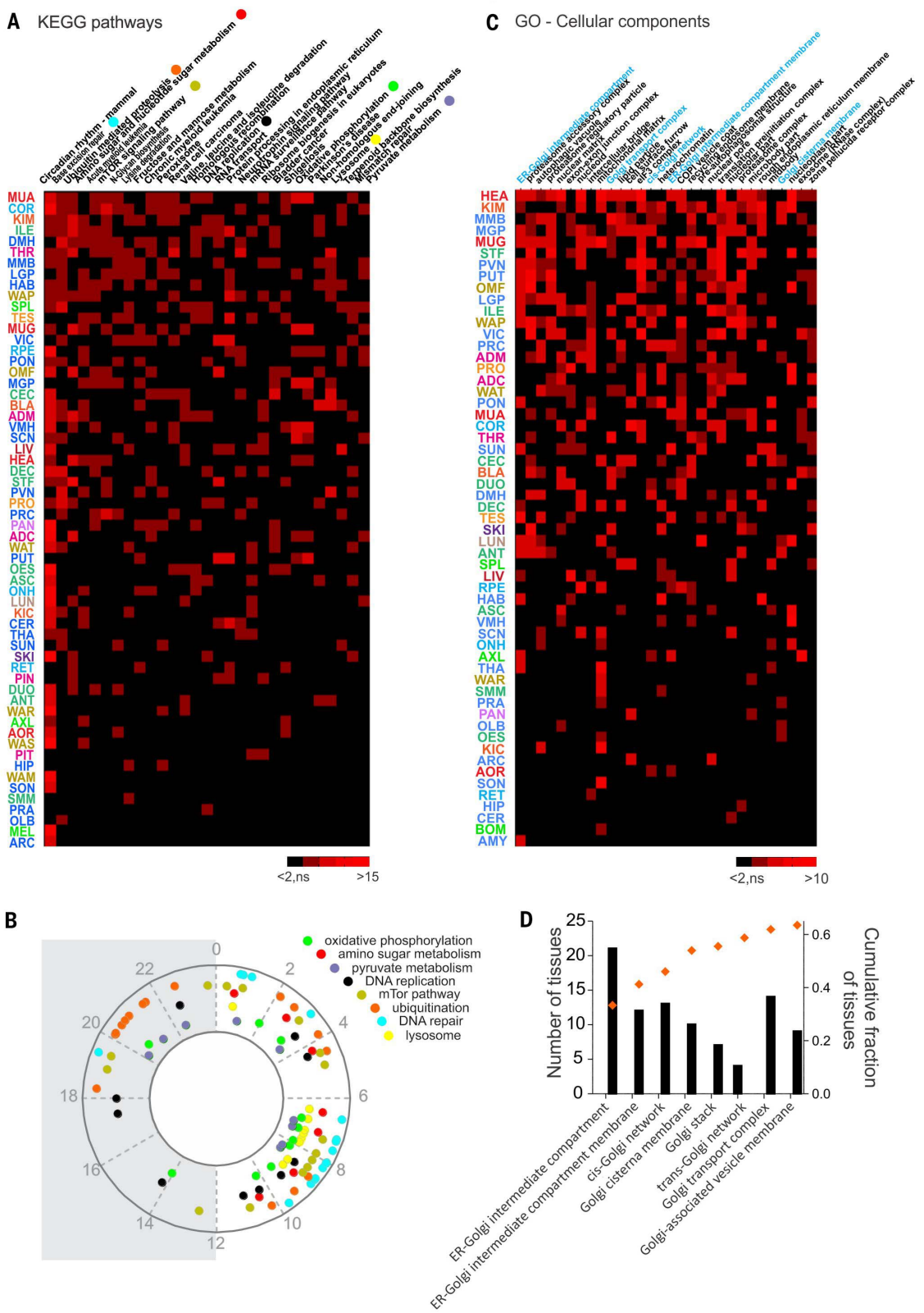


Fig. 6. Rhythmic transcription drives coordinated cellular organization and rhythmic physiology. (**A**) Heatmap of the KEGG pathways enriched for cycling genes in more than 10 tissues. Pathways satisfying the criteria of overrepresentation analysis (ORA) (gene number > 2, *z* score > 2, and permuted $P < 0.05$) are shown in red (red intensity codes for the *z* score). Insignificantly enriched pathways are represented in black. (**B**) Phase distribution over the 24-hour cycle of representative KEGG pathways. Phases were calculated and statistically tested with the PSEA tool. (**C**) Heatmap of the GO cell components terms enriched for cycling genes in more than 10 tissues. Red indicates the pathways satisfying the criteria of overrepresentation analysis as described for (A). (**D**) Histogram of the number of tissues showing overrepresentation for Golgi-related GO-cellular component terms sorted from the cis to the trans side of the Golgi apparatus. Superimposed is the cumulative fraction of the rhythmic regulation for each GO term in the 45 tissues (out of 64) in which at least one of these GO terms that are related to the Golgi apparatus was cycling.

mouse. Indeed, although the number of genes expressed in baboon and mouse tissues showed a positive correlation [coefficient of determination (R^2) = 0.6, P = 0.0145, except for the liver] (fig. S6A), the absolute numbers of cycling genes were not correlated (R^2 = 0.14, P = 0.29), with relatively few common rhythmic genes (Fig. 7, D and E, and fig. S6, B and C). The peak phases of expression of rhythmic genes in corresponding tissues in baboon and mouse showed distinctive features. Specifically, the primate showed a consolidated period of transcription during the day and a transcriptionally quiescent period during the night (when the animals are asleep), whereas there was no such period in rodents (Fig. 7F).

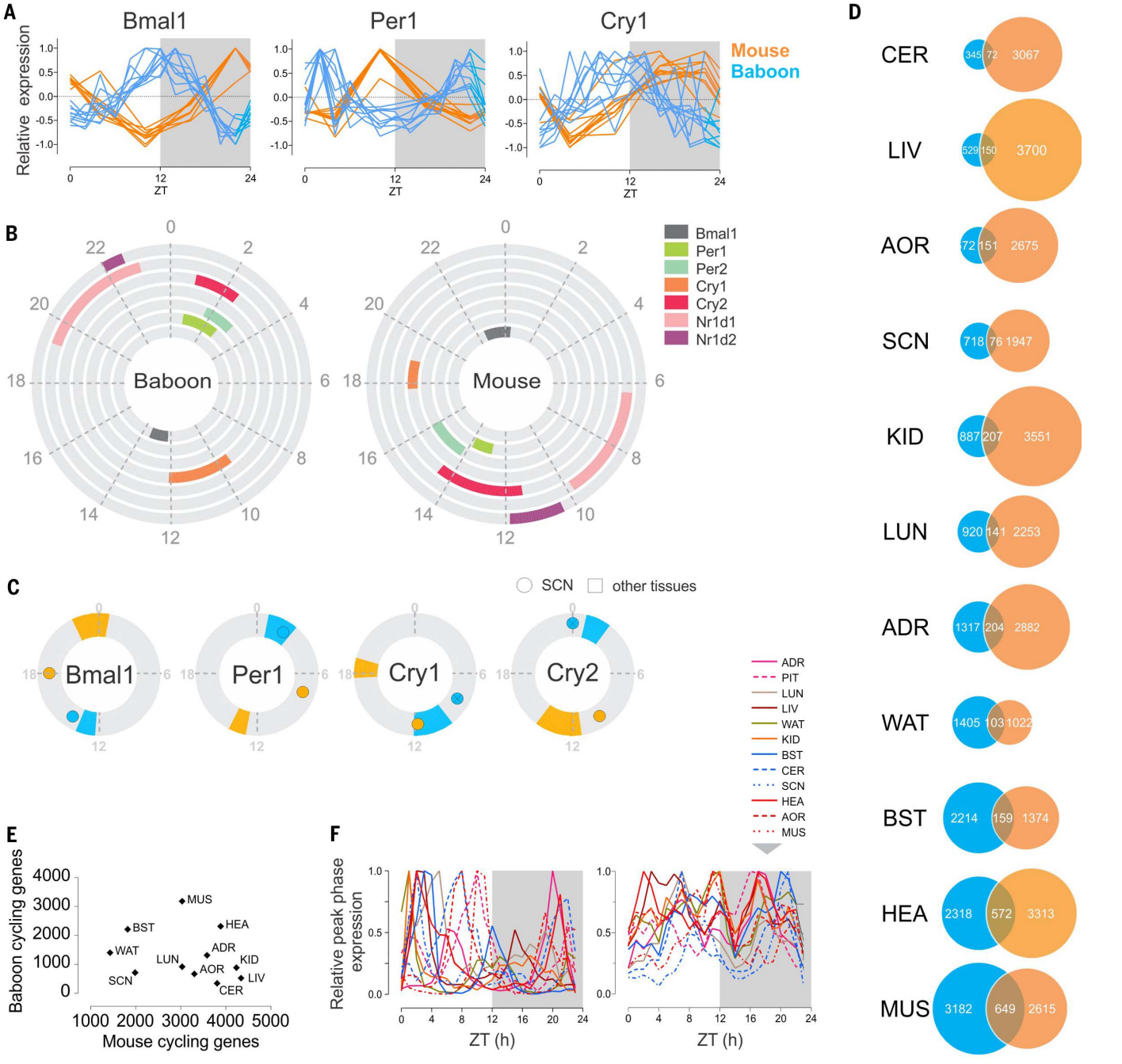


Fig. 7. Comparison of rhythmic transcriptome organization in the diurnal primate and nocturnal rodent. (**A**) Normalized daily expression profiles of *Bmal1*, *Per1*, and *Cry1* across 10 tissues in mouse and baboon (ADR, AOR, BST, CER, HEA, KID, LIV, LUN, MUS, and WAT). (**B**) Phase distribution of the main clock genes in the baboon (left) and the mouse (right). The colored fraction of the rings corresponds to the confidence interval (95%) of the phases of the corresponding clock genes in the tissues in which these genes were detected as cycling. (**C**) Comparison of the phase of *Bmal1*, *Per1*, *Cry1*, and *Cry2* in the SCN (dot) and in the other tissues (95% confidence interval) in mouse (orange) and baboon (blue). Baboon *Cry1* and *Cry2* are not detected as cycling in the SCN; their maximum expression is represented (crossed circle). (**D**) Numbers of cycling genes in selected tissues compared in the baboon and the mouse and overlap of common cycling genes in the two species. (**E**) A tissue-by-tissue comparison showing the lack of correlation of the number of cycling genes in the baboon and in the mouse (R^2 = 0.086, P = 0.38). (**F**) Normalized peak phases of gene expression in 12 tissues throughout the day (left in the baboon, right in the mouse).

Furthermore, among the common cyclers between mouse and baboon in the different tissues, we did not observe a consistent 12-hour difference in the peak phases of expression. This is in contrast to what would be expected from the hypothesis of a simple nocturnal-diurnal "switch" downstream of the SCN. For example, in HEA, 47% of shared rhythmic genes cycled with <6-hour phase difference between mouse and baboon, whereas in other organs, such as CER, 65% of shared genes cycled with a 10- to 15-hour phase difference (fig. S6D). This suggests a very diverse and autonomous temporal organization across peripheral tissues.

Our study provides a comprehensive spatiotemporal gene expression atlas in a diurnal primate, allowing the investigation of both tissue-specific and diurnal expression patterns underlying physiology and behavior. Although it has been assumed that almost every tissue in mammals harbors daily rhythms in gene expression, this study specifies the identities of rhythmic genes in a wide range of tissues covering all major functions. The abundance and complexity of gene expression in a given tissue had little correlation with the number or diversity of rhythmic genes in the same tissue. The temporal patterns of expression add a distinctive dimension to tissue and organ functional specializations that cannot be inferred from the tissue's transcriptome signature alone. The tissue-specific genes are the less likely to be circadianly expressed. Across tissues, the majority of rhythmic genes are among UEGs. At the organismal level, two-thirds of rhythmic genes are from the 10,989 UEGs. Similar to UEGs, the majority of core circadian clock components were detected in nearly all tissues, but only a few out of the 13 component showed rhythmic expression in any given tissue. The phase of expression of any clock component, when rhythmic, was similar across tissues. This would imply that a core circadian clock in a given tissue is composed of a specific combination of different paralogous pairs of activators and repressors, each oscillating at their respective phases. Such tissue-specific organization of the core clock and tissue-specific oscillations of UEGs lends support to the idea that a primordial circadian system, as found in unicellular cyanobacteria, tuned the rhythmic expression of a vast majority of the genome to orchestrate daily rhythms in almost all essential biochemical and cellular processes. With the evolution of metazoans and tissue specialization, the basic set of genes necessary for cellular function is expressed in all tissues, and the circadian clock still continues to oscillate in similar phase, but it recruits a tissue-specific gene repertoire among the UEGs for rhythmic expression.

The daily rhythms expressed in >80% of the PCGs, encoding diverse biochemical and cellular functions, constitutes by far the largest regulatory mechanism that integrates diverse biochemical functions within and across cell types. From a translational point of view, rhythmicity may have a major impact because 82.2% of genes coding for proteins that have been identified as druggable targets by the U.S. Food and Drug Administration (www.drugbank.ca) (*19*) show cyclic changes in transcription in at least one tissue (table S12). These percentages may be underestimations because it is likely that in heterogeneous tissue such as lymph nodes, genes oscillating in only a subset of cell types may not be scored as rhythmic. Because daily rhythms arise from interactions between the circadian clock, light–dark cycle, temperature cycles, and daily eating and sleeping patterns (fig. S7), the data set—which was collected under defined conditions known to contribute to daily rhythms—establishes a reference atlas to understand how genetic and environmental factors may contribute to circadian rhythm disruptions and lead to specific pathologies in humans.

REFERENCES AND NOTES

1. R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, J. B. Hogenesch, A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16219–16224 (2014). doi: 10.1073/pnas.1408886111; pmid: 25349387
2. M. F. Ceriani *et al.*, Genome-wide expression analysis in *Drosophila* reveals genes controlling circadian behavior. *J. Neurosci.* **22**, 9305–9319 (2002). pmid: 12417656
3. S. L. Harmer *et al.*, Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**, 2110–2113 (2000). doi: 10.1126/science.290.5499.2110; pmid: 11118138
4. H. Oster *et al.*, The functional and clinical significance of the 24-h rhythm of circulating glucocorticoids. *Endocr. Rev.* **38**, er20151080 (2016). doi: 10.1210/er.2015-1080
5. W. J. Schwartz, S. M. Reppert, S. M. Eagan, M. C. Moore-Ede, In vivo metabolic activity of the suprachiasmatic nuclei: A comparative study. *Brain Res.* **274**, 184–187 (1983). doi: 10.1016/0006-8993(83)90538-3; pmid: 6684493
6. C. S. Möller-Levet *et al.*, Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1132–E1141 (2013). doi: 10.1073/pnas.1217154110; pmid: 23440187
7. J. Z. Li *et al.*, Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9950–9955 (2013). doi: 10.1073/pnas.1305814110; pmid: 23671070
8. C. Vollmers *et al.*, Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metab.* **16**, 833–845 (2012). doi: 10.1016/j.cmet.2012.11.004; pmid: 23217262
9. M. Melé *et al.*, The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: 10.1126/science.aaa0355; pmid: 25954002
10. L. Pipes *et al.*, The Non-Human Primate Reference Transcriptome Resource (NHPRTR) for comparative functional genomics. *Nucleic Acids Res.* **41** (D1), D906–D914 (2013). doi: 10.1093/nar/gks1268; pmid: 23203872
11. GTEx Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi: 10.1126/science.1262110; pmid: 25954001
12. G. Wu, R. C. Anafi, M. E. Hughes, K. Kornacker, J. B. Hogenesch, MetaCycle: An integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016). doi: 10.1093/bioinformatics/btw405; pmid: 27378304
13. H. R. Ueda *et al.*, A transcription factor response element for gene expression during circadian night. *Nature* **418**, 534–539 (2002). doi: 10.1038/nature00906; pmid: 12152080
14. K. F. Storch *et al.*, Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002). doi: 10.1038/nature744; pmid: 11967526
15. G. Z. Wang *et al.*, Cycling transcriptional networks optimize energy utilization on a genome scale. *Cell Reports* **13**, 1868–1880 (2015). doi: 10.1016/j.celrep.2015.10.043; pmid: 26655902
16. A. C. Zambon *et al.*, GO-Elite: A flexible solution for pathway and ontology over-representation. *Bioinformatics* **28**, 2209–2210 (2012). doi: 10.1093/bioinformatics/bts366; pmid: 22743224
17. R. Zhang, A. A. Podtelezhnikov, J. B. Hogenesch, R. C. Anafi, Discovering biology in periodic data through Phase Set Enrichment Analysis (PSEA). *J. Biol. Rhythms* **31**, 244–257 (2016). doi: 10.1177/0748730416631895; pmid: 26955841
18. M. Hatori *et al.*, Lhx1 maintains synchrony among circadian oscillator neurons of the SCN. *eLife* **3**, e03357 (2014). doi: 10.7554/eLife.03357; pmid: 25035422
19. M. Uhlén *et al.*, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi: 10.1126/science.1260419; pmid: 25613900

ACKNOWLEDGMENTS

The data discussed in this study have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession no. GSE98965. Peripheral tissue samples are available from S.P. under a materials transfer agreement with Salk Institute. This work was supported by the Salk Institute Innovation and U.S. Department of Defense grant W81XWH-15-1-0169 to S.P.; Next Generation Sequencing Core Facility of the Salk Institute, with funding from NIHNational Cancer Institute Cancer Center Support Grant P30 014195; the Chapman Foundation and the Helmsley Charitable Trust grant; and ANR-09-MNPS-040-01 to H.M.C. G.B. was partly supported by Glenn Center for Aging Research, and L.M. was partially supported by Fyssen and Catharina foundations. We thank the veterinarians and animal technicians at the IPR for their assistance and J. A. Cooper for custom designing a baboon brain matrix for sectioning the brain, D. O'Keefe for manuscript editing, and J. Simon for his help with the illustrations.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/eaao0318/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*20*–*25*)
Tables S1 to S12
Databases 1 and 2

6 June 2017; accepted 24 January 2018
Published online 8 February 2018
10.1126/science.aao0318

RESEARCH ARTICLE

HUMAN GENOMICS

Phenotype risk scores identify patients with unrecognized Mendelian disease patterns

Lisa Bastarache,[1] Jacob J. Hughey,[1] Scott Hebbring,[2] Joy Marlo,[1] Wanke Zhao,[3] Wanting T. Ho,[3] Sara L. Van Driest,[4,5] Tracy L. McGregor,[5] Jonathan D. Mosley,[4] Quinn S. Wells,[4,6] Michael Temple,[1] Andrea H. Ramirez,[4] Robert Carroll,[1] Travis Osterman,[1,4] Todd Edwards,[4] Douglas Ruderfer,[4] Digna R. Velez Edwards,[7] Rizwan Hamid,[5] Joy Cogan,[5] Andrew Glazer,[4] Wei-Qi Wei,[1] QiPing Feng,[6] Murray Brilliant,[2] Zhizhuang J. Zhao,[3] Nancy J. Cox,[4] Dan M. Roden,[1,4,6] Joshua C. Denny[1,4]*

Genetic association studies often examine features independently, potentially missing subpopulations with multiple phenotypes that share a single cause. We describe an approach that aggregates phenotypes on the basis of patterns described by Mendelian diseases. We mapped the clinical features of 1204 Mendelian diseases into phenotypes captured from the electronic health record (EHR) and summarized this evidence as phenotype risk scores (PheRSs). In an initial validation, PheRS distinguished cases and controls of five Mendelian diseases. Applying PheRS to 21,701 genotyped individuals uncovered 18 associations between rare variants and phenotypes consistent with Mendelian diseases. In 16 patients, the rare genetic variants were associated with severe outcomes such as organ transplants. PheRS can augment rare-variant interpretation and may identify subsets of patients with distinct genetic causes for common diseases.

Classically, Mendelian diseases are thought to be rare, caused by variants with large effect sizes, and associated with considerable morbidity and mortality. Many are characterized by a range of clinical phenotypes, often affecting multiple organ systems. Several lines of evidence suggest that genes known to cause Mendelian disease also harbor variants that contribute to complex disease (*1*). Studies have identified clinical overlap in patients codiagnosed with Mendelian and complex diseases (*2*), and single-nucleotide polymorphisms found in genome-wide association studies (GWASs) are enriched for Mendelian loci (*3*). A review of evidence from GWASs and whole-exome sequencing studies found an overlap between primary immunodeficiency genes and complex inflammatory diseases (*4*). Collectively, this evidence suggests that variants in Mendelian disease–causing genes may be an underrecognized contributor to complex disease.

Until very recently, the phenotypic effects of rare genetic variants were ascertained primarily through family-based studies of patients with distinctive and often severe phenotypes. Population-level techniques, such as GWASs and phenome-wide association studies (PheWASs) (*5*, *6*), are not easily applied to rare variation because most studies are underpowered. Cohorts large enough to support GWASs of rare variants have only recently been assembled and have demonstrated the potential impact of rare variants on complex traits such as height, finding rare variants with effect sizes much greater than those of common variants (*7*).

Estimating the pathogenicity of rare variants remains a challenge and a barrier to use in the clinical setting (*8*). Many algorithms have been developed to predict variant pathogenicity (*9*–*11*), and consortia such as ClinGen (*12*) are aggregating knowledge to enable expert determinations. Resources such as ExAC (*13*) have helped refine variant interpretation. Some variants previously interpreted as pathogenic are too common in some populations to cause rare, life-threatening disorders (*14*), whereas others thought to be completely penetrant do not always cause disease (*15*). Initial studies suggest that electronic health records (EHRs) linked to genetic data may help drive genomic discovery and define clinical phenotypes associated with rare variants (*16*–*18*).

We have developed an approach that increases the power to detect rare-variant associations by leveraging the phenotypic patterns of Mendelian diseases. By mapping the clinical manifestations of a Mendelian disease to phenotypes extracted from the EHR, we can compute a "phenotype risk score" (PheRS) that expresses the degree to which an individual's symptoms overlap with a Mendelian disease. We defined a PheRS as a weighted aggregation of genetically related phenotypes, analogous to the genetic risk score approach for analyzing multiple variants against a single phenotype. PheRS was validated against clinically diagnosed cases and controls, and a genetic association study of PheRS profiles for 1204 Mendelian conditions identified both known and previously unknown associations with variants in target genes. The approach presents a method for measuring the phenotypic impact of rare variants and for identifying the heretofore underrecognized contribution of Mendelian disease genes to common medical conditions.

Constructing a phenotype risk score

The Online Mendelian Inheritance in Man (OMIM) database provides clinical synopses for thousands of monogenic diseases (*19*) that have been annotated using the Human Phenotype Ontology (HPO) (*20*). We created a map from HPO terms to consolidated billing codes from the EHR called phecodes. Phecodes enable high-throughput ascertainment of EHR phenotypes and have been widely used to replicate known genetic associations and discover new ones (*21*–*23*). By mapping HPO terms to phecodes, we can express "phenotype syndromes" patterned after Mendelian diseases in OMIM in terms of clinical phenotypes that can be rapidly derived from the EHR. The PheRS for a given Mendelian disease is defined as the sum of clinical features observed in a given subject weighted by the log inverse prevalence of the feature.

Validating PheRS

We compared the PheRSs of clinically diagnosed cases with matched controls for six Mendelian diseases. PheRS was a very strong predictor of case status for five of the diseases (Wilcoxon rank sum test; $P = 8 \times 10^{-42}$ to 5×10^{-320}) (Fig. 1, A and B). The exception was phenylketonuria ($P = 0.28$), which effectively served as a negative control because newborn screening and dietary avoidance of phenylalanine essentially eliminates disease manifestations in affected individuals (*24*). The PheRS for each Mendelian disease demonstrated specificity for the target disease: The cases for different Mendelian diseases had similar PheRS distributions to those for controls (Fig. 1C). The lone exception was that the PheRS for hereditary hemochromatosis (HH) was significantly elevated for cystic fibrosis (CF) cases versus controls. However, even in this instance, CF cases had PheRSs that were three times as high for CF compared with HH. A review of controls with a PheRS greater than the 75th percentile identified one individual (PheRS > 99th percentile) who was diagnosed with HH in the 6 months after the case/control ascertainment. Thus, for this individual,

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [2]Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI, USA. [3]Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. [4]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [5]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA. [6]Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA. [7]Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA.
*Corresponding author. Email: josh.denny@vanderbilt.edu

the PheRS suggested the diagnosis before it was made by providers.

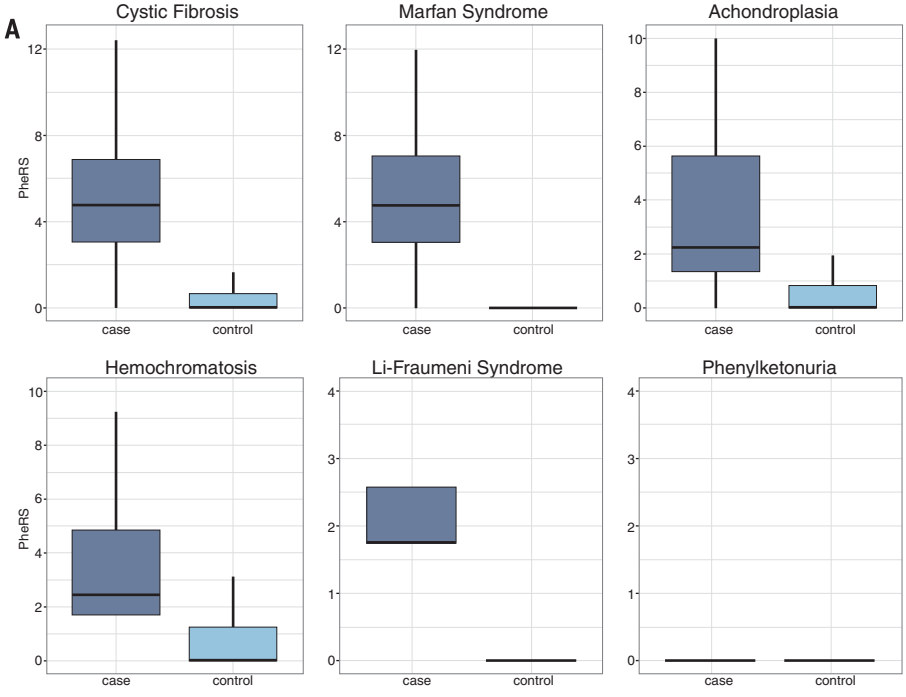
PheRS identifies potentially pathogenic variants in Mendelian disease genes

We conducted an association analysis based on a cohort of 21,701 adults of European ancestry genotyped on the Exome BeadChip (table S1). In this cohort, we computed PheRSs for 1204 Mendelian diseases (1096 causative genes) for which we had sufficient genotype data. We tested for association between PheRSs and 6188 rare variants (minor allele frequency < 1%) using linear regression, assuming a dominant genetic model. We only tested the PheRS for a particular Mendelian disease against variants in the gene or genes known to cause that disease. We found 18 significant associations between rare variants and PheRSs (false discovery rate $q < 0.1$; Table 1). All significant results had a positive beta coefficient, indicating that the variants were associated with an excess of Mendelian disease phenotypes. Four of the genes had an established dominant mode of inheritance, whereas the remaining 13 genes were known as exclusively or primarily recessive. Four were annotated in ClinVar as "pathogenic" or "likely pathogenic," and the Human Gene Mutation Database (HGMD) provided evidence of pathogenicity for an additional three variants (*25*). The phenotypic impact of the remaining nine variants have not, to our knowledge, been previously described.

Clinical chart review revealed that eight of the 807 individuals with statistically significant variants were diagnosed with the target Mendelian disease. Seven individuals with one of the two *CFTR* variants (p.G542* and p.R553*) were diagnosed with CF. Clinical genetic testing confirmed the variants called on the Exome BeadChip, including one homozygote for p.G542*, and established compound heterozygosity with ΔF508 for five others (Fig. 2A and table S2). All individuals diagnosed with CF had a PheRS greater than four standard deviations from expected values. Additionally, the highest-scoring heterozygote for p.E168Q in *HFE* was diagnosed with HH on the basis of clinical findings. The diagnosis of HH was considered but never confirmed for another p.E168Q heterozygote who died of end-stage liver disease.

Although the majority of patients with significant variants were undiagnosed, these individuals had a high burden of severe end points related to the Mendelian diseases. Of the 40 heterozygotes for *HFE* p.E168Q, four had liver transplants (10 versus 1.2% in background; $P = 2.1 \times 10^{-3}$; Fisher's exact test). Individuals with variants in two genes associated with renal failure had elevated rates of kidney transplant: Five of 36 patients (14%) with *AGXT* p.A295T received transplants (another is awaiting transplant), as well as two of 15 patients (13%) with *DGKE* p.W322* (versus 3% in background; $P = 6.9 \times 10^{-3}$ and $P = 0.088$, respectively; Fisher's exact test). Four of 69 *TG* p.G77S heterozygotes underwent thyroidectomies (6 versus 2% of noncarriers; $P = 0.039$; Fisher's exact test). These are end-stage phenotypes, potentially resulting from the effects of these variants, and did not have an increased prevalence in other significant variant carriers (table S3). Additionally, we found that the population attributable fraction for the constituent PheRS phenotypes averaged 0.5%, with a maximum of 4.5%, suggesting that common diseases in adult populations may, in some cases, be attributed to variants in Mendelian genes (fig. S1).

An examination of PheWAS results using Fisher's exact test for variants identified in the discovery analysis revealed that constituent phenotypes were often marginally significant ($P < 0.05$), while not crossing the Bonferroni correction level for a single PheWAS. For *CFTR* p.G542*, three features used in the PheRS for CF achieved marginal significance (bronchiectasis, disease of the pancreas, and chronic airway obstruction) (Fig. 2B). However, the constituent phenotypes for CF were only statistically significant when they were analyzed collectively as a PheRS. The association of p.G542* with the PheRS for CF was similar to the association with the phenotype of CF itself (PheRS by linear regression, $P = 3 \times 10^{-8}$, versus CF diagnosis by Fisher's exact test, $P = 8 \times 10^{-7}$). Similarly, although individuals with variant p.W322* in *DGKE* had an excess of nephrotic syndrome features (Fig. 2C), these phenotypes were not significantly associated on their own in the PheWAS analysis (Fig. 2D). A similar pattern was observed for the remaining variants identified in the discovery analysis (figs. S2 to S17). A PheWAS analysis of all 6188 variants tested in the discovery analysis and 1734 phecodes did not yield any significant ($q < 0.1$) associations.



B

Disease	Abbrev.	Cases	P
Cystic fibrosis	CF	733	5×10^{-220}
Marfan syndrome	MS	244	2×10^{-175}
Achondroplasia	ACH	101	4×10^{-51}
Hereditary hemochromatosis	HH	83	8×10^{-43}
Li-Fraumeni syndrome	LF	21	4×10^{-64}
Phenylketonuria	PKU	133	0.280

C

PheRS definition

Case/control definition	CF	MS	ACH	HH	LF	PKU
CF	4.50	0.00	0.00	1.43	0.00	0.00
MS	0.00	4.39	0.00	0.00	0.00	0.00
ACH	0.00	0.00	2.11	0.00	0.00	0.00
HH	0.00	0.00	0.00	1.70	0.00	0.00
LF	0.00	0.00	0.00	0.00	1.74	0.00
PKU	0.00	0.00	0.00	0.00	0.00	0.00

Fig. 1. PheRSs capture the diagnostic fingerprint of Mendelian disease in EHR data. Scores for six Mendelian diseases were calculated for clinically diagnosed cases and controls matched by age, sex, race, and record length. (**A**) Boxplots of PheRSs for cases and controls for each disease. (**B**) Number of cases and statistical significance between cases and controls (Wilcoxon rank sum test) for each disease. (**C**) Matrix of standardized differences in location (pseudomedian) of the PheRSs between cases and controls (by row) and for each Mendelian disease definition (by column).

Replication of previously unrecognized associations

We attempted to replicate significant associations from the discovery analysis in two independent cohorts: a European ancestry cohort from Marshfield Clinic (n = 9441) and a non-European ancestry cohort from Vanderbilt (n = 3820) (tables S4 and S5). Each was tested as in the discovery cohort, using linear regression assuming a dominant model, adjusted for age and sex. Only variants with at least 10 heterozygotes or homozygotes for the rare allele were tested. In the Marshfield cohort, both attempted associations were replicated: p.G77S in *TG* with thyroid dyshormonogenesis PheRS ($P = 5.0 \times 10^{-4}$) and p.R507H in *FAN1* with karyomegalic interstitial nephritis PheRS ($P = 8.2 \times 10^{-3}$) (table S6). In the Vanderbilt non-European ancestry cohort, we replicated two of three associations: p.A993A in *KIF1A* with spastic paraplegia PheRS ($P = 1.9 \times 10^{-3}$) and p.A295T in *AGXT* with primary hyperoxaluria type 1 PheRS ($P = 3.9 \times 10^{-3}$). The association between p.R507H in *FAN1* and karyomegalic interstitial nephritis PheRS was not replicated in the non-European ancestry cohort, potentially owing to the small number of individuals with the allele in the replication cohort (n = 15).

Sequencing individuals with hitherto unrecognized variants

To test for additional rare variants segregating with high-PheRS individuals, we analyzed the whole-exome sequences of 84 individuals from the discovery analysis for seven of the significant variants (table S7), including individuals with elevated (n = 36) and nonelevated (n = 48) PheRSs. A total of four individuals were found to carry a second rare, nonsynonymous variant in the target gene (Fig. 3 and tables S8 and S9). Two were possible compound heterozygotes (phase could not be determined in this analysis) (*PLCG2* and *AGXT*), and two were homozygotes for the variant identified in the discovery analysis (*DGKE* and *AGXT*, confirming the results from genotyping). Three of the four individuals with confirmed second variants had the highest PheRS for their respective diseases among those selected for whole-exome sequencing.

The heterozygote for *AGXT* p.A295T who was found to have an additional rare *AGXT* variant (p.R381K) through whole-exome sequencing had the highest PheRS for primary hyperoxaluria type 1, a recessive condition characterized by nephrocalcinosis and oxalate nephrolithiasis; an EHR review revealed that he had calcium oxalate crystals evident in urinalysis. The second-highest-scoring individual, a confirmed heterozygote, was diagnosed with hyperoxaluria, which was attributed to his Crohn's disease. The p.A295T homozygotes in the discovery and replication cohorts were no more symptomatic than their heterozygous counterparts. This evidence, along with the persistence of the signal after removing individuals with second variants, suggests that p.A295T may act as a strong risk factor for hyperoxaluria, with more severe manifestations occurring in individuals with additional genetic or environmental risk factors.

Table 1. Significant associations between phenotype risk scores (PheRSs) for Mendelian disease and rare variants. Shown are significant results from the analysis of 7520 PheRS-variant pairs, generated using linear regression assuming a dominant model, adjusted for age and sex. All associations with a false discovery rate of $q < 0.1$ are included. The established mode of inheritance is listed in the "OMIM reported inheritance" column; AD indicates autosomal dominant, AR indicates autosomal recessive; and AR* indicates that disease has also been reported in heterozygotes. ClinVar designations are included, when available: P, pathogenic; LP, likely pathogenic; LB, likely benign; U, uncertain significance. Variants with relevant phenotype associations in HGMD are indicated with a "Y." Results from applying American College of Medical Genetics and Genomics (ACMG) interpretations are given in the last column; an arrow indicates that the interpretation changed in light of evidence presented in this paper. rsIDs are assigned by the dbSNP database. HOM, homozygote; HET, heterozygote.

Gene	Variant	rsID	HOM/ HET	Associated Mendelian Disease	OMIM Reported inheritance	Phenotype categories in PheRS	Beta	P	ClinVar	HGMD	ACMG
CFTR	c.1624G>T p.Gly542Ter	rs113993959	1/27	Cystic fibrosis	AR		1.39	2.9×10^{-8}	P	Y	P
CHRNA4	c.1448G>A p.Arg483Gln	rs55855125	1/21	Nocturnal frontal lobe epilepsy, 1	AD		0.58	9.0×10^{-8}	U		U
DGKE	c.966G>A p.Trp322Ter	rs138924661	1/14	Nephrotic syndrome, type 7	AR		1.31	2.8×10^{-7}	LP	Y	LP→P
SUOX	c.228G>T p.Arg76Ser	rs202085145	0/24	Sulfocysteinuria	AR		0.82	1.7×10^{-6}	U		U→P
CFTR	c.1657C>T p.Arg553Ter	rs74597325	0/12	Cystic fibrosis	AR		1.81	2.1×10^{-6}	P	Y	P
KIF1B	c.2021C>T p.Thr674Ile	rs41274468	0/21	Charcot-Marie-Tooth disease, 2A1	AD		0.79	5.3×10^{-6}			U
VWF	c.5851A>G p.Thr1951Ala	rs144072210	0/21	Von Willebrand disease	AR*		0.53	8.6×10^{-6}		Y	U
KIF1A	c.2676C>T p.Ala993=	rs116297894	1/25	Spastic paraplegia-30	AR		0.84	1.3×10^{-5}	LB		LB→U
F10	c.872G>A p.Arg291Gln	rs149212700	0/15	Factor X deficiency	AR*		0.62	1.9×10^{-5}			U
HFE	c.502G>C p.Glu168Gln	rs146519482	0/40	Hemochromatosis	AR		1.08	4.0×10^{-5}	U	Y	U
TG	c.229G>A p.Gly77Ser	rs142698837	0/69	Thyroid dyshormonogenesis	AR		0.26	6.0×10^{-5}		Y	U→P
SH2B3	c.1183G>A p.Glu395Lys	rs148636776	0/22	Familial erythrocytosis, 1	AD		1.48	6.1×10^{-5}			U→P
SPTBN2	c.7109G>A p.Arg2370His	rs145522851	0/11	Spinocerebellar ataxia	AR*		0.75	9.0×10^{-5}			U→LP
FAN1	c.1520G>A p.Arg507His	rs150393409	0/434	Interstitial nephritis, karyomegalic	AR		0.15	9.9×10^{-5}			LB→U
PANK2	c.1561G>A p.Gly521Arg	rs137852959	0/26	HARP syndrome	AR		0.58	1.1×10^{-4}	P	Y	P
SH2B3	c.1183G>A p.Glu395Lys	rs148636776	0/22	Essential thrombocythemia	AD		0.33	1.4×10^{-4}			U→P
AGXT	c.883G>A p.Ala295Thr	rs13408961	1/35	Primary hyperoxaluria, type I	AR		0.82	1.7×10^{-4}	U/LB		LB→U
PLCG2	c.751A>G p.Ile251Val	rs190840748	0/10	Familial cold autoinflammatory syn. 3	AD		0.70	1.9×10^{-4}			U

Neoplastic
Endocrine/Metabolic/Blood
Nervous/Psychiatric/Sensory
Circulatory/Respiratory
Digestive/Genitourinary
Musculoskeletal/Dermatologic
Other symptoms/Injuries

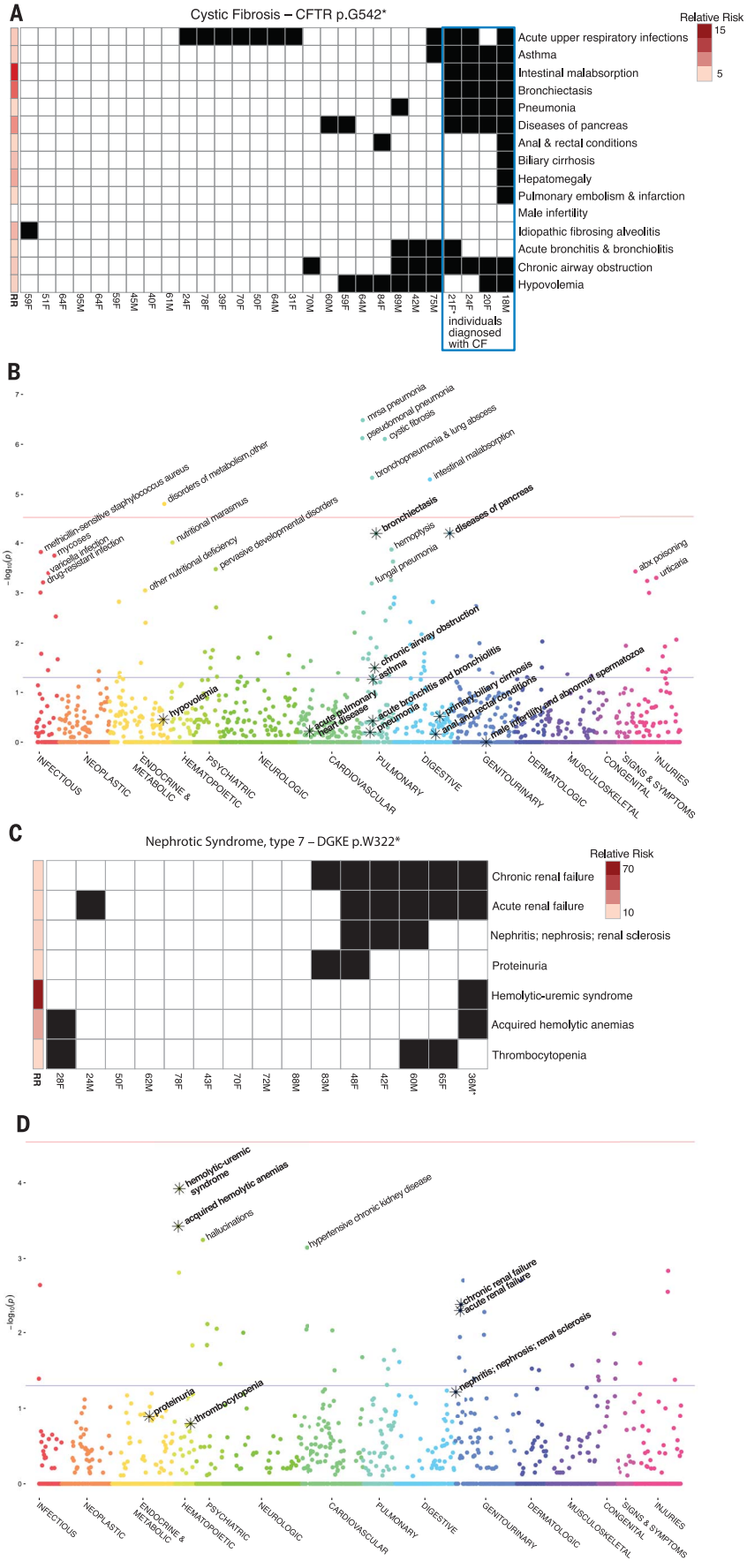


Fig. 2. Phenotypes and PheWAS for two variants associated with PheRS for cystic fibrosis and nephrotic syndrome. For phenotype grids (**A**) and (**C**), each row corresponds to a phenotype used in the PheRS; each column represents an individual who is heterozygous or homozygous (starred) for the variant. The bar on the left of the grid indicates the relative risk (RR) for each phenotype compared with the wild type. In grid (A), individuals clinically diagnosed with cystic fibrosis are enclosed in a blue box. PheWAS plots (**B**) and (**D**) show the PheWAS (Fisher's exact *P* value) for the variant in (A) and (C), respectively. The constituent phenotypes that define the PheRS are starred. All associations with $P < 0.001$ are labeled. The horizontal red and blue lines show the Bonferroni correction threshold for an individual PheWAS and the nominal (uncorrected) $P = 0.05$, respectively. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; E, Glu; G, Gly; H, His; I, Ile; K, Lys; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. An asterisk signifies a termination codon.

The highest scorer for familial cold autoinflammatory syndrome 3 (FCAS3), a dominant condition caused by variants in *PLCG2*, presented in the emergency room with a systemic "urticarial-type rash," for which a cause was never identified, and continued to present with blistering rashes and lip and tongue swelling. Whole-exome sequencing revealed that this patient harbors a second rare variant (p.R687S) in the SH2 domain of *PLCG2*. A nearby variant (p.S707Y), also in the SH2 domain, has been implicated in a related dominant disease that has overlapping features with FCAS3 (OMIM #614878) (*26*).

The confirmed homozygote for p.W322* in *DGKE*, a recessive gene that causes nephrotic syndrome type 7, was diagnosed with hemolytic uremic syndrome as a child and received a kidney transplant in his teens. A genetic etiology for his symptoms was never explored.

Sequencing confirmed the variants called on the Exome BeadChip for *SUOX*, *SH2B3*, *SPTBN2*, and *TG* and did not reveal any additional rare variants in the target genes. For *SH2B3*, the lack of a second variant is consistent with an established dominant inheritance pattern, as well as the high proportion of heterozygotes in our cohort (20 of 22) with at least one feature of familial erythrocytosis. Individuals without a second variant in *AGXT* and *DGKE*, both of which are associated with recessive conditions, also had elevated PheRSs (Fig. 3). This stands in contrast to the heterozygotes for the *CFTR* variants, who did not have a significantly elevated PheRS ($P = 0.51$; linear regression assuming a dominant model, adjusted for age and sex), consistent with a recessive inheritance model. These findings suggest a blurring of the distinction between dominant and recessive labels for some genes.

Sequencing did not reveal any additional rare nonsynonymous variants in the 36 individuals with nonelevated PheRSs. In general, individuals with the highest PheRSs were more likely to be

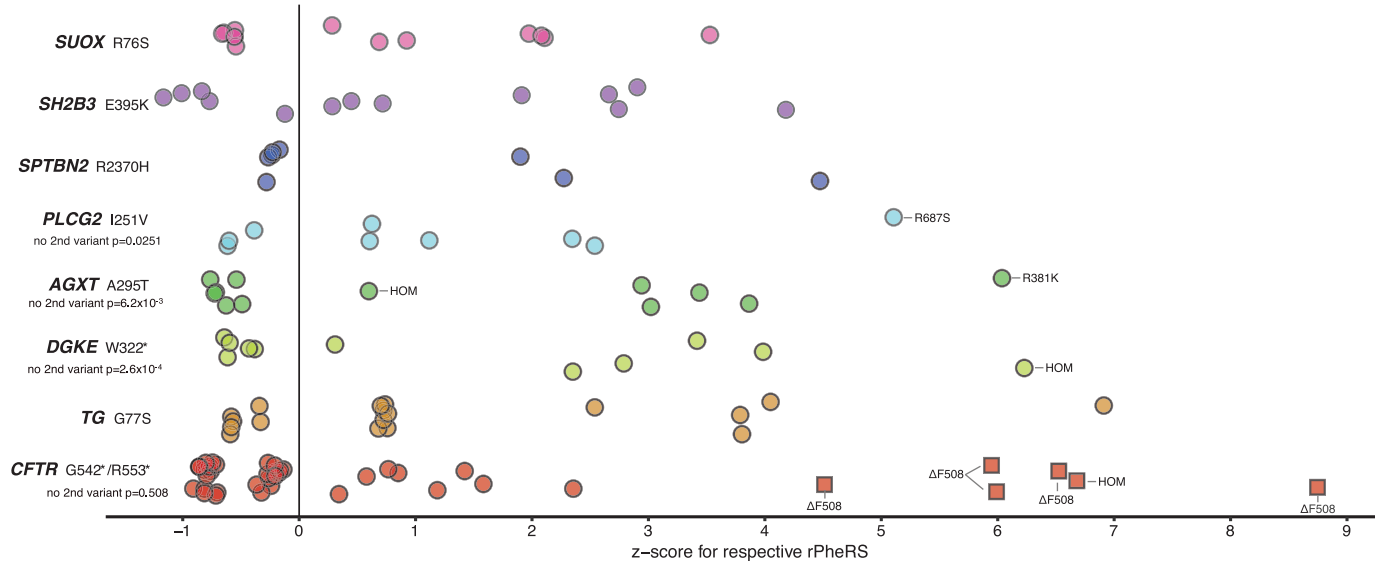


Fig. 3. Whole-exome sequencing reveals second variants among individuals with high PheRSs and demonstrates disease risk in heterozygotes. Each point represents an individual who is heterozygous or homozygous for the variant labeled on the left. The *x* axis represents the *z*-score for the PheRS relative to what is expected given age and sex (using the residual from the PheRS). All individuals carry at least one copy of the variant indicated on the left; additional variants identified by whole-exome sequencing or clinical chart review are labeled for each individual; homozygotes confirmed by sequencing are labeled "HOM." Additional *CFTR* variants were ascertained from clinical testing in the EHR; all other individuals were sequenced for this study. Clinically diagnosed individuals are represented as squares; all others are shown as circles. Where additional variants were found, the association test from the discovery analysis was repeated after dropping individuals with a second variant (*P* values generated using linear regression assuming a dominant model, adjusted for age and sex), and the *P* value is recorded under the gene/variant label.

clinically diagnosed or have additional genetic variants related to their symptoms (fig. S18).

Biologic validation of *SH2B3*, *TG*, and *SUOX* associations

We selected three candidate previously unrecognized associations for biologic validation: *SH2B3*, *SUOX*, and *TG*. SH2B3 is a negative regulator of cytokine signaling in hematopoietic cells that operates through a direct interaction between its SH2 domain and JAK2 to attenuate JAK2-mediated activation of proliferative pathways (*27*). The variant identified in this study, p.E395K, is located in a region of the protein that is critical for its inhibitory function (*28*) and is near known disruptive variants (*29*). Human embryonic kidney (HEK) 293T cells stimulated with erythropoietin showed an increase in phosphorylated extracellular signal–regulated kinase (pERK) levels that was quenched in the presence of wild-type SH2B3 but not quenched with both the known p.R392E functional mutation and our p.E395K variant (Fig. 4, A and B).

Splicing prediction programs suggested a probable reduction in 5′ donor strength for *SUOX* p.R76S and possible generation of an exonic cryptic splice acceptor site by *TG* p.G77S. *SUOX* p.R76S is located at the conserved −1 position of the 5′ donor of exon 5. We demonstrated that the *SUOX* variant caused a decrease in exon inclusion from 96 to 35% (unpaired two-tailed *t* test; $P < 0.001$; Fig. 4C). No transcripts aside from the exon-included and exon-skipped transcripts were detected. Similarly, *TG* p.G77S resulted in altered splicing. The basal rate of exon inclusion was reduced from 65% for the wild-type *TG* exon to only 26% inclusion in the p.G77S exon (unpaired two-tailed *t* test; $P < 0.001$). These ratios were consistent across a range of cDNA concentrations and polymerase chain reaction (PCR) cycle numbers (Fig. 4, C and D).

Comparison of PheRS with existing methods to determine variant pathogenicity

Across all PheRS variant associations with nominal $P < 0.05$ ($n = 454$), functional annotations were significantly correlated with PheRS effect size (Wilcoxon rank sum test); splice donor/acceptor and stop-gain variants tended to have the largest effect size, followed in decreasing order by missense, splice region, synonymous, and intron/untranslated region variants (fig. S19A). Thirteen of 14 functional prediction methods trended or associated with the probability of finding associations with the PheRS; predictions from CADD (combined annotation-dependent depletion) (*9*), SiPhy (*30*), and Polyphen2 HVAR (*10*) were statistically significant ($P < 0.05$; Fisher's exact test; fig. S19B).

Discussion

In our validation study, PheRS was very effective in identifying patients with diagnosed Mendelian disease by using only the phenotypic signatures. Applying PheRS to a genotyped population, we found an increased burden of phenotypes among individuals with rare variants in Mendelian disease genes. Sequencing identified or confirmed second rare variants in four individuals, three of whom had the highest PheRS among all heterozygotes or homozygotes for that variant. In vitro studies provided supporting evidence of pathogenicity for all three variants tested.

Although our approach relies on many decades of accumulated knowledge about the phenotypic imprint of Mendelian disease, the method itself is simple to implement. Our ability to replicate results in an external cohort suggests that it is portable and would therefore be applicable to data sets such as those of the Million Veteran Program, UK Biobank, and the All of Us Research Program (All of Us is a service mark of the U.S. Department of Health and Human Services). Applied to such large populations, this method could facilitate the discovery of pathogenic variants, refine estimates of penetrance across diverse populations, and provide a more nuanced understanding of inheritance patterns, which this study suggests may be more complex than merely "recessive" or "dominant" for some genes. Incorporation of richer EHR data, such as laboratory results and clinical notes (*31*), could increase the resolving power of PheRS. Furthermore, this method may be used with other combinations of phenotypes that do not follow established Mendelian patterns, perhaps based on undiagnosed patients with unusual presentations.

The American College of Medical Genetics and Genomics established guidelines for variant interpretation that reflect the need to combine multiple lines of evidence, including population-based genotype-phenotype correlations (*32*). Our method provides a high-throughput means to generate such evidence. Using these guidelines, 10 of the

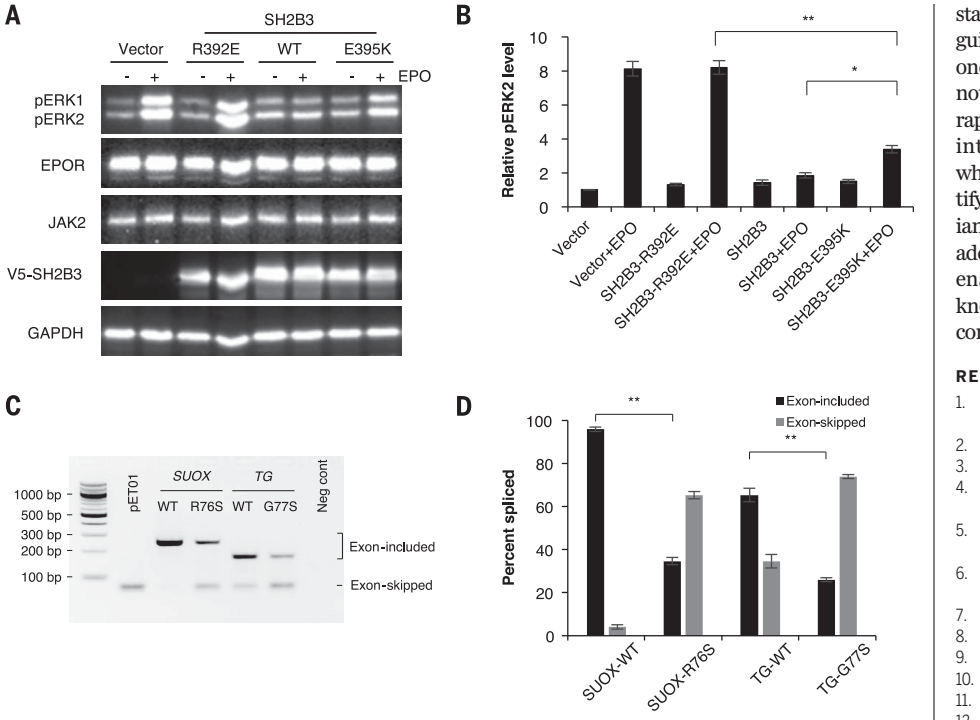


Fig. 4. PheRS enriches for variants with altered function in vitro. Representative Western blots (**A**) and mean pERK2 levels normalized to erythropoietin receptor (EPOR) expression (**B**) in EPO-stimulated HEK293T cells transiently transfected with wild-type (WT) or variant SH2B3 constructs, EPOR, and JAK2. As expected, known functional mutation SH2B3-R392E fails to inhibit EPO-stimulated ERK phosphorylation. Similarly, SH2B3-E395K shows ~1.8-fold elevation (relative to WT SH2B3) of EPO-stimulated ERK activation at 10 min. Reverse transcription PCR analysis (**C**) and quantification (**D**) of WT versus variant splicing of *SUOX* and *TG* exons in HEK293T cells transiently transfected with empty minigene vector pET01, pET01 containing exons of interest flanked by 100 base pairs (bp) of intronic sequence, or negative control pIRES2–EGFP (enhanced green fluorescent protein). Absolute change in the percent of exon inclusion was −61% for the *SUOX* variant and −39% for the *TG* variant. Means ± SEM; $n = 4$ [(A) and (B)] or 5 [(C) and (D)] independent experiments; unpaired two-tailed t test; $^{*}P = 0.003$, $^{**}P < 0.001$.

variants from the discovery analysis were interpreted as having "uncertain significance." By adding data from the PheRS analysis, in combination with evidence from our in vitro studies, four of these variants could be converted to "likely pathogenic" or "pathogenic" (Table 1 and table S10).

Our findings suggest that the phenotypic burden of rare variants in Mendelian genes may be greater than previously thought. A combination of PheRS and sequencing identified symptomatic individuals with genetics consistent with established inheritance patterns—heterozygous individuals for dominant genes (*SH2B3* and *PLCG2*) and individuals with confirmed second variants in recessive genes (*DGKE* and *AGXT*)—none of whom were diagnosed with a genetic condition. A much larger number of individuals were heterozygous for variants in genes with a presumed recessive inheritance and yet still had symptoms consistent with the Mendelian disease pattern. Although we cannot exclude the possible influence of structural or noncoding variants, the evidence suggests that these variants increase risk in heterozygotes. These individuals tend to have disease that is mild compared with the classic presentations but severe relative to the general population. For example, homozygous pathogenic mutations in *TG* are associated with congenital goiter, which often progresses to thyroid carcinoma; our most severely affected heterozygote received a thyroidectomy at age 26 for goiter and thyroid carcinoma.

This work adds to the evidence that Mendelian and complex disease are not dichotomous, but rather exist on a spectrum. As a method that is both high-throughput and sensitive to the vast knowledge already acquired, PheRS is a tool that may help bridge the gap between Mendelian and complex disease. A consequential question is whether the treatments designed for a Mendelian condition could be effective in individuals with nontraditional molecular presentations. Of the 17 diseases represented among those patients with suspected but undiagnosed Mendelian disease, 11 have specific treatments available (table S11), some of which could alter the long-term course of the disease.

The impact that this approach will have on accelerating precision medicine depends on three interrelated challenges. First, we must integrate statistical associations generated with PheRS into guidelines used for variant interpretation. Second, as we collect stronger evidence for the phenotypic effects of rare variants, we must learn to rapidly and effectively integrate that knowledge into clinical care. Third, we must determine whether PheRS can be used to prospectively identify patients whose symptoms are caused by variants in Mendelian genes. If these challenges are addressed, approaches like ours may ultimately enable the conversion of big data not just to knowledge but also to improved care and outcomes for patients.

REFERENCES AND NOTES

1. J. R. Lupski, J. W. Belmont, E. Boerwinkle, R. A. Gibbs, *Cell* **147**, 32–43 (2011).
2. D. R. Blair *et al.*, *Cell* **155**, 70–80 (2013).
3. T. Groza *et al.*, *Am. J. Hum. Genet.* **97**, 111–124 (2015).
4. D. Langlais, N. Fodil, P. Gros, *Annu. Rev. Immunol.* **35**, 1–30 (2017).
5. J. C. Denny, L. Bastarache, D. M. Roden, *Annu. Rev. Genomics Hum. Genet.* **17**, 353–373 (2016).
6. W. S. Bush, M. T. Oetjens, D. C. Crawford, *Nat. Rev. Genet.* **17**, 129–145 (2016).
7. E. Marouli *et al.*, *Nature* **542**, 186–190 (2017).
8. D. G. MacArthur *et al.*, *Nature* **508**, 469–476 (2014).
9. M. Kircher *et al.*, *Nat. Genet.* **46**, 310–315 (2014).
10. I. A. Adzhubei *et al.*, *Nat. Methods* **7**, 248–249 (2010).
11. N. M. Ioannidis *et al.*, *Am. J. Hum. Genet.* **99**, 877–885 (2016).
12. H. L. Rehm *et al.*, *N. Engl. J. Med.* **372**, 2235–2242 (2015).
13. M. Lek *et al.*, *Nature* **536**, 285–291 (2016).
14. A. K. Manrai *et al.*, *N. Engl. J. Med.* **375**, 655–665 (2016).
15. R. Chen *et al.*, *Nat. Biotechnol.* **34**, 531–538 (2016).
16. S. L. Van Driest *et al.*, *JAMA* **315**, 47–57 (2016).
17. I. S. Kohane, *Nat. Rev. Genet.* **12**, 417–428 (2011).
18. D. C. Crawford *et al.*, *Front. Genet.* **5**, 184 (2014).
19. OMIM (Online Mendelian Inheritance in Man), http://omim.org/.
20. S. Köhler *et al.*, *Nucleic Acids Res.* **42**, D966–D974 (2014).
21. J. C. Denny *et al.*, *Nat. Biotechnol.* **31**, 1102–1110 (2013).
22. A. Verma *et al.*, *PLOS ONE* **11**, e0160573 (2016).
23. W.-Q. Wei *et al.*, *PLOS ONE* **12**, e0175508 (2017).
24. E. L. Macleod, D. M. Ney, *Ann. Nestlé [Engl.]* **68**, 58–69 (2010).
25. P. D. Stenson *et al.*, *Hum. Genet.* **136**, 665–677 (2017).
26. Q. Zhou *et al.*, *Am. J. Hum. Genet.* **91**, 713–720 (2012).
27. N. Maslah, B. Cassinat, E. Verger, J.-J. Kiladjian, L. Velazquez, *Leukemia* **31**, 1661–1670 (2017).
28. W. Tong, J. Zhang, H. F. Lodish, *Blood* **105**, 4604–4612 (2005).
29. C. Camps *et al.*, *Haematologica* **101**, 1306–1318 (2016).
30. M. Garber *et al.*, *Bioinformatics* **25**, i54–i62 (2009).
31. S. J. Hebbring *et al.*, *Bioinformatics* **31**, 1981–1987 (2015).
32. S. Richards *et al.*, *Genet. Med.* **17**, 405–423 (2015).

ACKNOWLEDGMENTS

The authors thank B. Li and Q. Wei for technical assistance with the whole-exome sequence variant–calling pipeline. **Funding:** This work was supported by grants R01-LM010685, K22-LM011939, and T15-LM007359 from the National Library of Medicine; U01 grants supporting Vanderbilt's participation in the eMERGE (Electronic Medical Records and Genomics) network (HG004603, HG006378, and HG008672); and grants T32-HG008341 and U01-HG009086 from National Human Genome Research Institute. BioVU received and continues to receive support through the National Center for Research Resources (UL1-RR024975), which is now the National Center for Advancing Translational Sciences (UL1-TR000445). Other support comes from grants R01-GM114128, P50-GM115305, and R01-GM120523 from the National Institute for General Medical Sciences; R01-HL133786 from the National Heart, Lung, and Blood Institute; an American Heart Association career development award (16FTF30130005); and research funds from the University of Oklahoma Health Sciences Center. **Author contributions:** L.B., J.J.H., and J.C.D. conceived of and implemented PheRS. M.B. and S.H. assisted with the replication of primary results. L.B., R.H., T.L.M., J.C., and J.C.D. researched and interpreted rare variants. J.M., J.C., and A.G. designed and conducted the SUOX and TG in vitro studies. W.Z., W.T.H., and Z.J.Z. designed and conducted the SH2B3 in vitro

studies. W.-Q.W., Q.F., D.R., T.E., and D.R.V.E. assisted with designing and analyzing the sequence data. T.E. and J.D.M. provided support for statistical analyses. T.O., A.H.R., J.M., L.B., and J.C.D. reviewed medical records. S.L.V.D., J.D.M., Q.S.W., M.T., R.C., N.J.C., and D.M.R. advised on the interpretation of the results and methods refinement. L.B., J.J.H., D.M.R., and J.C.D. drafted the manuscript. All authors provided critical feedback and helped shape the analysis and interpretation of the results of this study. **Competing interests:** S.L.V.D. received an honorarium from Merck as an invited speaker. After completing her work on this project, T.L.M. began working for Alnylam Pharmaceuticals and received stock in the company. Neither Merck nor Alnylam supported or were involved in this research. The remaining coauthors have no competing interests to report. **Data and materials availability:** Software for PheWAS analysis and visualization is available at https://github.com/PheWAS/PheWAS. Software to generate PheRSs and produce grid visualizations is available at https://github.com/labastar/PheRS. Genetic and phenotypic data used in this study are available at dbGap under accession number phs001516.v1.p1.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1233/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S17
References (*33*–*44*)

14 November 2016; resubmitted 25 August 2017
Accepted 22 January 2018
10.1126/science.aal4043

REPORT

POLYMERS

Random heteropolymers preserve protein function in foreign environments

Brian Panganiban,[1] Baofu Qiao,[2] Tao Jiang,[1] Christopher DelRe,[1] Mona M. Obadia,[3] Trung Dac Nguyen,[2] Anton A. A. Smith,[1] Aaron Hall,[1] Izaac Sit,[1] Marquise G. Crosby,[4] Patrick B. Dennis,[4] Eric Drockenmuller,[3] Monica Olvera de la Cruz,[2,5] Ting Xu[1,6,7,8]*

The successful incorporation of active proteins into synthetic polymers could lead to a new class of materials with functions found only in living systems. However, proteins rarely function under the conditions suitable for polymer processing. On the basis of an analysis of trends in protein sequences and characteristic chemical patterns on protein surfaces, we designed four-monomer random heteropolymers to mimic intrinsically disordered proteins for protein solubilization and stabilization in non-native environments. The heteropolymers, with optimized composition and statistical monomer distribution, enable cell-free synthesis of membrane proteins with proper protein folding for transport and enzyme-containing plastics for toxin bioremediation. Controlling the statistical monomer distribution in a heteropolymer, rather than the specific monomer sequence, affords a new strategy to interface with biological systems for protein-based biomaterials.

Nature's building blocks, such as proteins and biomachinery, have many features unmatched by synthetic counterparts, including chemical diversity, hierarchical structure, programmed system dynamics, and efficient energy conversion. Despite years of effort to stabilize proteins outside of their native environments, there has been limited development in interfacing biological and synthetic components without compromising their structures and inherent functions. Nonaqueous enzymology via reverse micelles can only maintain a fraction (<20%) of native activity (*1*); amphipols solubilize membrane proteins, but not for water-soluble proteins in organic solvents (*2*); polymer conjugation relies on accessibility of protein-functional groups (*3*); and sol-gel confinement limits protein accessibility and postintegration (*4*).

A chaperone-like polymeric shell outside of a protein may effectively improve protein solubility and stability in organic solvents by providing a barrier to resist both organic solvent exposure and protein conformational change. To form such a nanoscopic polymeric shell, protein-polymer interactions need to be strong enough to favor adsorption yet soft enough not to outcompete the forces governing protein folding. The structure and activity of natural building blocks are governed by multiple noncovalent interactions that are subject to change with small perturbations (*5*). When a homopolymer is conjugated to a protein surface, monomer–amino acid interactions can affect protein folding (*6*, *7*) and deform the polymer chain conformation (*8*). In nature, intrinsically disordered proteins adopt local chain conformations, often achieved by multiple weak binding sites, and mediate various processes (*9*, *10*). Amphiphilic heteropolymers mimicking disordered proteins may offer a versatile approach for protein solubilization and stabilization given, the diversity and complexity of proteins.

Protein surfaces are chemically diverse and heterogeneous (Fig. 1A). Surface analysis of folded water-soluble proteins based on hydrophobicity or charge shows characteristic patch size distributions. The typical patch size is 1 to 2 nm in diameter, with an interpatch distance of 1 to 2 nm (Fig. 1B and fig. S1). Sequence analysis of water-soluble proteins, performed by assigning amino acids as either hydrophilic or hydrophobic, using glycine as the reference (*11*), shows that the block length of amino acids with similar hydrophobicity tends to be less than 10 (fig. S2). Matching the statistical chemical pattern has been shown to be critical in de novo protein design (*12*, *13*) and in modulating polymer/surface interactions (*14*). Rather than synthesizing sequence-specific polymers (*15*), synthetic heteropolymers with similar chemical features and spatial distributions of side chains to match the surface pattern of natural proteins may behave like disordered proteins.

Recent developments in reversible deactivation radical polymerization make it feasible to synthesize random heteropolymers with reliable control over the statistical monomer distribution (*16*–*18*). Four methacrylate-based monomers are selected to impart chemical diversity in heteropolymers and to optimize short-range polymer-protein interactions (Fig. 1C and fig. S3): methyl methacrylate (MMA), oligo(ethylene glycol) methacrylate [OEGMA; number-average molecular weight (M_n) = 500 Da], 2-ethylhexyl methacrylate (2-EHMA), and 3-sulfopropyl methacrylate potassium salt (3-SPMA). MMA is chosen to tailor the overall hydrophobicity for protein solubilization, to anchor the chain at the polar-nonpolar interface, and to reduce the entropic penalty associated with adjusting its local conformation. OEGMA is chosen to leverage the well-known ability of poly(ethylene glycol) (PEG) to stabilize proteins. 2-EHMA and 3-SPMA are chosen to interact with the hydrophobic and positively charged residues on the protein surface, respectively. The selection of the monomer ratio is guided by the calculated solubility parameters (*19*, *20*) and through experimental screenings to achieve polymer dispersion in both aqueous and organic media with the best retention in enzyme activity. Our best-performing heteropolymer, called "RHP," has a compositional ratio of 5(MMA):2.5(OEGMA):2(2-EHMA):0.5(3-SPMA). Using the well-established Mayo-Lewis equation, we estimated the statistical monomer distribution along the RHP chains (*21*, *22*). The histogram of monomer block size confirms the absence of long blocks of the same monomer that could interfere with native protein structure (figs. S5 to S7). The RHP used has a M_n of ~30 kDa and a dispersity of 1.3. At least 12 batches of RHP were synthesized with excellent reproducibility (fig. S11).

Computational studies were performed on mixtures of a common enzyme, horseradish peroxide (HRP), and RHP in both an aqueous buffer solution and toluene in order to elucidate the polymer-protein adsorption mechanism. Explicit solvent all-atom molecular dynamics (MD) simulations were performed by using the CHARMM 36 potential (supplementary materials, section S3). The RHPs are built with compositions and degrees of polymerization close to the corresponding experimental values. The last simulation snapshots of the RHP/HRP mixture in toluene and in water are shown in Fig. 2, A and B, respectively. In water, RHP and HRP are loosely complexed, and only ~40% of HRP surface is covered by RHP, which is in line with the experimental observations that RHP is soluble in water. However, the HRP surface is fully covered by RHPs in toluene; the complex is stable, and no protein structure change was observed over the simulation duration of 0.6 μs. The radial distribution of RHPs around the protein's center of mass was calculated in toluene (Fig. 2C) and in water (fig. S17E). In toluene, the hydrophilic monomers of the polymer (OEGMA and 3-SPMA) were found to

[1]Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. [3]Université Lyon, Université Lyon 1, CNRS, Ingénierie des Matériaux Polymères, UMR 5223, F-69003 Lyon, France. [4]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433, USA. [5]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [6]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA. [7]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [8]Tsinghua-Berkeley Shenzhen Institute, University of California, Berkeley, Berkeley, CA 94720, USA.
*Corresponding author. Email: tingxu@berkeley.edu

point inward, adjacent to the protein, whereas the hydrophobic monomers (MMA and 2-EHMA) were located on the outside, in contact with the toluene. By contrast, the favored orientations of the RHP side chains become weaker in water. These results suggest that once positioned close to the protein surface, the RHPs can adjust their conformations to maximize protein-polymer interactions, which correlates well with our hypothesis. Correlations between the protein surface groups and their nearest monomer neighbors are computed by categorizing Ala, Val, Ile, Leu, Met, Phe, Tyr, Trp, Gly, and Pro amino acids as hydrophobic and the others as hydrophilic. Approximately 70% of the protein surface is covered by the hydrophilic monomers; 50 ± 4% originates from hydrophilic monomer–hydrophilic amino acid interactions, contributing –800 ± 300 kJ/mol to the energy (Fig. 2D). This shows the importance of the local protein surface-polymer interactions in stabilizing the protein structure. The shell formed by the polymer around the HRP core in organic solvents improves protein solubility and stability by providing a barrier to resist both organic solvent exposure and protein conformational change.

The role of the RHPs' composition on their ability to encapsulate the protein in water and toluene was investigated by using a coarse-grained (CG) model based on the data gathered from all-atom MD simulations and experiments (Fig. 2E and supplementary materials, section S4). A representative snapshot of the RHP encapsulating the protein is given in Fig. 2F. The surface coverage increases as the attraction strength between the adsorbing monomers and protein-attractive sites, ε_{Hh}, increases (Fig. 2G) for all solvent selectivity conditions (captured by the attraction strength between the polymer solvophobic beads, ε_{hh}). For sufficiently high ε_{hh}, there exists an optimal value of the fraction of adsorbed monomers (ϕ_A) that maximizes the surface coverage (Fig. 2H). The spatial correlations between the adsorbing monomers and the protein-attractive sites (fig. S22A) demonstrate that the RHPs tend to adopt energetically favored conformations once in contact with the protein to compensate for the entropic penalty associated with confinement to the surface (*23*).

We experimentally tested our hypothesis that RHP is capable of interacting with proteins and mediating their interactions with the local environment. We performed cell-free synthesis of membrane proteins that eliminates potential interference or assistance from the host cell physiology for protein folding and lipid insertion during translation (*24*). Once the plasmids for a model transmembrane protein—oligopeptide/proton symporter PepTso or water channel aquaporin Z (AqpZ)—are added, the translation and folding status of the membrane protein is scored according to the fluorescence intensity of a C terminus–fused green fluorescent protein (GFP). Western blotting analysis (Fig. 3A) using antibodies to GFP confirmed PepTso-GFP expression regardless of the presence of RHP. Without RHP, no GFP fluorescence is detected, suggesting that protein folding is not correct to gain GFP fluorescence (*25*). Control experiments using amphipol (A8-*35*), which is effective in stabilizing membrane proteins after protein expression, show little GFP fluorescence for AqpZ-GFP (*2*). However, there is a ~15-fold increase in GFP fluorescence when PepTso-GFP or AqpZ-GFP is expressed in the presence of RHP at a RHP:ribosome mole ratio of 50:1 (Fig. 3B). The resultant GFP fluorescence is more than an order of magnitude higher than that of commonly used liposome. To further verify the protein folding, the cell-free synthesized PepTso was reconstituted in liposomes and tested in a pyranine-based proton transport assay (Fig. 3C) (*26*). Proton transport was detected by adding valinomycin to the outer solution containing a dipeptide, Ala-Ala (Fig. 3D). The cell-free synthesized PepTso, in the presence of RHP, can fold properly to retain transport function in a lipid environment. Thus, RHP is able to chaperone proper protein folding and solubilize membrane proteins in aqueous solution but will not outcompete the protein-lipid interaction, interfere with lipid insertion, or compromise transport function.

We investigated protein dispersion and stabilization in organic solvents, a requirement to interface proteins with synthetic building blocks toward protein-based materials. Previous nonaqueous enzymology studies of HRP observed only less than 1% of its native activity in hydrophobic solvents that can extract the heme cofactor (*27*). For all studies, proteins and RHPs were first cosolubilized in deionized water, lyophilized, and subsequently resuspended in solvents.

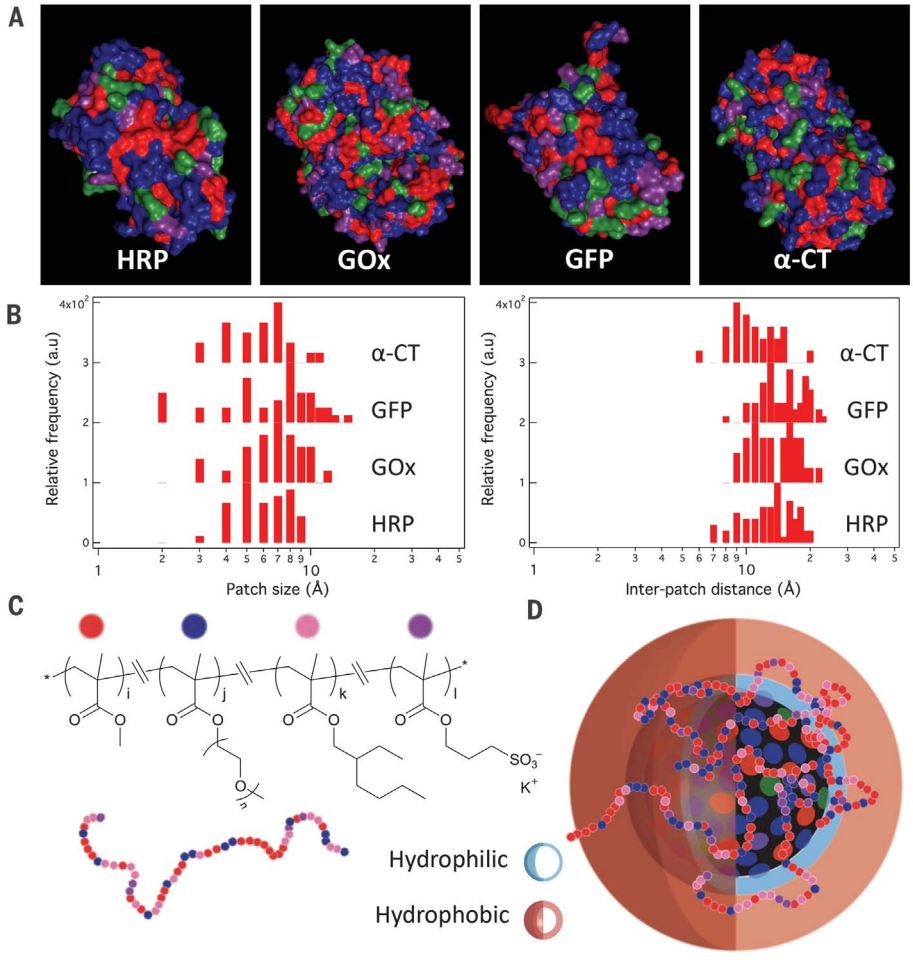


Fig. 1. Design of random heteropolymers based on protein surface pattern for protein solubilization and stabilization in organic solvents. (**A**) As seen in the color-coded proteins—HRP(1H55), GOx(1CF3), GFP(2HRW), and α-CT(1YPH)—the protein surfaces are chemically heterogeneous, with characteristic patch size and interpatch distance. Neutral hydrophilic, blue; hydrophobic, red; positively charged, green; negatively charged, purple. (**B**) The histograms of the patch diameter and the interpatch distance show that the length centers around 1 to 2 nm. Analysis of hydrophobic patches is shown. Analysis of negatively and positively charged patches is shown in fig. S1. (**C**) Designed random heteropolymer with statistical distribution of monomers with varied hydrophobicity matching that of the protein. (**D**) The RHP can coassemble with protein (schematically shown as a patchy particle) in organic media in which the random heteropolymer adjusts its local conformation to maximize protein-heteropolymer interaction without denaturing the protein's local structure.

The RHP/HRP complexes were readily soluble in common solvents for material synthesis and processing such as toluene and chloroform. Transmission electron microscopy (TEM) results (Fig. 4A) show that RHP/HRP complexes form nanoparticles with a diameter of ~50 to 60 nm. Fourier transform-infrared spectroscopy (FTIR) spectra were collected after dissolving RHP/HRP complexes in toluene over 24 hours. The amide I band and its corresponding negative second derivative revealed minimal change in the HRP's secondary structure in toluene (Fig. 4B). There is little change (<2 nm) in the peak position in the ultraviolet (UV)–visible spectra of the heme cofactor in HRP over 24 hours (Fig. 4C),

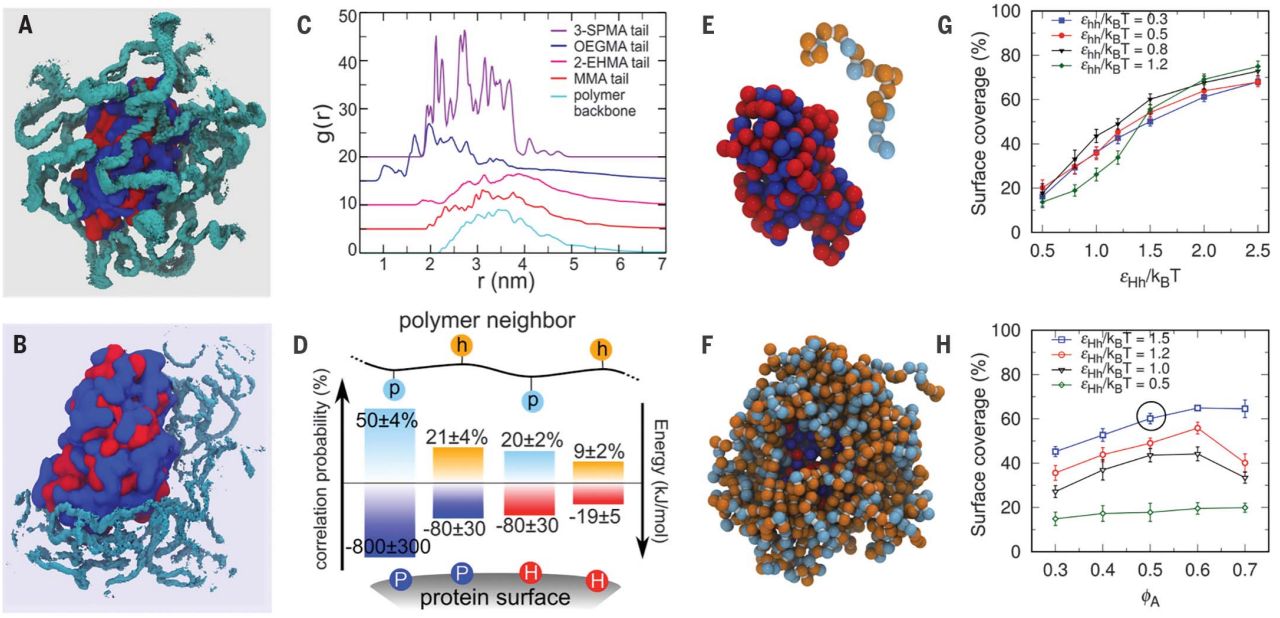
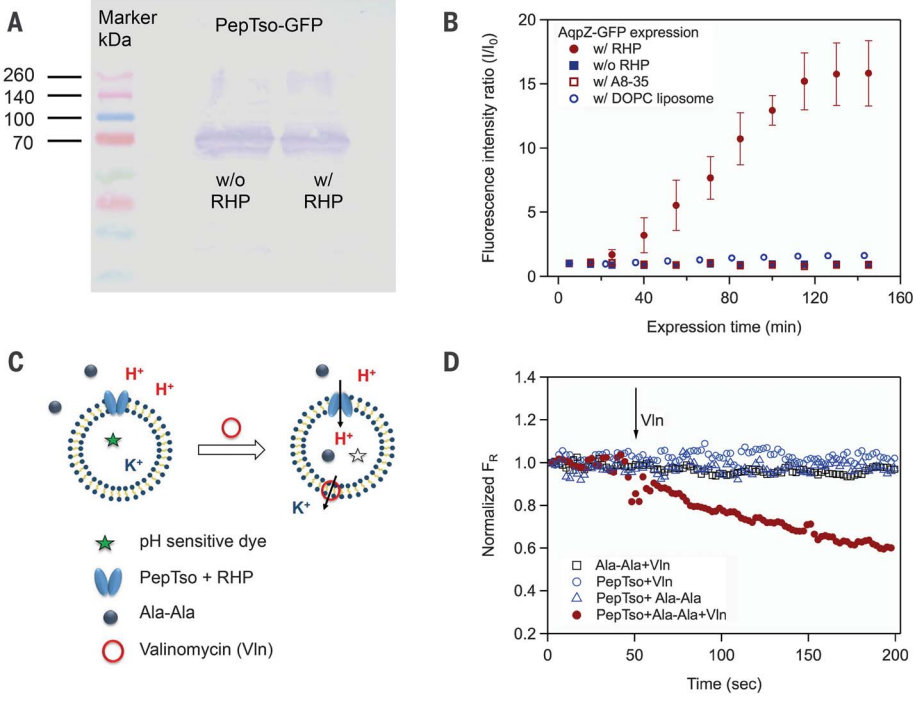


Fig. 2. All-atom and CG molecular dynamics simulation results. (**A** and **B**) Three-dimensional spatial distribution of polymer backbone (cyan) around HRP in (A) toluene and (B) water. The isosurfaces are plotted for densities 50 times the average. The hydrophilic and hydrophobic protein surfaces are colored in blue and red, respectively. (**C**) Radial density profiles of the polymer backbone and the tail atoms of the four monomers (fig. S12) around the center of mass of HRP backbone in toluene. (**D**) Correlation probability (in percent) and energy (kilojoule per mole) between the hydrophilic (P, blue) and hydrophobic (H, red) protein surface and the hydrophilic (p, light blue) and hydrophobic (h, orange) polymer nearest neighbors in toluene. The uncertainties refer to the standard deviations. (**E**) CG models of HRP and a RHP chain, where the red beads represent the protein-attractive sites and the orange beads represent the adsorbing monomers. (**F**) A snapshot of the RHP chains covering the HRP. (**G**) The surface coverage as a function of the adsorption strength. (**H**) The surface coverage as a function of the MMA composition. An example of the system corresponding to the circled data point is illustrated in (F).

Fig. 3. Random heteropolymers can effectively mediate interaction between membrane protein and aqueous solution and assist protein folding during translation in cell-free synthesis. (**A**) Western blot analysis of protein expression with and without RHP in a cell-free synthesis assay. The left lane shows color-coded protein mass standards. (**B**) The level of proper protein folding of GFP-fused AqpZ is significantly enhanced with the presence of RHP (n = 3 independent measurements) in comparison with other surfactants, such as amphipol A8-35 and dioleoylphosphatidylcholine (DOPC) unilamellar liposome. (**C**) Schematics of assay to evaluate PepTso transport properties upon insertion into liposomes. (**D**) Cell-free synthesized PepTso in the presence of RHP can fold and inset into lipids properly, retaining transport function across the membrane.



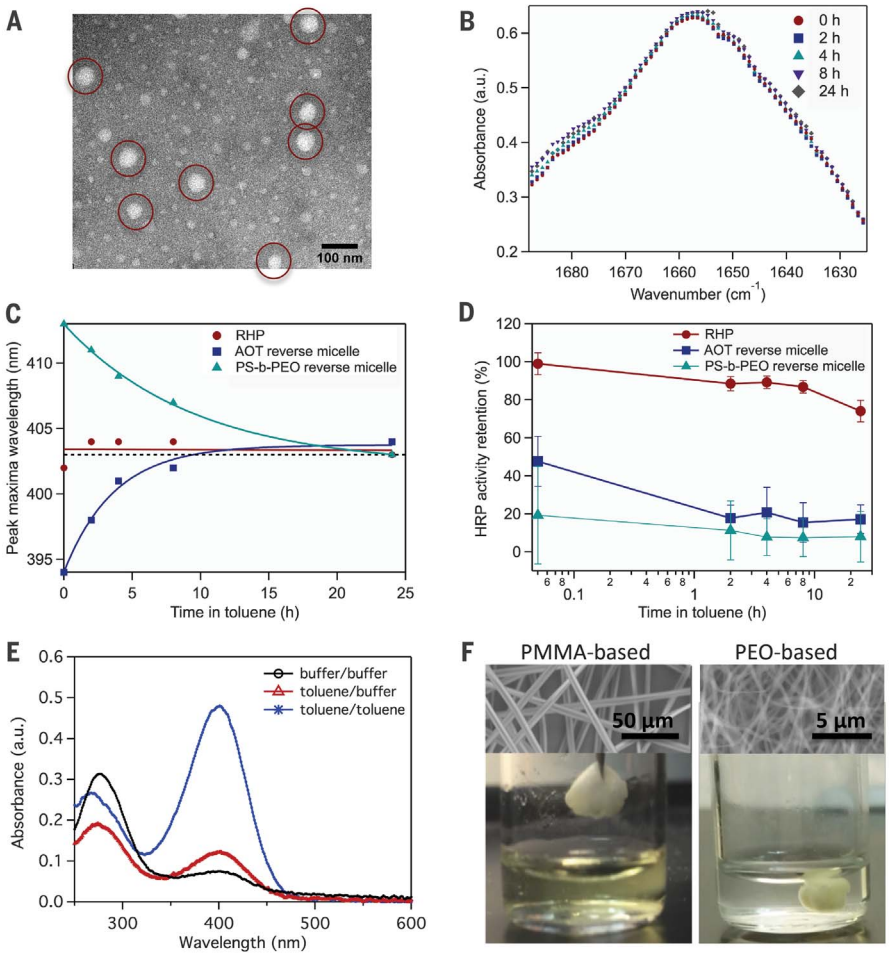


Fig. 4. Random heteropolymers preserve enzymes in organic solvents and enable processing of protein-containing functional materials. (**A**) TEM image of RHP/HRP dried from toluene solution. The RHP/HRP mixture forms ~50- to 60-nm-diameter nanoparticles (red circles). (**B**) FTIR spectra of the RHP/HRP complex in toluene over 24 hours, suggesting minimal changes in the protein secondary structure. (**C**) The peak position of the UV-visible spectrum of the heme cofactor in RHP/HRP complex in toluene over 24 hours, confirming the integrity of its protein tertiary structure and enzymatic pocket. (**D**) In the presence of RHP, ~80% of native HRP activity is retained after being stored in toluene over 24 hours (n = 3 independent measurements). Results from HRP encapsulated in reverse micelles based on sodium bis(2-ethylhexyl) sulfosuccinate (AOT) (<25%) or polystyrene-block-poly(ethylene oxide) (PS-*b*-PEO) polymeric surfactant (<18%) are shown for comparison. (**E**) Enzymatic activity of OPH as a function of storage media for OPH and substrate. Activity seven times higher than control experiments in buffer can be obtained when OPH and the pesticide methyl parathion are cosolubilized in toluene. (**F**) Electrospun PEO- or PMMA-based fiber mats containing RHP/OPH complex can perform on-demand bioremediation. The hydrolysis by-products are trapped in the RHP/OPH/PEO fiber mat for easy removal.

confirming the retention of HRP tertiary structure and stability of the heme-binding pocket in toluene, which is critical for enzymatic activity. To evaluate the retention of the HRP activity after suspension in toluene, aliquots of RHP/HRP toluene solution were dispersed in aqueous solution in order to perform a colorimetric assay. With the presence of RHP, ~80% of HRP native activity was maintained after 24 hours suspended in toluene (Fig. 4D). This is several orders of magnitude higher than HRP alone in nonaqueous media and at least four times higher than the best reported value when using molecular and/or polymeric surfactants (*1*).

We explored the versatility and universality of RHP by extending the study to other proteins. PEGylation is effective for protein dispersion and stabilization but is not effective for proteins with a small number of functionalizable groups, such as glucose oxidase (GOx) (*28*). (fig. S24). The PEGylation is sufficient for GOx dispersion, but the proteins retain less than 10% of native activity after 2 hours (fig. S24). By contrast, RHP/GOx shows ~50% of native activity after 24 hours of storage in toluene. Proteins with different structures, such as β-barrel GFP, are also tested. After RHP/GFP complex is solubilized in organic solvents, there is no detectable shift in fluorescent emission peak maxima and minimal decrease in the emitted fluorescence over 24 hours (fig. S25). These results confirm that the environment in the β-barrel interior remains the same and suggest that there is minimal toluene penetration into the protein core.

The RHP-enabled protein solubilization and stabilization can lead to technologically relevant protein-based materials. Organophosphorus hydrolase (OPH) was chosen for its excellent ability to degrade organophosphates, which are commonly used as insecticides and chemical warfare agents (*29*). However, OPH becomes inactive even in its partially folded dimeric state (*30*), and organophosphates typically have poor solubility in aqueous solution. There is a need to retain OPH activity while cosolubilized with organophosphates in order to realize on-demand bioremediation for these acute toxins. We evaluated OPH activity using a 10 mM preparation of the well-known pesticide methyl parathion (MP). RHP/OPH can be readily solubilized in toluene and chloroform and resuspended in buffer solution. RHP/OPH retains 80 ± 5.6% (n = 3 independent measurements) of the initial OPH activity after 24 hours' suspension in toluene. When methyl parathion was codissolved with RHP/OPH in toluene, dried, and assayed, the OPH activity was more than seven times higher than that of pure OPH in aqueous solution (Fig. 4E). This is attributed to the higher substrate concentration in toluene and RHP's ability to stabilize OPH in both aqueous and organic media. The successful dispersion and stabilization of RHP/OPH complexes enables coprocessing of OPH and synthetic polymers (Fig. 4F). Fiber mats based on either polyethylene oxide (PEO) or PMMA were prepared by means of electrospinning and tested for bioremediation. Both fiber mats were active and degraded MP, weighing approximately 1/10 of the total fiber mat, in a few minutes. In particular, when the RHP/OPH/PEO fiber mat was soaked in MP-containing toluene solution and assayed in buffer, the hydrolysis by-products could be trapped for easy removal. This enables in situ toxin remediation without preprocessing, transfer, or contact with the agent before and after detoxification.

Our studies confirm that random heteropolymers designed on the basis of statistical monomer distribution are effective at maintaining protein function in foreign environments. Using only one design of random heteropolymer, we demonstrated successful cell-free synthesis of membrane proteins with proper folding and retention of protein activity in organic solvents for a wide range of proteins. These hybrid materials will not only take advantage of the precision and efficiency of natural building blocks but also enable reactions and processes on demand where common chemistry necessities are unavailable.

REFERENCES AND NOTES

1. A. M. Klibanov, *Nature* **409**, 241–246 (2001).
2. J.-L. Popot *et al.*, *Cell. Mol. Life Sci.* **60**, 1559–1574 (2003).
3. E. M. Pelegri-O'Day, H. D. Maynard, *Acc. Chem. Res.* **49**, 1777–1785 (2016).

4. R. B. Bhatia, C. J. Brinker, A. K. Gupta, A. K. Singh, *Chem. Mater.* **12**, 2434–2441 (2000).
5. K. A. Dill, *Biochemistry* **29**, 7133–7155 (1990).
6. J. Y. Shu *et al*., *Biomacromolecules* **11**, 1443–1452 (2010).
7. N. Dube, A. D. Presley, J. Y. Shu, T. Xu, *Macromol. Rapid Commun.* **32**, 344–353 (2011).
8. R. Lund, J. Shu, T. Xu, *Macromolecules* **46**, 1625–1632 (2013).
9. H. J. Dyson, P. E. Wright, *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
10. Z. Liu, Y. Huang, *Protein Sci.* **23**, 539–550 (2014).
11. O. D. Monera, T. J. Sereda, N. E. Zhou, C. M. Kay, R. S. Hodges, *J. Pept. Sci.* **1**, 319–329 (1995).
12. W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri, A. Lombardi, *Annu. Rev. Biochem.* **68**, 779–819 (1999).
13. D. A. Moffet, M. H. Hecht, *Chem. Rev.* **101**, 3191–3203 (2001).
14. P. Mansky, Y. Liu, E. Huang, T. P. Russell, C. J. Hawker, *Science* **275**, 1458–1460 (1997).
15. J. F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **341**, 1238149 (2013).
16. J. Chiefari *et al*., *Macromolecules* **31**, 5559–5562 (1998).
17. G. Moad, E. Rizzardo, S. H. Thang, *Acc. Chem. Res.* **41**, 1133–1142 (2008).
18. A. E. Smith, X. Xu, C. L. McCormick, *Prog. Polym. Sci.* **35**, 45–93 (2010).
19. A. F. M. Barton, *Handbook of Polymer-Liquid Interaction Parameters and Solubility Parameters* (Taylor & Francis, 1990).
20. J. N. Israelachvili, *Intermolecular and Surface Forces* (Elsevier Science, 2015).
21. B. G. Manders, W. Smulders, A. M. Aerdts, A. M. vanHerk, *Macromolecules* **30**, 322–323 (1997).
22. F. R. Mayo, F. M. Lewis, *J. Am. Chem. Soc.* **66**, 1594–1601 (1944).
23. T. Ge, M. Rubinstein, *Macromolecules* **48**, 3788–3801 (2015).
24. D. Schwarz *et al*., *Nat. Protoc.* **2**, 2945–2957 (2007).
25. A. Müller-Lucks, S. Bock, B. Wu, E. Beitz, *PLOS ONE* **7**, e42186 (2012).
26. J. L. Parker, J. A. Mindell, S. Newstead, *eLife* **3**, e04273 (2014).
27. A. M. Klibanov, *Trends Biotechnol.* **15**, 97–101 (1997).
28. A. D. Presley, J. J. Chang, T. Xu, *Soft Matter* **7**, 172–179 (2011).
29. B. K. Singh, A. Walker, *FEMS Microbiol. Rev.* **30**, 428–471 (2006).
30. J. K. Grimsley, J. M. Scholtz, C. N. Pace, J. R. Wild, *Biochemistry* **36**, 14366–14374 (1997).
31. Data are available from the Dryad Digital Repository; https://doi.org/10.5061/dryad.j1pt126.

ACKNOWLEDGMENTS

We thank A. Presley for PEGylated GOx studies and V. Wu for assistance with coding of heteropolymers sequence simulation. **Funding:** The work was supported by the U.S. Department of Defense (DOD), Army Research Office, under contract W911NF-13-1-0232. B.Q., T.D.N., and M.O.C. thank the support of the Department of Energy award DE-FG02-08ER46539 and the Sherman Fairchild Foundation. E.D. acknowledges support from the Auvergne-Rhône-Alpes region. A.A.A.S. was supported by The Danish Council for Independent Research Award DFF–5054-00215. T.X. acknowledges support from Dreyfus foundation via Camille-Dreyfus Teacher-Scholar Award. E.D. and T.X. acknowledge support from the Fulbright foundation. M.O.C. and T.X. acknowledge support from the Miller Institute. C.D. is supported by the DOD through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. **Author contributions:** T.X. conceived the idea and guided the project. B.P. performed studies using RHP with HRP, CT, GOx, and GFP. T.J. performed cell-free synthesis of membrane proteins and evaluated their activity. C.D. performed studies using OPH and fabricated and tested the OPH-containing fabrics. B.Q. performed all-atom simulation studies. T.D.N. performed CG simulation studies. B.Q., T.D.N., and M.O.C. analyzed simulation data. E.D. and M.M.O. screened different compositions of RHPs to afford the synthesis and detailed characterization of the first positive hit. A.A.A.S. and A.H. analyzed the RHP monomer distribution, quantified composition drift in RHP, and synthesized different batches of RHPs to verify reproducibility. P.B.D. and M.G.C. provided OPH. All authors participated in writing the manuscript. All authors have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials. For reproduction purposes, raw data used to generate figures are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.j1pt126) (*31*). Random heteropolymers are available from T.X. under a materials transfer agreement with the University of California, Berkeley. **Competing interests:** T.X., B.P., and T.J. are inventors on a patent application submitted by the University of California, Berkeley, that covers random heteropolymers to preserve proteins in foreign environments.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1239/suppl/DC1
Materials and Methods
Figs. S1 to S28
Tables S1 to S4
References (*32*–*52*)
Movies S1 to S2

7 June 2017; accepted 25 January 2018
10.1126/science.aao0335

SURFACE CHEMISTRY

Real-time imaging of adatom-promoted graphene growth on nickel

Laerte L. Patera,[1,2]* Federico Bianchini,[1,3] Cristina Africh,[2]† Carlo Dri,[1,2]‡ German Soldano,[4] Marcelo M. Mariscal,[4] Maria Peressi,[1,5]† Giovanni Comelli[1,2]

Single adatoms are expected to participate in many processes occurring at solid surfaces, such as the growth of graphene on metals. We demonstrate, both experimentally and theoretically, the catalytic role played by single metal adatoms during the technologically relevant process of graphene growth on nickel (Ni). The catalytic action of individual Ni atoms at the edges of a growing graphene flake was directly captured by scanning tunneling microscopy imaging at the millisecond time scale, while force field molecular dynamics and density functional theory calculations rationalize the experimental observations. Our results unveil the mechanism governing the activity of a single-atom catalyst at work.

Reduced dimensionality in the catalyst structure often leads to increased chemical reactivity (*1*, *2*). More specifically, atomic steps of metal surfaces were experimentally identified as the active sites for several heterogeneous catalytic reactions (*3*, *4*), and this identification was theoretically rationalized in terms of a characteristic upshift of the d-band center (*5*). Such undercoordinated sites play also a crucial role in governing the stability of the active phase (*4*). Single atoms, representing the lowest possible coordination, can exhibit even higher specificity and efficiency compared to stepped surfaces (*1*). For example, isolated single metal atoms anchored onto oxide nanocrystallites promote CO oxidation (*6*) and the water–gas-shift reaction (*7*). Recent studies also revealed the peculiar activity of single-atom alloys in boosting the selective hydrogenation of alkenes and hydrocarbons (*8*).

Mobile adatoms, commonly present on metal surfaces at elevated temperatures, are involved in several chemical processes occurring on surfaces and at the solid-liquid interface (*9*, *10*). However, the lack of direct experimental observation of the dynamics of these peculiar processes at the atomic scale, generally owing to the difficulty of achieving simultaneous spatial and temporal resolution, limits the description and understanding of the surface reaction pathways. For the specific case of the chemical vapor deposition (CVD) growth of graphene on transition metals, a catalytic role of single metal atoms was predicted by theorists, but the detailed growth mechanism at the atomic scale remains unclear (*11*).

Nickel (Ni) is one of the most used substrates to synthesize graphitic carbon single layers (*12*). We report the characterization of graphene growth on Ni(111) in real time by means of in situ high-speed scanning tunneling microscopy (STM) measurements. By monitoring the layer formation at the atomic scale and with a time resolution down to milliseconds, we observed at the kink sites of the graphene edges single Ni atoms that are directly involved in the growth process. Simulations of possible reaction paths based

[1]Department of Physics, University of Trieste, Via Valerio 2, I-34127 Trieste, Italy. [2]Istituto Officina dei Materiali–Consiglio Nazionale delle Ricerche (IOM-CNR) Laboratorio TASC, S.S. 14 km 163.5 in AREA Science Park, Basovizza, I-34149 Trieste, Italy. [3]Centre for Materials Science and Nanotechnology (SMN), Department of Chemistry, University of Oslo, Post Office Box 1126, N-0318 Oslo, Norway. [4]Instituto de Investigaciones en Físico-Química de Córdoba, Consejo Nacional de Investigaciones Científicas y Técnicas (INFIQC CONICET) and Departamento de Química Teorica y Computacional, Facultad de Ciencias Quimicas, Universidad Nacional de Córdoba, Cdad. Universitaria, 5000 Córdoba, Argentina. [5]Istituto Officina dei Materiali–Consiglio Nazionale delle Ricerche (IOM-CNR) DEMOCRITOS National Simulation Center, via Bonomea 265, I-34136 Trieste, Italy. *Present address: Institute of Experimental and Applied Physics, University of Regensburg, D-93053 Regensburg, Germany. **†Corresponding author. Email: africh@iom.cnr.it (C.A.); peressi@ts.infn.it (M.P.)** ‡Present address: ELETTRA-Sincrotrone Trieste, S.S. 14 km 163.5, Basovizza, I-34149 Trieste, Italy.

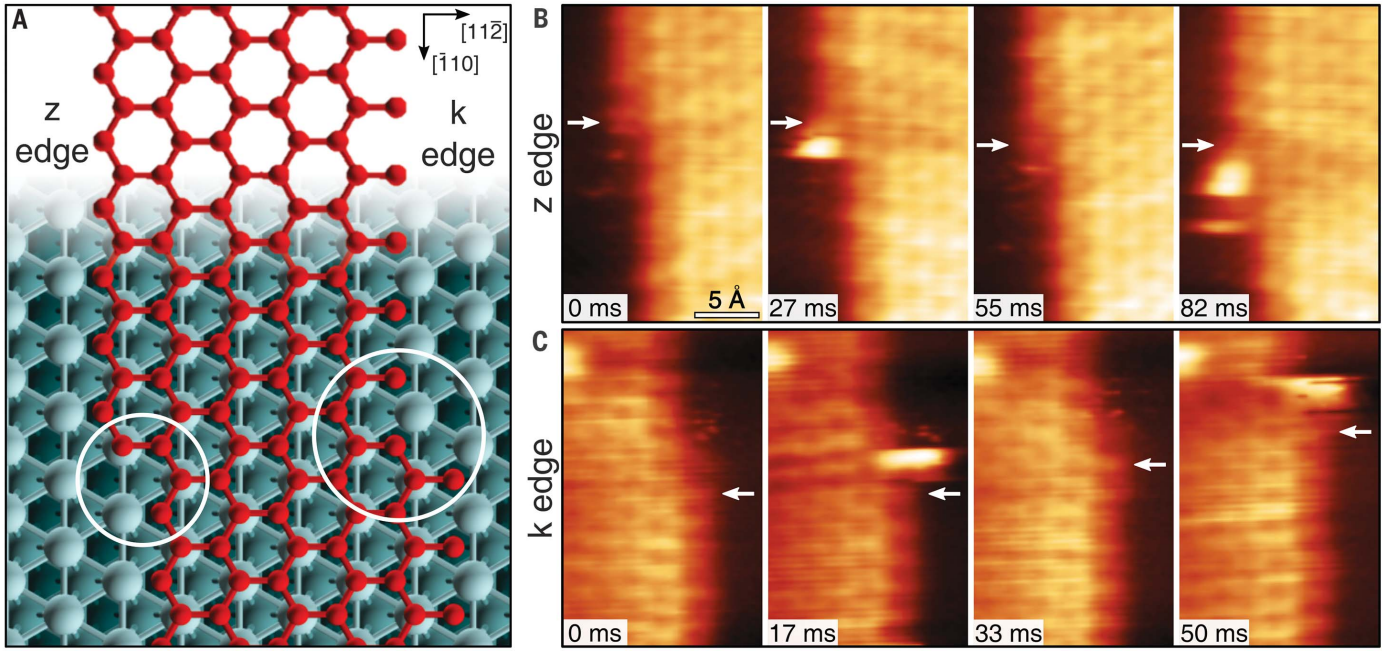


Fig. 1. Graphene growth along z and k edges. (**A**) Zigzag (z) and Klein (k) edges of a top-fcc EG layer on Ni (111). At both edges, the kink structures are highlighted by circles. (**B**) High-speed STM sequence acquired at 710 K in quasi-constant height mode at the z edge, from movie S1 (V = 20 mV; I = 8 nA; frame rate = 36.5 Hz]. White arrows indicate the position of C atoms in fcc-hollow sites near the kink. (**C**) Same as in (B) but for the k edge, from movie S2 (V = 20 mV; I = 7 nA; frame rate = 60 Hz).

on density functional theory (DFT) revealed that Ni adatoms drive the carbon (C) atom addition mechanism by substantially lowering the corresponding reaction barriers.

Islands of monolayer epitaxial graphene (EG) were synthesized on Ni(111) by subsurface C segregation at temperatures between 700 and 740 K (*13*). Three coexisting stable chemisorbed configurations of EG were observed (*14*), with the top/face-centered cubic geometry (top-fcc) as the most abundant phase. During growth, the island edges evolved too rapidly to be resolved by standard STM imaging (typical image acquisition rates of ≈0.02 frames/s). Nevertheless, in a previous study, we succeeded in identifying the top-fcc edge structure by performing high-speed STM measurements (*15*), exploiting the FAST module developed in our laboratory (*16*, *17*). The module allows for image acquisition rates up to 100 frames/s with commercial microscopes already optimized for structural and reactivity studies. The edge terminations of EG islands on Ni(111) during growth at temperatures higher than 570 K (*15*) were identified as the zigzag (z) and Klein (k) structures shown in Fig. 1A. For both edges, kink sites obtained by interrupting the last two C rows were predicted to play an important role in graphene edge expansion (*18*, *19*).

To identify the underlying growth mechanisms during the C-attachment process at both edges, we acquired STM movies with high spatial and temporal resolution and long enough for a statistically relevant analysis.

Series of consecutive images along z [movie S1 (*20*)] and k (movie S2) edges were acquired with frame rates of 36.5 and 60 Hz, respectively. The use of different frame rates (from 4 to 10 Hz) and tunneling current setpoints (from 500 pA to 10 nA) did not yield substantial differences, indicating a negligible effect of the fast scanning tip on the observed processes. This consistency presumably resulted from the short tip-surface interaction time (typically ≈1.5 to 10 μs/ pixel) and from the thermal excitation that at elevated temperatures likely overwhelmed tip-induced effects (*21*).

Movie S1 initially shows the formation at the z edge of two kink sites, acting as nucleation centers for graphene growth (fig. S1). In the image sequence in Fig. 1B, growth proceeds from the kink site in the middle by orderly completion of the two interrupted carbon rows while moving the kink ahead. During the z edge growth, the observed structure of the kink remains equivalent to the initial one, probably because of the preference for fcc-hollow site termination (*15*). A similar evolution was predicted by previous DFT calculations for the z edge of EG on Cu(111), where the growth would start from the hexagon forming the kink, and continuous incorporation of C atoms would form a new z chain on the graphene growth front (*19*). The image sequence in Fig. 1C, extracted from movie S2, shows the growth mechanism at the k edge, similar to the one described above for the z edge, i.e., proceeding from the kink site and parallel to the edge. Our DFT calculations (fig. S2) support this observation, showing that the incorporation of new C atoms takes place at the kink and that, after addition of a first C atom in a top site, the energetics favor the addition of a second C atom in a fcc-hollow site.

A closer examination of both sequences in Fig. 1, B and C, reveals the presence of bright objects at the kink sites. These features appeared and disappeared: They were present in some frames (at 27 and 82 ms in Fig. 1B, and at 17 and 50 ms in Fig. 1C) and were absent in others (at 0 and 55 ms in Fig. 1B, and at 0 and 33 ms in Fig. 1C). They were typically imaged for few scanlines only, indicating a short residence time of the originating species (at the millisecond time scale; see supplementary text), and their appearance did not depend on scanning parameters and tip conditions. Regarding the nature of the imaged species, one possibility would be that they are mobile C clusters, as described for graphene growth on Ru(0001) and Rh(111) (*22*, *23*). However, for Rh(111), where STM images have been acquired, the clusters have a height similar to that of the graphene layer, at variance with what

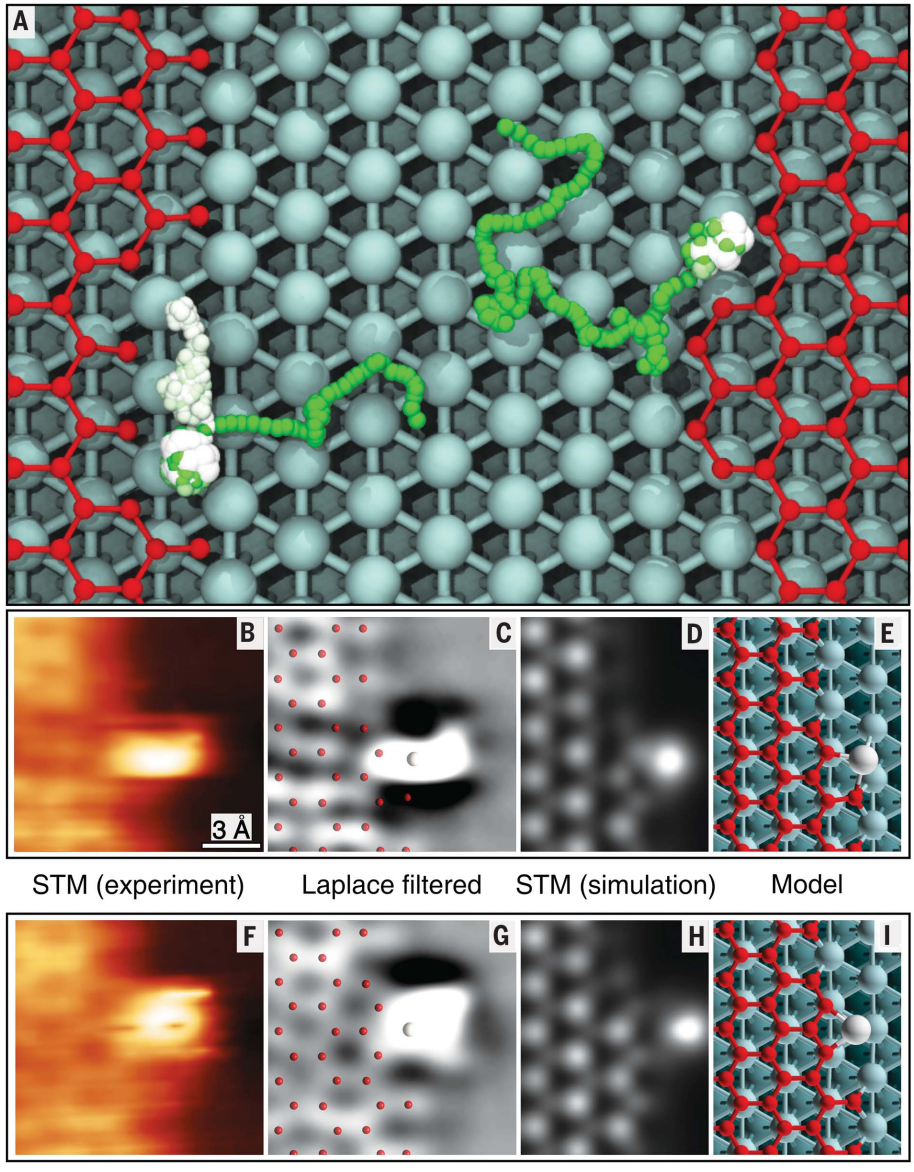


Fig. 2. Nickel adatoms at the graphene edges. (**A**) Ni adatom diffusing on a surface in a region delimited by graphene z (right) and k (left) edges with kinks. Two representative trajectories obtained by MD simulations with ReaxFF performed at 710 K for 100 ps are shown. Color palette for Ni trajectories: from green (initial position) to white (final position). The final steps are highlighted by increasing the ball size. (**B** to **I**) Short-lived configurations of Ni adatom at k-edge kinks: (B and F) High-speed STM images from movie S2 (*20*) (V = 20 mV; I = 7 nA); (C and G) Laplace-filtered version of images from (B) and (F) with superimposed ball models; (D and H) constant-height STM simulated images based on the calculated geometries (E and I).

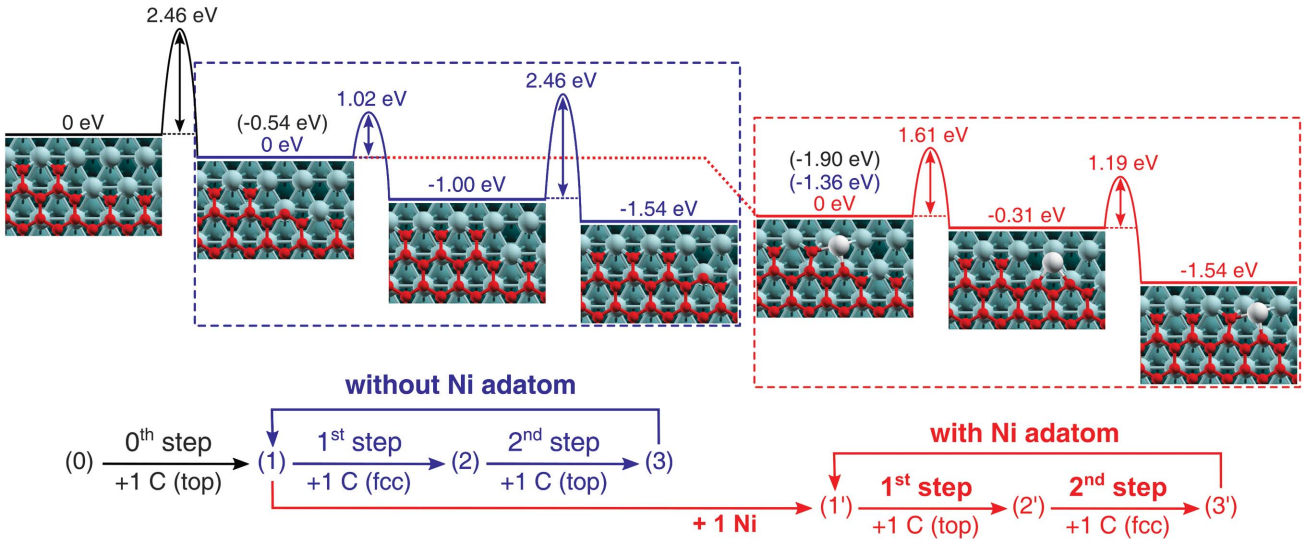


Fig. 3. Alternative graphene growth pathways as predicted by DFT. Energy zero corresponds to the bare kink configuration (0), plus one Ni adatom on the surface far from the edge and a reservoir of sub-fcc C atoms far from the edge. For clarity, the energy scale is reset for each cycle, producing the blue and red energy scales, respectively.

we observe. Apparent height and lateral dimensions of the mobile species are, on the contrary, comparable to those of the static pointlike defects observed after CVD graphene growth on Ni(111) in the same temperature range, which were previously identified as substitutional Ni atoms trapped in the graphene network during the growth process (*13*). We therefore attribute the bright features in Fig. 1, B and C, to mobile Ni adatoms, in line with previous DFT-based prediction (*24*). This hypothesis is supported by the well-known presence of diffusing adatoms on metal surfaces at elevated temperatures (*25*). Other dashed features with lower apparent height (comparable to or lower than that of carbon atoms in graphene) were visible in some frames and can be attributed either to Ni imaged by the tip in an off-center position (*26*), or to moving C atoms. A further confirmation of our identification of the bright features in the STM images as Ni adatoms comes from molecular dynamics (MD) simulations with state-of-the-art reactive force field (ReaxFF) performed at 710 K [Fig. 2A; details in (*20*)]. Mobile Ni adatoms moved randomly over the bare metal surface until they reached a graphene edge. There, they diffused parallel to the edge, with considerably longer residence time in the kink sites.

A notable feature in the STM image series is that the presence of the Ni adatoms at the kinks is in most of the cases accompanied by C dimer attachment nearby (see frame at 55 ms in Fig. 1B, and frames at 33 and 50 ms in Fig. 1C), suggesting a catalytic role of the single metal atom. This observation is consistent with previous DFT calculations for graphene on Cu(111) (*19*), which showed that the barrier for C incorporation at the z edge was strongly reduced by Cu adatom passivation at armchair-like sites, present near a hexagon (a kink) sticking out of the edge. Our interpretation is also in line with recent in situ transmission electron microscopy reports of the chemical activity of single atoms in facilitating addition and removal of C atoms from the edges of suspended graphene flakes (*27*). These previous results, although notable, are not related to a real growth process, but rather are related to attachment-detachment events of individual C atoms, and thus cannot be directly extended to technologically relevant CVD growth processes on solid metal surfaces.

That the Ni adatoms were almost exclusively imaged at the kink sites suggests that they represent a minimum-energy configuration for Ni; nevertheless, the short residence time implies that the bond can be easily broken at these temperatures, avoiding the poisoning of the active kink site. No stable Ni trapping was observed in our time series, at variance with the defect formation pathway predicted by Wang *et al.* (*24*).

From the STM sequence corresponding to the k edge, two short-lived Ni adatom configurations can be identified at the kink sites (Fig. 2, B, C, F, and G). DFT simulations of these configurations yielded the two stable structures shown in Fig. 2, E and I, where a Ni adatom and one or two top C atoms were added to the bare kink. The corresponding constant-height simulated images in Fig. 2, D and H, match the experimental STM data. DFT calculations indicate a pronounced mutual stabilization effect between C and Ni adatoms at the kink (from the energy comparison of configurations in figs. S2 and S3). These results suggest that Ni had a strong promoting role in the addition of C atoms to the kink and hence in the edge growth, as previously predicted for Cu(111) (*19*). Similar conclusions could be drawn for the z edge, where a single intermediate state was imaged and described by DFT calculations (fig. S4).

The stable configurations observed by STM and confirmed by theory can serve as a guide for a DFT description of the entire growth process, discriminating among many possible alternative paths and highlighting the catalytic role of the Ni adatom. In Fig. 3, we report the energy diagram corresponding to selected C-attachment pathways to the k edge with and without the presence of the Ni adatom. The energy gain upon addition of a C atom is referred to a subsurface position far from the edge ["sub-fcc," which is the most stable position on a clean Ni(111) surface]; for the addition of a Ni adatom, to a position on the surface far from the edge.

We take as initial configuration [labeled (0) in Fig. 3] a k edge with all terminating C atoms in fcc-hollow. The addition of a first C atom on the top site (1) involves a considerable energy barrier, 2.46 eV. The growth process can then follow two alternative cycles, either without (blue box) or with (red box) a Ni adatom.

Considering the process without a Ni adatom, the attachment of the second C atom in a fcc-hollow position (2) requires overcoming a barrier of 1.02 eV (see fig. S5 for details on energy barriers). Because of the epitaxial match between graphene and Ni(111), the edge configuration is now equivalent to the initial one. Further addition of a C top atom (3) requires again overcoming the 2.46-eV barrier, completing the cycle that can be repeated in a continuous loop, but always including this high energy barrier.

In the alternative path, a Ni adatom attaches without an energy barrier to the kink decorated by the top C atom (1), yielding an energy gain of 1.36 eV (1′). The most favorable configuration for attachment of a second C atom is again a top site, with the Ni adatom diffusing laterally, bonding to the two top C atoms (2′). In the last step of the growth cycle, another C atom attaches

in the fcc-hollow position, extending the graphene edge and pushing aside the Ni adatom (3′). The energy barriers involved in the two steps of this cycle are 1.61 and 1.19 eV, respectively. Ni adatoms, spontaneously binding to the kink site, thus reduce the rate-limiting energy barrier of the cyclic process by about 35% (from 2.46 to 1.61 eV), acting as single-atom catalysts for the graphene growth process.

Because the Ni attachment is barrierless, the detachment barriers from the two stable states at the kink are 1.36 and 1.35 eV, obtained as the DFT energy differences of the corresponding equilibrium states, i.e., configurations (1′) and (2′) of Fig. 3, and configurations (1) in Fig. 3 and (e) in fig. S2, respectively. These values yield Ni residence time values on the millisecond scale, compatible with the values extracted from the STM data (fig. S6).

REFERENCES AND NOTES

1. X. F. Yang *et al*., *Acc. Chem. Res.* **46**, 1740–1748 (2013).
2. A. Baraldi *et al*., *J. Phys. Chem. C* **115**, 3378–3384 (2011).
3. T. Zambelli, J. Wintterlin, J. Trost, G. Ertl, *Science* **273**, 1688–1690 (1996).
4. B. L. M. Hendriksen *et al*., *Nat. Chem.* **2**, 730–734 (2010).
5. B. Hammer, J. K. Nørskov, *Nature* **376**, 238–240 (1995).
6. B. Qiao *et al*., *Nat. Chem.* **3**, 634–641 (2011).
7. J. Lin *et al*., *J. Am. Chem. Soc.* **135**, 15314–15317 (2013).
8. G. Kyriakou *et al*., *Science* **335**, 1209–1212 (2012).
9. Y. C. Yang, A. Taranovskyy, O. M. Magnussen, *Angew. Chem. Int. Ed.* **51**, 1966–1969 (2012).
10. Z. Feng *et al*., *ACS Nano* **9**, 8697–8709 (2015).
11. X. Zhang, J. Xin, F. Ding, *Nanoscale* **5**, 2556–2569 (2013).
12. R. Rosei *et al*., *Phys. Rev. B* **28**, 1161–1164 (1983).
13. L. L. Patera *et al*., *ACS Nano* **7**, 7901–7912 (2013).
14. F. Bianchini, L. L. Patera, M. Peressi, C. Africh, G. Comelli, *J. Phys. Chem. Lett.* **5**, 467–473 (2014).
15. L. L. Patera *et al*., *Nano Lett.* **15**, 56–62 (2015).
16. F. Esch *et al*., *Rev. Sci. Instrum.* **82**, 053702 (2011).
17. https://fastmodule.iom.cnr.it/
18. V. I. Artyukhov, Y. Liu, B. I. Yakobson, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15136–15140 (2012).
19. H. Shu, X. Chen, X. Tao, F. Ding, *ACS Nano* **6**, 3243–3250 (2012).
20. See supplementary materials.
21. T. Mitsui, M. K. Rose, E. Fomin, D. F. Ogletree, M. Salmeron, *Science* **297**, 1850–1852 (2002).
22. E. Loginova, N. Bartelt, P. J. Feibelman, K. F. McCarty, *New J. Phys.* **10**, 093026 (2008).
23. B. Wang *et al*., *J. Phys. Chem. C* **121**, 9413–9423 (2017).
24. L. Wang, X. Zhang, H. L. W. Chan, F. Yan, F. Ding, *J. Am. Chem. Soc.* **135**, 4476–4482 (2013).
25. G. Ehrlich, *Surf. Sci.* **246**, 1–12 (1991).
26. J. Wintterlin *et al*., *Surf. Sci.* **394**, 159–169 (1997).
27. J. Zhao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15641–15646 (2014).

ACKNOWLEDGMENTS

We thank C. Cepek, J. Repp, and A. Peronio for fruitful discussions and K. C. Prince for carefully reading our manuscript. We thank the Detectors and Instrumentation Group at ELETTRA-Sincrotrone Trieste and F. Esch for their contribution to the FAST module development. **Funding:** We acknowledge funding from the EU-H2020 Research and Innovation programme under grant agreement no. 654360 Nanoscience Foundries and Fine Analysis-Europe; Italian Ministry of Foreign Affairs, through the Executive Programme with Argentina 2014–2016 (PGR00190); University of Trieste through the program "Finanziamento di Ateneo per progetti di ricerca scientifica–FRA 2015"; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Fondo para la Investigación Científica y Tecnológica (FONCyT) PICT-2015-2191, and Secretaría de Ciencia y Tecnología (SeCyT)-Universidad Nacional de Córdoba (UNC) (Argentina). Computational resources were obtained from the CINECA Italian Consortium of Universities through the Italian SuperComputing Resource Allocation (ISCRA) initiative and the agreement with the University of Trieste and from Centro de Computación de Alto Desempeño (CCAD), Universidad Nacional de Córdoba (Argentina). **Author contributions:** L.L.P. and C.A. conceived the experiments. L.L.P performed the experiments and analyzed the experimental data. M.P. conceived the simulations. F.B. and G.S. performed and analyzed the simulations. C.D. developed the FAST movie analysis software. L.L.P., C.A., M.M., M.P., and G.C. discussed the results. L.L.P., C.A., M.P., G.C. co-wrote the paper. All authors revised the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data are reported in the main text and supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1243/suppl/DC1
Materials and Methods
Figs. S1 to S8
Movies S1 and S2
References (*28*–*31*)

1 June 2017; resubmitted 4 December 2017
Accepted 25 January 2018
10.1126/science.aan8782

ENZYMOLOGY

Organometallic and radical intermediates reveal mechanism of diphthamide biosynthesis

Min Dong,[1] Venkatesan Kathiresan,[2] Michael K. Fenwick,[1] Andrew T. Torelli,[1] Yang Zhang,[1] Jonathan D. Caranto,[1] Boris Dzikovski,[1] Ajay Sharma,[2] Kyle M. Lancaster,[1] Jack H. Freed,[1] Steven E. Ealick,[1]* Brian M. Hoffman,[2]* Hening Lin[1,3]*

Diphthamide biosynthesis involves a carbon-carbon bond-forming reaction catalyzed by a radical S-adenosylmethionine (SAM) enzyme that cleaves a carbon-sulfur (C–S) bond in SAM to generate a 3-amino-3-carboxypropyl (ACP) radical. Using rapid freezing, we have captured an organometallic intermediate with an iron-carbon (Fe–C) bond between ACP and the enzyme's [4Fe-4S] cluster. In the presence of the substrate protein, elongation factor 2, this intermediate converts to an organic radical, formed by addition of the ACP radical to a histidine side chain. Crystal structures of archaeal diphthamide biosynthetic radical SAM enzymes reveal that the carbon of the SAM C–S bond being cleaved is positioned near the unique cluster Fe, able to react with the cluster. Our results explain how selective C–S bond cleavage is achieved in this radical SAM enzyme.

Diphthamide is a posttranslationally modified histidine residue on archaeal and eukaryotic translation elongation factor 2 (EF2), a protein essential for ribosomal protein synthesis (*1*–*3*). Diphthamide biosynthesis involves at least four steps and seven proteins (*4*). The first step of the biosynthesis is the transfer of a 3-amino-3-carboxypropyl (ACP) group from S-adenosyl-L-methionine (SAM) to a histidine residue of EF2 (Fig. 1). The enzyme that performs this reaction is a Dph2 (diphthamide biosynthesis protein 2) homodimer in archaea, such as *Pyrococcus horikoshii* (*Ph*Dph2) (*5*) or a Dph1-Dph2 heterodimer in eukaryotes (*6*). Dph2 binds an essential [4Fe-4S] cluster and is thought to use the reduced (1+) state of the cluster to cleave the $C_{\gamma,Met}$–S bond of SAM to generate a 3-amino-3-carboxypropyl (ACP) radical. The formation of the ACP radical is supported by the *Ph*Dph2-catalyzed generation of 2-aminobutyric acid and homocysteine sulfinic acid in the absence of the substrate protein, *Ph*EF2 (*5*). Additional support for a radical mechanism was provided by the reaction with a carboxyallyl SAM analog, SAM_{CA} (*7*). However, no radical intermediate has been directly observed in the reaction with SAM itself, and the detailed reaction mechanism remains unknown.

Despite using the same components, a [4Fe-4S] cluster and SAM, *Ph*Dph2 is structurally unrelated to the much larger family of 5′-deoxyadenosine radical (5′-dA•) forming radical SAM (RS) enzymes. In those enzymes, the $C_{5',Ade}$–S bond of SAM is cleaved, generating a 5′-dA• that then initiates downstream reactions (Fig. 1) (*8*). The question of how the *Ph*Dph2 homodimer or eukaryotic Dph1-Dph2 bind SAM and achieve a different SAM cleavage pattern remains unanswered. Here, we report biochemical and spectroscopic studies of two kinetically competent intermediates that, together with x-ray crystal structures of archaeal Dph2 homodimers in complex with SAM, allow us to propose a reaction mechanism for this class of RS enzymes.

We used rapid freeze-quench (RFQ) to arrest reaction mixtures, which were then analyzed by electron paramagnetic resonance (EPR) and electron nuclear double-resonance (ENDOR) spectroscopies for radical intermediates. Because the optimum temperature of *Ph*Dph2 is above 60°C, we opted to use the yeast Dph1-Dph2 system, which is active at room temperature and thus more convenient for RFQ experiments. A solution containing Dph1-Dph2 and SAM was mixed with dithionite and freeze-quenched at 500 ms to 4 min. Each sample displayed a new EPR signal, with *g*-parallel value ($g_{||}$) = 2.036, *g*-perpendicular value ($g_{\perp}$)= 2.005 at 12 K (fig. S1). The intensity of this signal reached a maximum at 2 s quench time and decreased at longer delays (Fig. 2A). This species is not an ACP radical, which would not show such a high *g* shift and would exhibit resolvable ^{1}H hyperfine splittings from the methylene radical protons; nor is it an S-based radical, which could exhibit such *g* values but again would give resolved hyperfine ^{1}H splittings. Based on these observations and experience from previous work using SAM_{CA} (*7*), we interpreted this new species as having an iron-sulfur cluster-based organometallic structure. Importantly, the EPR spectrum and *g* values of this species resemble those of the organometallic intermediate that was detected in the RS enzyme, pyruvate formate lyase-activating enzyme (PFL-AE) and assigned to an Fe-5′–C bond between the deoxyadenosyl group and the unique cluster iron (*9*). As confirmation, when this intermediate was generated with ^{57}Fe-enriched enzyme, its EPR spectrum exhibited ^{57}Fe-hyperfine line broadening (fig. S1). Thus, the reaction catalyzed by Dph1-Dph2 produces an organometallic intermediate. Its *g* values follow the pattern of a $[4Fe\text{-}4S]^{3+}$ cluster, $g_{||} > g_{\perp} > 2$ (*10*), where $g_{||}$ is sensitive to coordination at the unique Fe, suggesting a possible formal description of the intermediate as containing a carbanion bound to an oxidized cluster.

To establish the structure of this intermediate, we collected RFQ ENDOR samples prepared with (methionine-$^{13}C_5$)–SAM, in which the methionine carbons of SAM were uniformly labeled with ^{13}C. Field-modulated continuous wave (CW) ENDOR spectra obtained at 2 K of this sample exhibits a ^{13}C doublet, with a ^{13}C hyperfine coupling constant of $A \approx 7.8$ MHz (Fig. 3A). This coupling constant is comparable to $A_{iso} = 9.4$ MHz observed for the Fe–[5′-^{13}C]–deoxyadenosyl bond

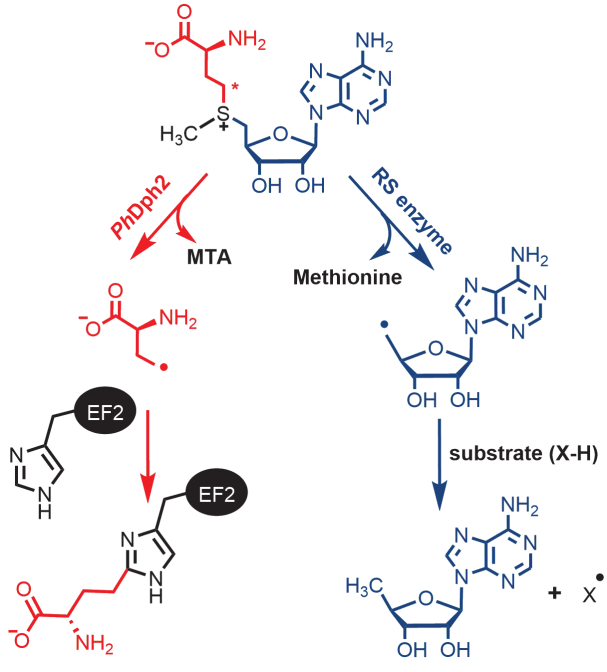


Fig. 1. *Ph*Dph2 and 5′-dA•–forming radical SAM enzyme catalyzed reactions. The red asterisk labels $C_{\gamma,Met}$ of SAM.

[1]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA. [2]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [3]Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA.
*Corresponding author. Email: see3@cornell.edu (S.E.E.); bmh@northwestern.edu (B.M.H.); hl379@cornell.edu (H.L.)

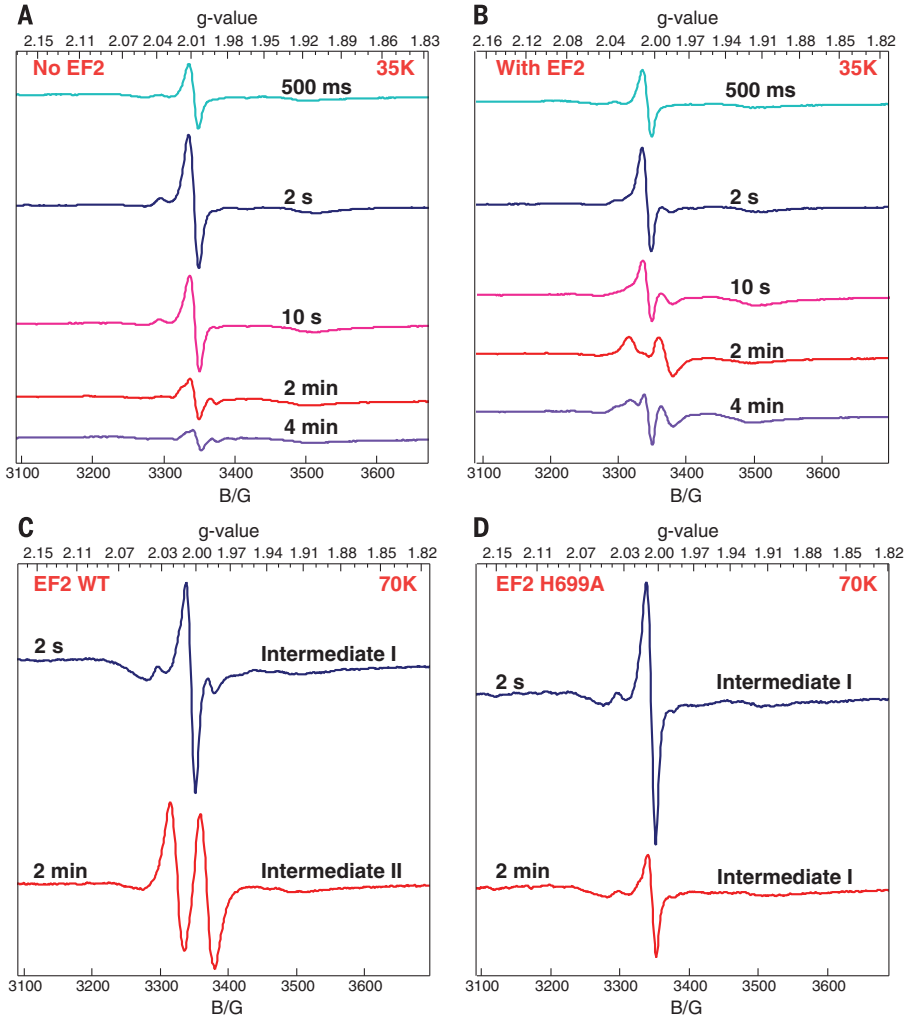


Fig. 2. X-band CW EPR spectra of RFQ samples showing the formation of the organometallic intermediate I and organic radical intermediate II. B/G (magnetic field/Gauss). (**A**) Reaction of reduced Dph1-Dph2 with SAM quenched at various time points (T = 35 K). (**B**) Reaction of reduced Dph1-Dph2 with SAM and EF2 quenched at various time points (T = 35 K). (**C**) Reaction of reduced Dph1-Dph2 with SAM and wild-type EF2 quenched at 2 s and 2 min (T = 70 K). (**D**) Reaction of reduced Dph1-Dph2 with SAM and EF2 H699A mutant quenched at 2 s and 2 min (T = 70 K). In early-time spectra, a variable contribution at g = 2 from an unidentified organic radical enhances the intensity of the $g_{\perp}$ feature of intermediate I relative to that at $g_{||}$ and distorts its shape; fig. S1 presents spectra with little of the radical and thus a more faithful lineshape of intermediate I.

in the PFL-AE intermediate and together with the ^{57}Fe hyperfine broadening provides strong support that this intermediate contains an Fe–C bond. However, in this case, the Fe-bonded carbon must originate from the ^{13}C-labeled methionine, presumably the $C_{\gamma,Met}$.

In testing whether this organometallic species (denoted intermediate I) was an active reaction intermediate, we collected a RFQ EPR time course in the presence of the substrate protein, EF2. Intermediate I accumulated in the first 10 s (Fig. 2B). At later time points, the intermediate I signal diminished concomitantly with the appearance of a doublet radical signal centered at g = 2 (Fig. 2B). This new signal, denoted intermediate II, reached a maximum at 2 min quench time and decreased at 4 min (Fig. 2B) in a process that ultimately leads to product formation.

The intermediate II EPR signal was distorted by saturation effect at 35 K (Fig. 2B) but was readily observed at 70 K (Fig. 2C), where spin relaxation is faster. Tentative assignment of the doublet splitting to coupling to a single proton with a large hyperfine splitting ($A \approx 120$ MHz) led us to hypothesize that this intermediate is an organic radical generated when the organometallic intermediate I reacted with the enzymatic target, histidine 699 (H699) of yeast EF2. To test this, we repeated the RFQ experiment with the EF2 His^{699} to alanine mutant (H699A). With this EF2 mutant, only the organometallic intermediate I was detected at both 2 s and 2 min (Fig. 2D), and no intermediate II signal was detected. This experiment provides strong support that intermediate I is chemically competent.

To test the proposed structure, intermediate II was RFQ-trapped with (methionine-$^{13}C_5$)–SAM. The resulting X-band EPR spectrum showed additional splitting of $A_{iso}(^{13}C) \approx 59$ MHz and broadening compared with that of the natural-abundance SAM (Fig. 3B), demonstrating that the radical incorporates the ACP group of SAM. The EPR-resolved doublet, $A(^1H) \approx 120$ MHz, observed for intermediate II collapsed to a singlet when intermediate II was prepared with $^2H_5$-His–labeled EF2 protein (Fig. 3B), confirming that the splitting observed with the natural-abundance EF2 was associated with ^{1}H of the histidine residue on EF2. This was further confirmed by ^{1}H ENDOR of intermediate II with EF2, which exhibited a nonexchangeable ^{1}H signal with the same coupling (fig. S2); an exchangeable coupling, $A(^1H) \approx 15$ MHz not resolved in the EPR spectrum was also seen. These results confirm that intermediate II is a radical produced by the reaction of an ACP radical from intermediate I with the imidazole ring of His^{699} of the EF2 substrate.

Given the structure of the final diphthamide product, a candidate for the structure of intermediate II is an ACP-modified histidine radical, as shown in Fig. 4. The shape of the ^{1}H doublet in the EPR spectrum (Fig. 3B) and the strongly coupled ^{1}H ENDOR signal (fig. S2) suggest that the proton coupling is essentially isotropic, as expected for a proton β to the spin in the N-2p-π orbital (Fig. 4 and fig. S3) (*11*); the resolved ^{13}C splitting in the EPR spectrum of intermediate II (Fig. 3B) prepared with (methionine-$^{13}C_5$)–SAM must come from the carbon of the ACP fragment bound to the His, again β to the spin site. The individual peaks of the ^{1}H doublet are broad enough to contain unresolved ^{14}N splitting introduced by the spin density on the ^{14}Nδ of His (Fig. 4). As shown in fig. S4, the EPR spectrum of the natural-abundance radical is well simulated with a strong, nearly isotropic ^{1}H coupling ($A_1 = A_2$ = 122 MHz; A_3 = 103 MHz) and a highly anisotropic ^{14}N

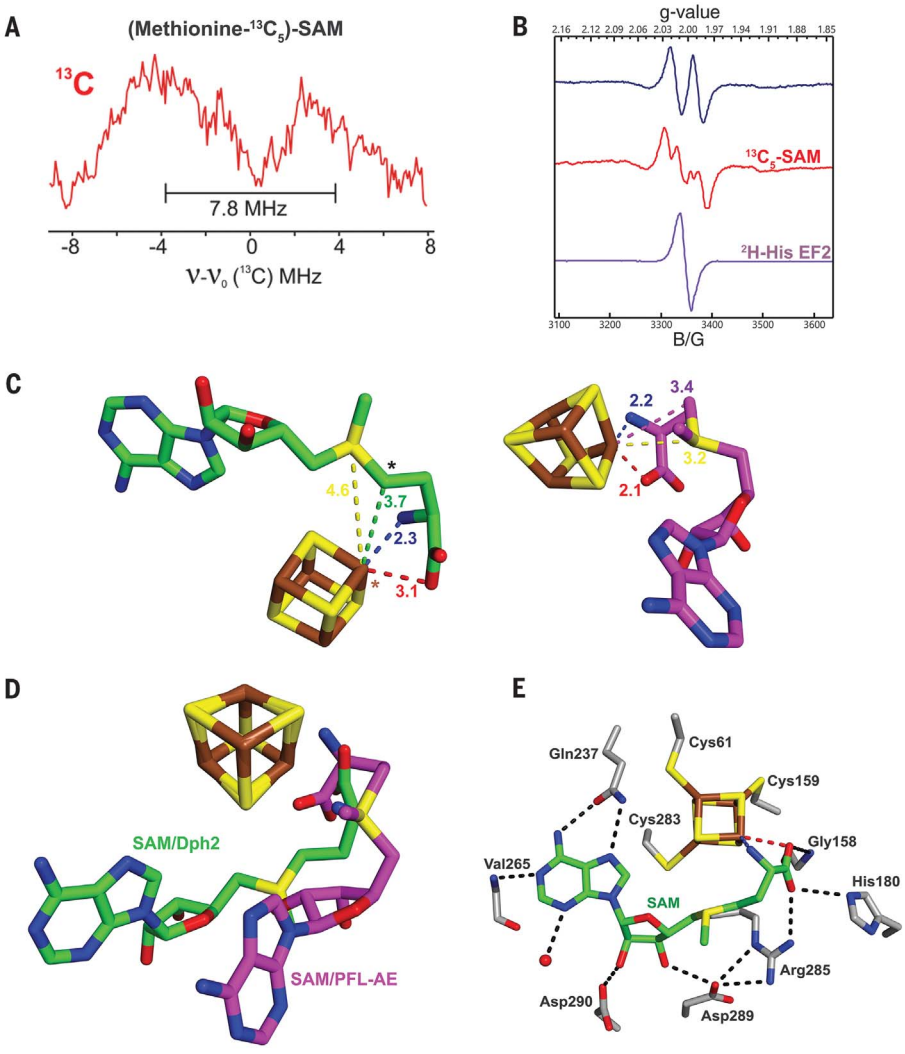


Fig. 3. EPR and ENDOR spectra of isotope-labeled intermediates and crystal structures of SAM-bound *Cmn*Dph2. (**A**) 35 GHz CW ^{13}C ENDOR spectra for intermediate I with (methionine-$^{13}C_5$)–SAM. (**B**) Isotope EPR study of the organic radical intermediate II on EF2 quenched at 2 min (T = 70 K). Intensity adjusted for comparison. (**C**) Cluster and bound SAM in the structure of *Cmn*Dph2 (left), in comparison with cluster and bound SAM in the structure of PFL-AE (right, PDB 3CB8) (*13*). Distances are given in angstroms. The black and brown asterisks label $C_{\gamma,Met}$ of SAM and the differentiated iron of the [4Fe-4S] cluster, respectively. (**D**) Overlay of SAM and cluster in *Cmn*Dph2 and PFL-AE structures. (**E**) *Cmn*Dph2 active site showing the binding of SAM. Hydrophobic interactions with purine ring of SAM are not shown for clarity.

hyperfine coupling ($A_1 = A_2 = 4$ MHz, $A_3 = 54$ MHz). ^{15}N features seen in the ENDOR of the ^{15}N-His EF2 sample (fig. S5) likely are a superposition of signals associated with the A_1/A_2 components of this histidine nitrogen (assigned as Nδ in Fig. 4), and those from the less strongly coupled His ^{15}Nε. Finally, weakly coupled ^{13}C ENDOR signals seen when intermediate II was prepared with ^{13}C-His EF2 (figs. S5 and S6) are assigned to His carbons bonded to the spin-bearing Nδ (Fig. 4) and account for the slight broadening in the components of the ^{1}H doublet in this isotopologue; an exchangeable ^{1}H seen in ENDOR (fig. S2) is thought to reside on Nε of His.

The time scale for intermediate II formation measured by EPR tracked that of intermediate I decay, whereas the reaction product, ACP-modified EF2 as quantified by mass spectrometry, accumulates on the same time scale as the intermediate II decay (fig. S7). These results indicate that both intermediate I and intermediate II are formed in a kinetically competent fashion. Thus, we have trapped two sequential intermediates, an organometallic intermediate I, which gives rise to an organic radical intermediate II, which goes on to complete the first step of diphthamide biosynthesis.

To gain further understanding into the reaction mechanism, we obtained crystal structures of the RS enzymes involved in diphthamide biosynthesis in complex with SAM. A 2.3-Å resolution structure of [4Fe-4S]$^{2+}$–loaded *Ph*Dph2 in complex with SAM showed some evidence of SAM cleavage. A 2.1-Å resolution structure of *Candidatus Methanoperedens nitroreducens* Dph2 (*Cmn*Dph2) showed mostly uncleaved SAM. Both structures showed average N–Fe and O–Fe distances of 2.3 and 3.1 Å, respectively (Fig. 3C and fig. S8), suggesting that, similar to 5′-dA•–forming RS enzymes, the amino and carboxylate groups of SAM coordinate the unique Fe of the [4Fe-4S] cluster in Dph2. This is consistent with our recent study showing that both the amino and carboxylate of SAM are important for the *Ph*Dph2 and Dph1-Dph2–catalyzed reactions (*12*).

However, SAM in both *Ph*Dph2 and *Cmn*Dph2 binds with a distinct geometry compared with that in 5′-dA•–forming RS enzymes such as PFL-AE (*13*) (Fig. 3D). We further obtained additional crystal structures using both *Ph*Dph2 and *Cmn*Dph2 that support the geometry shown in Fig. 3, C and D (see figs. S9 to S11). The SAM binding site of *Cmn*Dph2 is shown in Fig. 3E. Gln^{237} forms hydrogen bonds with N6 and N7 of the adenine ring and the amide nitrogen of Val^{265} hydrogen bonds to N1. The adenine ring is sandwiched between Phe^{58} and Ile^{286}. Asp^{289} and Asp^{290} form hydrogen bonds with the hydroxyl groups of the ribosyl moiety, whereas Arg^{285}, His^{180}, and Gly^{158} form hydrogen bonds with the carboxylate of SAM.

Most strikingly, the $C_{\gamma,Met}$ distance to the unique Fe, as averaged over the *Ph*Dph2 and *Cmn*Dph2 structures, is 3.7 Å, which is closer than S_{Met} to the Fe (the S_{Met}–Fe distance is 4.6 Å). In contrast, in the PFL-AE structure, the S_{Met} is closer to the unique Fe than $C_{5',Ade}$, with an S_{Met}–Fe distance of 3.2 Å (Fig. 3C). The observation that in *Ph*Dph2 and *Cmn*Dph2, $C_{\gamma,Met}$ was closer than S_{Met} to the unique Fe suggests that during the $C_{\gamma,Met}$–S_{Met} bond cleavage, the electron is transferred from the Fe–S cluster via $C_{\gamma,Met}$, whereas for PFL-AE (and generally believed for all 5′-dA•–forming RS enzymes), the electron is transferred from the Fe–S cluster via S_{Met} (*8*).

Based on the two intermediates that we have detected and the structures of *Ph*Dph2 and *Cmn*Dph2 in complex with SAM, we propose a reaction mechanism for the first step of

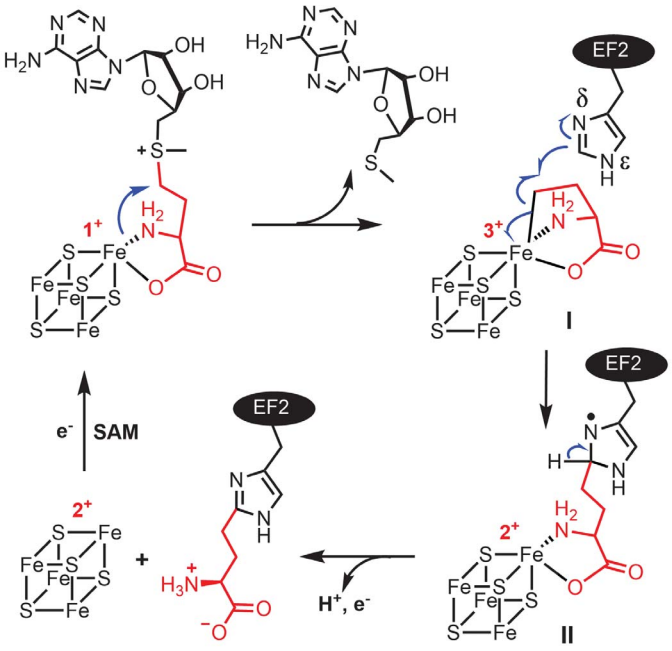


Fig. 4. The proposed reaction mechanism of diphthamide biosynthetic RS enzymes.

diphthamide biosynthesis (Fig. 4). The unique iron of a $[4Fe\text{-}4S]^{+}$ cluster in an archaeal Dph2 homodimer or eukaryotic Dph1-Dph2 heterodimer attacks the γ carbon of the methionine in SAM, generating a 3-amino-3-carboxypropyl-$[4Fe\text{-}4S]^{3+}$ organometallic intermediate I and 5′-methylthioadenosine (MTA). The organometallic intermediate I is essentially a stabilized ACP radical. In the presence of the substrate EF2, the Fe–C bond can break homolytically, allowing the ACP radical to react with the imidazole ring of the histidine residue of EF2 and generate the organic radical, intermediate II. Intermediate II then loses a proton and an electron to form an ACP-modified histidine.

In the study of PFL-AE, it was noted that the Fe-[5′-C]-deoxyadenosyl organometallic intermediate could be generated through a radical reaction or nucleophilic reaction (*9*). In our case, the formation of intermediate I could occur via homolytic reductive cleavage of SAM to form an ACP radical, which then reacts with the unique Fe (stepwise one-electron transfer) to form the organometallic intermediate I. Alternatively, a concerted one-step, two-electron transfer (nucleophilic attack on $C_{\gamma,Met}$) can form intermediate I directly. The fact that $C_{\gamma,Met}$ was only 3.7 Å away from the unique Fe, whereas the sulfonium moiety was 4.6 Å away (Fig. 3C), favors the latter mechanism, although we cannot eliminate the homolytic cleavage mechanism.

The structure also provides a simple solution to the question, how does the archaeal Dph2 homodimer or eukaryotic Dph1-Dph2 heterodimer cleave a different C–S bond in SAM compared with all 5′-dA•–forming RS enzymes? If the cluster interacts with and transfers electrons to $C_{\gamma,Met}$ instead of S_{Met}, the only C–S bond that can be broken in SAM is the $C_{\gamma,Met}$–S bond. Previously, Kampmeier proposed that a stereo-electronic control mechanism could explain the different cleavage patterns (*14*). This model rationalizes the C–S bond cleavage in RS enzymes using a radical displacement reaction, with the formation of an Fe–S bond accompanying the cleavage of the C–S bond (fig. S12). This radical displacement reaction also requires that the C–S–Fe atoms involved in the release of a radical on carbon are arranged roughly colinearly. Thus, if SAM in archaeal Dph2 homodimer or eukaryotic Dph1-Dph2 is bound in a conformation with S_{Met} close to the Fe–S cluster and colinear $C_{\gamma,Met}$–S–Fe arrangement, it would lead to the cleavage of $C_{\gamma,Met}$–S. However, our structure indicates that in archaeal Dph2 homodimer, the $C_{\gamma,Met}$ is close to the unique Fe. The $C_{\gamma,Met}$ mediates electron transfer from the Fe–S cluster, dictating which bond is cleaved. This is an important difference between diphthamide biosynthetic RS enzymes and 5′-dA•–forming RS enzymes. Why does nature choose this way to control the bond cleavage in diphthamide biosynthetic RS enzymes? Perhaps conformational constraints on SAM prevent the $C_{\gamma,Met}$–S–Fe colinear arrangement proposed by Kampmeier, whereas the roughly colinear S–$C_{\gamma,Met}$–Fe arrangement as revealed by the structures is feasible (fig. S12).

In summary, this study has explained how diphthamide biosynthetic RS enzymes create the EF2 product through cleavage of the $C_{\gamma,Met}$–S bond of SAM in a process that involves first an organometallic intermediate, then a radical intermediate that collapses to product. Together with the recent study showing an organometallic intermediate in a 5′-dA•–forming RS enzyme PFL-AE, these studies further suggest that an organometallic intermediate may serve as a stabilized form of the highly reactive primary organic radical and provide a strong parallel between RS enzymes and adenosylcobalamin-dependent enzymes.

REFERENCES AND NOTES

1. E. A. Robinson, O. Henriksen, E. S. Maxwell, *J. Biol. Chem.* **249**, 5088–5093 (1974).
2. B. G. Van Ness, J. B. Howard, J. W. Bodley, *J. Biol. Chem.* **255**, 10717–10720 (1980).
3. B. G. Van Ness, J. B. Howard, J. W. Bodley, *J. Biol. Chem.* **255**, 10710–10716 (1980).
4. R. Schaffrath, W. Abdel-Fattah, R. Klassen, M. J. Stark, *Mol. Microbiol.* **94**, 1213–1226 (2014).
5. Y. Zhang *et al.*, *Nature* **465**, 891–896 (2010).
6. M. Dong *et al.*, *J. Am. Chem. Soc.* **136**, 1754–1757 (2014).
7. M. Dong *et al.*, *J. Am. Chem. Soc.* **138**, 9755–9758 (2016).
8. J. B. Broderick, B. R. Duffus, K. S. Duschene, E. M. Shepard, *Chem. Rev.* **114**, 4229–4317 (2014).
9. M. Horitani *et al.*, *Science* **352**, 822–825 (2016).
10. G. E. Cutsail 3rd, J. Telser, B. M. Hoffman, *Biochim. Biophys. Acta* **1853**, 1370–1394 (2015).
11. A. Carrington, A. D. McLachlan, *Introduction to Magnetic Resonance with Applications to Chemistry and Chemical Physics* (Harper & Row, New York, 1967).
12. M. Dong *et al.*, *J. Am. Chem. Soc.* **139**, 5680–5683 (2017).
13. J. L. Vey *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16137–16141 (2008).
14. J. A. Kampmeier, *Biochemistry* **49**, 10770–10772 (2010).

ACKNOWLEDGMENTS

We thank M. Luo at Memorial Sloan–Kettering Cancer Center for providing the methionine adenosyltransferase I plasmid. We thank P. Doan for help in confirming the organometallic nature of intermediate I and M. Oliver Ross for assistance in collecting ENDOR spectra. We thank S. Zhang (Proteomics and Metabolomics Facility at Cornell) for help with the mass spectrometry experiment. M. D. thanks T. L. Grove and R. D. Britt for helpful discussion. **Funding**: This work was supported by grants from the National Institutes of Health National Institute of General Medical Sciences (GM088276 to H.L., P41GM103521 to J.H.F., P41GM103403 to S.E.E, GM 111097 to B.M.H., and GM124908 to K.M.L.). This work is based on research conducted at the Northeastern Collaborative Access Team beamlines. The Pilatus 6M detector on 24-ID-C beamline is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. **Author contributions**: M.D. contributed to experimental design, performed the biochemical experiments, and prepared the proteins for spectroscopic studies. V.K. and A.S. performed the ENDOR studies. M.K.F., A.T.T., and Y.Z. carried out the crystallization and structure determination of all the proteins. J.D.C. and K.M.L. provided critical support for the RFQ experiments and EPR simulation and performed the fitting of kinetic data. M.D., B.D., and J.H.F performed the X-band EPR measurements. S.E.E. directed the x-ray structural studies; B.M.H. directed the ENDOR studies and contributed to experimental design; H.L. directed the biochemical studies, contributed to experimental design, and coordinated collaborations. M.D., H.L., B.M.H., and S.E.E. wrote the manuscript with input from V.K., M.K.F., A.T.T., Y.Z., and J.D.C. All authors reviewed, edited, and approved the manuscript. **Competing interests**: The authors declare no competing financial interests. **Data and materials availability**: Crystallographic models have been deposited in the Protein Data Bank under accession codes 6BXK for *Ph*Dph2/MTA, 6BXL for *Ph*Dph2/SAM, 6BXM for *Cmn*Dph2/SAM1, 6BXN for *Cmn*Dph2/SAM2, and 6BXO for *Cmn*Dph2/SAH. Raw data and materials associated with the main text and the supplementary materials are available for sharing upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1247/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*15*–*35*)

12 August 2017; accepted 30 January 2018
10.1126/science.aao6595

INDUCED SEISMICITY

Oklahoma's induced seismicity strongly linked to wastewater injection depth

Thea Hincks,[1,2] Willy Aspinall,[1,2,3] Roger Cooke,[4,5] Thomas Gernon[6]*

The sharp rise in Oklahoma seismicity since 2009 is due to wastewater injection. The role of injection depth is an open, complex issue, yet critical for hazard assessment and regulation. We developed an advanced Bayesian network to model joint conditional dependencies between spatial, operational, and seismicity parameters. We found that injection depth relative to crystalline basement most strongly correlates with seismic moment release. The joint effects of depth and volume are critical, as injection rate becomes more influential near the basement interface. Restricting injection depths to 200 to 500 meters above basement could reduce annual seismic moment release by a factor of 1.4 to 2.8. Our approach enables identification of subregions where targeted regulation may mitigate effects of induced earthquakes, aiding operators and regulators in wastewater disposal regions.

Oklahoma's ~900-fold increase in the annual rate of seismicity since 2009 (*1*) (fig. S1) makes it the most seismically active region in the contiguous United States. Wastewater disposal via deep injection wells explains the surge in earthquake activity (*1*–*7*). In many cases, pressurized fluids migrate from target formations downward into the crystalline basement (*1*, *7*–*10*), where most earthquakes occur (*11*–*18*). The disposed fluids may locally increase pore fluid pressure and reduce the effective stress along faults (*2*, *9*, *11*), resulting in induced seismicity (*2*, *19*). In September 2016, a magnitude (M_w) 5.8 earthquake centered near Pawnee caused injury and damage to buildings, prompting the immediate suspension of 37 disposal wells (*20*). The number of damaging earthquakes (*5*, *6*, *21*) resulted in litigation against well operators (*22*), increased risk to critical infrastructure (*23*), and considerable financial loss in the reinsurance industry (*24*).

The escalating seismicity compelled Oklahoma regulators to introduce emergency directives aimed at managing injection (*25*), informed by an empirical understanding of how seismicity responds to changes in injection rates or volumes (*2*, *4*, *6*). These measures appear to have decreased the earthquake count ($M_w \geq 3$) in the period 2015 (*21*) to the present (fig. S1). Considering the purported strong association between seismicity and oil production in Oklahoma (*26*), a major driver of this trend could have been abated injection due to relatively low oil prices. However, a case has been made (*27*) that the probabilities of occurrence of larger-magnitude earthquakes are not decreasing as rapidly as other models suggest (*1*). The total seismic moment release has declined only modestly despite decreased injection volumes (fig. S1). The ultimate impact of reduced injection may take years to materialize given observed lagging in the response of seismicity levels to injection (*11*, *27*, *28*).

The Oklahoma Corporation Commission (OCC) requires operators to prove that their injection does not encounter basement lithologies (*29*), but debate remains on the precise influence of injection depth. Although a study encompassing the wider Central and Eastern United States (CEUS) found no significant correlation between proximity to basement and earthquake occurrence (*4*), other studies suggest that injections near, or within, crystalline basement might increase the likelihood of seismicity locally (*12*, *30*, *31*). Therefore, we need to analyze the joint effects of spatial, operational, geologic, and seismicity parameters (*4*, *19*, *21*).

We developed a Bayesian network (BN) to quantitatively evaluate correlations between well operational parameters, geologic setting, and seismicity, focusing on the "Oklahoma induced seismicity zone" (*5*) (Fig. 1). Our BN framework capitalizes on mathematical advances (*32*) and innovative software, UNINET (*33*), implemented in R (*34*). UNINET models joint dependency using continuous data distributions, bypassing the limiting assumptions of discrete approximations (*32*). Monthly underground injection control (UIC) well data for January 2011 to September 2016 were obtained from OCC (*35*), including injected volume, pressure, duration of well activity, and total well depth. Earthquake location and magnitude data, from the Advanced National Seismic Systems' (ANSS) earthquake catalog, ComCat (*36*), were used to calculate total seismic moment release. We considered earthquakes of $M_w \geq 2$, with a hypocentral depth <10 km (fig. S3), and a hypocenter falling within 20 km of an injection well (*32*). We chose this distance on the basis of the documented occurrence of associated earthquake swarms within 10 km (*37*) to 20 km (*3*) of high-rate disposal wells, supported by additional calculations that indicate strong correlations between injection and moment up to ~20 km (*32*). Basement depths at well sites were estimated by spline interpolation (*32*, *38*) using 1232 depth records from Oklahoma Geological Survey (*39*) (fig. S5).

Over the study period, >10,000 active class II UIC wells in Oklahoma (Fig. 1, A and B) collectively injected on average ~2.3 billion barrels (bbl) of fluid per year via saltwater disposal (SWD) and enhanced oil recovery (EOR) (*35*) (figs. S1 and S2 and table S1). We considered class II injections into the full range of target geologic formations, to capture all potential contributions to induced seismicity (*32*). Wastewater disposal is particularly intensive within the Cambrian–Ordovician Arbuckle Group, a sedimentologically and structurally heterogeneous sequence of dolomitized carbonates and breccias (*40*) ~0.15 to 2 km thick throughout most of the study area. Sedimentary cover is hydraulically connected to Precambrian crystalline basement (*1*, *7*–*9*) (fig. S6) via an extensive fault network (*14*, *17*, *41*, *42*) (Fig. 1C).

High-rate injection has been linked to increased seismicity (*4*). However, analysis using BN and regression models indicate that other parameters have an important, but largely unquantified, impact on seismic hazard. We explored correlations between key operational and spatial characteristics and seismicity, comparing a number of network configurations with a simple linear regression (Fig. 2) to determine the most appropriate model (*32*). We defined nodes for key observables, including well operating parameters (depth, monthly, annual and cumulative injected volume), geologic parameters (distance to basement), and total seismic moment release within 20 km of the well in the year ahead (Fig. 2, A and B) (*32*)—a standard time scale for operational and seismic hazard forecasts (*5*, *6*, *21*). To improve forecasting for deeper and higher-volume wells associated with higher seismicity (Fig. 3 and fig. S7) (*32*), our models focused on wells ≥1 km deep and cases of monthly injection >10,000 bbl, i.e., those of greater concern to regulators (*20*). We used UNINET in "data mining" mode to estimate the set of joint normal copulae (multivariate probability distributions) that most closely represented the input data (*32*). We used the resulting models to estimate moment release for any given set of operating conditions.

The saturated BN (Fig. 2A) (*32*) gave the most complete representation of the observed data under stationary conditions, outperforming the regression model (in-sample testing, table S3 and figs. S9 and S10). However, the unsaturated BN (Fig. 2B) provided better forecasting under the changing injection regime of 2015 to 2016 (fig. S1B). This is because rank correlations between volume variables in the saturated BN preserved relationships (associated with increasing injection) that were no longer valid under post–2015 operating conditions.

[1]School of Earth Science, University of Bristol, Bristol BS8 1RJ, UK. [2]Cabot Institute, Royal Fort House, University of Bristol, Bristol BS8 1UJ, UK. [3]Aspinall & Associates, Cleveland House, High Street, Tisbury SP3 6HF, UK. [4]Delft University of Technology, Mekelweg 4, 2628 CD, Delft, Netherlands. [5]Resources for the Future, 1616 P Street NW, Washington, DC 20036, USA. [6]Ocean and Earth Science, University of Southampton, Southampton SO14 3ZH, UK.
*Corresponding author. Email: thomas.gernon@noc.soton.ac.uk

We used two test data sets for model verification (fig. S11 and table S3). The first demonstrated operational use of the BN to forecast future moment release, learning with observations from January 2011 to June 2015, and forecasting from July to December 2015. The second test used 90% of the full data set for learning (January 2011 to December 2015, randomly sampled) and the remaining 10% for verification (fig. S11) (*32*). We compared forecast results using the BN and regression models (Fig. 2, figs. S12 and S13, and table S6) and found that in both cases, the BN gave the best performance.

The spatial distribution of earthquakes (Fig. 1, B and C, and fig. S2) implies important underlying geologic characteristics influencing seismicity. The BN cannot explicitly represent these characteristics because of the variability and paucity of geologic data at relevant depths and resolution across the region. To address this issue, we applied the models iteratively to identify where (spatially) they systematically under or overforecast, because of latent features outside the scope of the model. We used both BN and regression models to estimate seismic moment for a subset of the training data and used the forecast error to compute a "geospatial correction" map—a proxy for unobserved geologic data (*32*) (fig. S8). We repeated the learning step with the training data using the geospatial correction as an additional input node. We used this revised network (Fig. 2) for final forecast verification (figs. S10, S12, and S13) (*32*).

The geospatial correction shows the strongest (negative) correlation (–0.55) with annual moment release (Fig. 2C), supporting the observation that lithologic (*9*) and fault network characteristics (*14*, *16*–*18*) collectively play a critical role in determining susceptibility to induced seismicity. Root mean squared errors, mean absolute errors, and median average deviation for the BN and regression models, both with and without the geospatial correction node, show the improvement in forecast skill resulting from this additional step (table S6, Test 2). This degree of improvement suggests that the geospatial correction effectively characterizes the impact of important, but unobservable (latent), factors. It could additionally prove valuable in mapping spatial variation in induced earthquake hazard (*21*).

Our analysis with UNINET shows that distance to the basement interface is more strongly correlated (–0.34) with annual seismic moment release than injected volume; and cumulative injected volume has a stronger correlation (+0.26) with annual seismic moment release than annual (+0.20) or monthly (+0.18) volume (Fig. 2C, empirical rank correlations). However, we expect these volume correlations to change with time owing to the effects of regulation, market-driven changes in oil production and associated disposal, and time lags between injection and induced seismicity. For example, cumulative volume will become less well correlated with annual moment if the current trend of reduced seismicity continues (fig. S1A).

UNINET uses joint normal copulae (*32*,*33*). As several input distributions are highly skewed and seismic moment is censored (because of catalog incompleteness for $M_w < 2.5$; fig. S4), the normal copula cannot perfectly represent the data. We see this in the difference between the empirical and BN unconditional rank correlations (table S5) and cumulative distribution functions (CDFs; Fig. 2 and figs. S12 and S13). However, calculations suggest that using optimal copulae would yield only a small improvement, negated by greatly increased computational complexity (*32*).

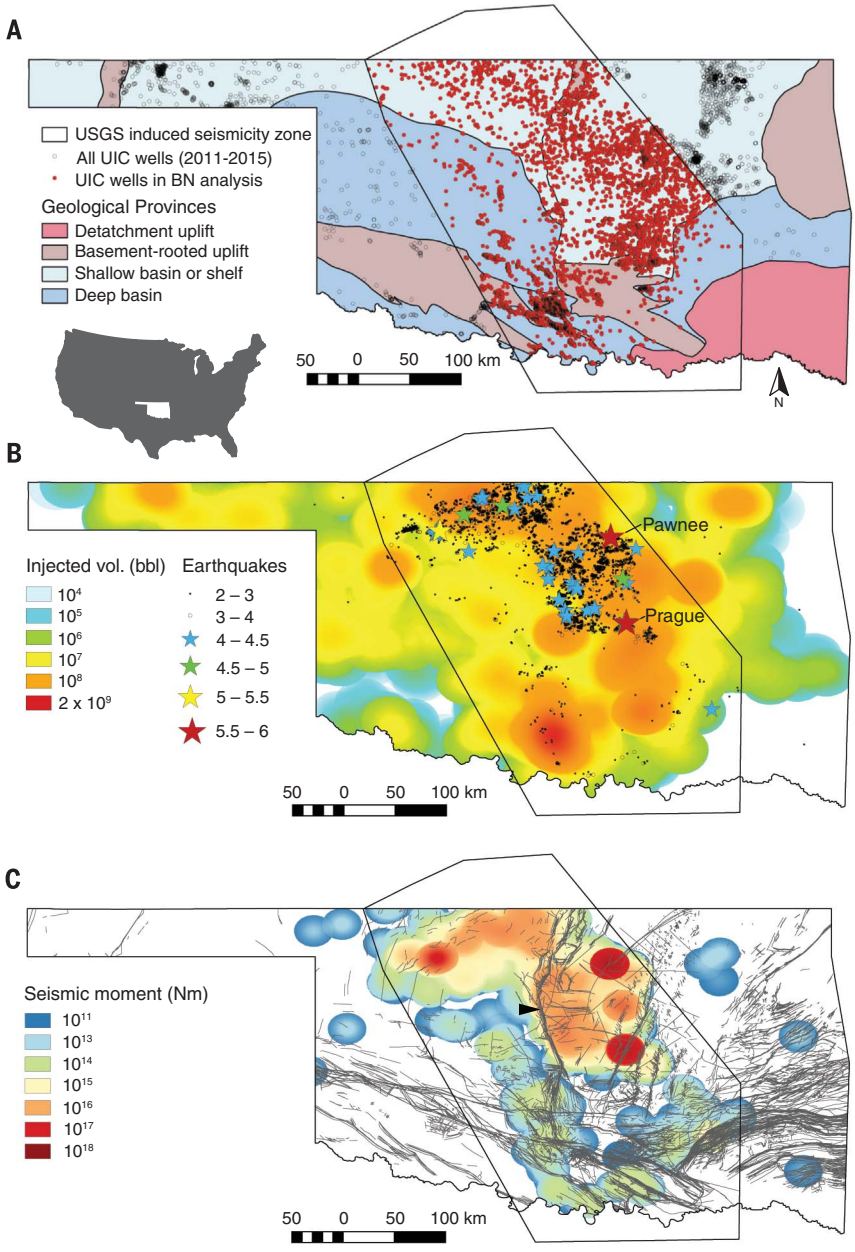
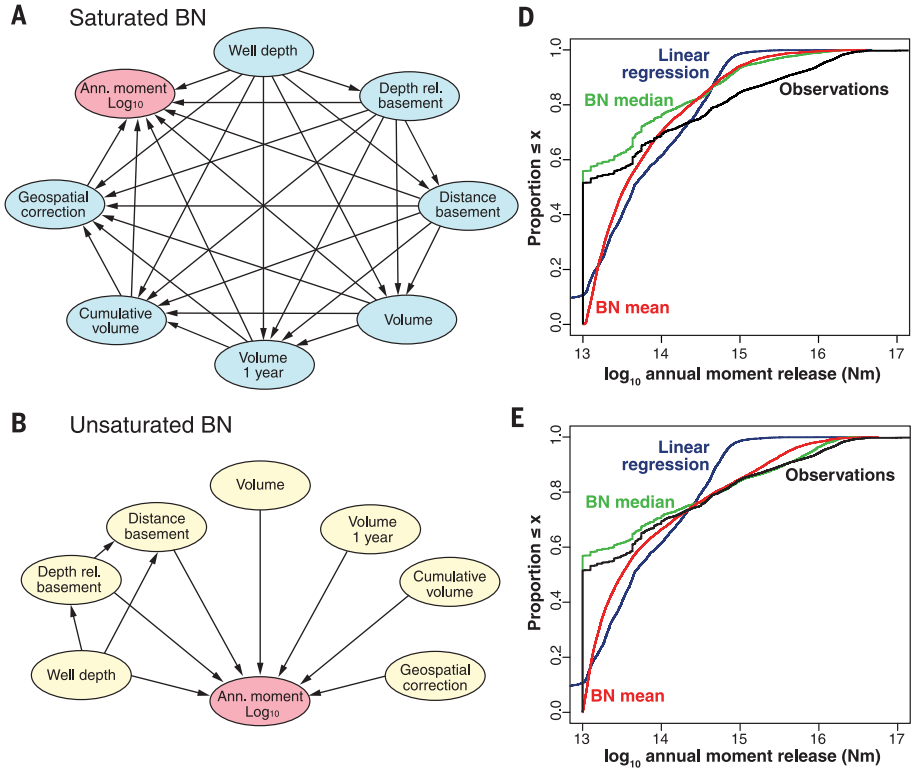


Fig. 1. Wastewater injection and seismicity in Oklahoma. (**A**). Map showing the surface expression of key geologic provinces of Oklahoma (*41*) and class II UIC wells (*35*) (related to oil and gas production). Black polygon defines the "Oklahoma induced seismicity zone" delineated by the USGS (*5*). (**B**) Map showing total injected volume (bbl) for 2011 to 2015 (30-km Gaussian kernel density). Symbols denote locations of earthquakes from 2011 to 2016, including the 5 November 2011 Prague (*7*, *10*, *11*) and 3 September 2016 Pawnee (*18*, *21*, *28*, *44*) events. (**C**) Map showing total seismic moment release from 2011 to 2016 (20-km Gaussian kernel density), with mapped faults in the sedimentary cover including the Nemaha uplift (arrow) also shown (*42*) [this does not include all large-scale basement structures (*17*)]. Yearly injection volumes and seismic moment release are shown in fig. S2.



C Rank correlation coefficients for Log_{10} annual moment release

	Empirical rank correlation	Saturated BN rank correlation	Unsaturated BN rank correlation
Well depth	0.20	0.17	0.17
Depth rel. basement	-0.20	-0.16	-0.16
Dist. basement	-0.34	-0.34	-0.34
Volume	0.18	0.19	0.11
Volume 1 year	0.20	0.22	0.10
Cumulative volume	0.26	0.26	0.15
Geospatial correction	-0.55	-0.52	-0.48

Fig. 2. BN structure, unconditional rank correlation coefficients, and cumulative distribution functions (CDFs) for Test 2. (**A**) The saturated network and (**B**) the unsaturated network. Arcs denote influence between nodes. (**C**) Unconditional rank correlation coefficients for all nodes with log_{10} annual moment release (empirical data and BNs). This correlation is a measure of the strength of influence of each observable on the node of interest (moment release). The BN uses joint normal copulae to represent the empirical data—the closer the values in the BN rank correlation matrix are to the empirical rank correlation matrix, the better the model approximates the data (matrix determinants can also be compared). Full correlation matrices are provided in table S5. (**D** and **E**) CDFs of log_{10} annual moment release for the BN forecast (mean and median estimate), linear regression model, and empirical data for (D) saturated BN and (E) unsaturated BN. This function shows how closely the model output distribution matches the observations. Results are shown for Test 2–learning using 90% of the observations, testing with the remaining 10% (see fig. S11).

The asymmetry of the data also means that a linear model will not perform as well as the BN (e.g., Fig. 2, D and E).

We demonstrate how the BN could be used to quantify the expected impact of UIC regulations with three applied examples. We can (i) reduce injection depth by a fixed amount [to simulate well plug back (*29*) or optimal depth of a new well], (ii) cap injection volume according to imposed regulatory limits (*25*, *43*), and (iii) restrict injection depth in the Oklahoma induced seismicity zone, by enforcing a minimum permissible distance above the basement interface. We intend these examples as demonstrations of how the BN could be used in practice. Before using this approach to inform operational or regulatory practices, further detailed simulations would be required to account for local-scale susceptibility to induced seismicity.

For a subset (10%) of well records (fig. S11), we reduced well depth by 1000 m, and the resulting "test data" were used to produce a set of revised forecasts for annual moment release. We calculated the change in mean moment for each individual case due to the change in depth (Fig. 4). For above-basement wells, raising the well by 1000 m shifted injection further from the basement (shallower) and led to a reduction in seismic moment release. We found the greatest reduction in wells that inject higher volumes (Fig. 4A). Conversely, increasing depth to inject closer to (or into) the basement led to an increase in seismic moment release (fig. S14).

To model the potential impact of regulating volume, we applied a daily limit of 15,000 bbl per well, as imposed by Oklahoma regulators (*25*), similar to the Kansas 16,000 bbl/day limit for areas of "seismic concern" (*43*). We applied this limit to the full data set and used the revised monthly injections to recalculate annual and cumulative volumes for each well (*32*). We then used the BN to simulate the revised annual moment release for all individual monthly well records with reduced injection (Fig. 4B).

As expected, our simulations showed that wells subjected to the largest drop in monthly and annual injected volume yielded the greatest reduction in moment release (Fig. 4B). We found a small number of exceptions in which large changes in mean annual moment were predicted for wells with relatively low annual injection rates (Fig. 4B). These cases largely corresponded to wells injecting close to the basement (fig. S15), where more modest changes in injection volume can yield proportionally larger decreases in moment (Fig. 4). This demonstrates the importance of quantifying the joint effect of operational parameters (depth, cumulative volume and injection rates) on future seismicity levels.

We performed a final set of simulations to quantify the statewide impact of depth restriction, using depth limits of (i) 200 m and (ii) 500 m above basement, wherever this could be practically applied. Using well records for December 2015 to predict annual moment release in the year ahead, we found an indicative (mean) reduction in total annual moment release for the Oklahoma induced seismicity zone of a factor of ~1.4 for the 200-m limit and a factor of ~2.8 for the 500-m limit (*32*).

A distinct advantage of the BN is the ability to update the model and produce revised forecasts as data become available. For example, using newly released (September 2017) OCC injection data (*35*), we tested BN forecasting for 2016, first using the model as developed above (fig. S16) and second, using an updated geospatial correction and moving window for learning, which (by partially accounting for temporal changes in the system) improves performance (figs. S17 and S18) (*32*). The empirical and unconditional BN rank correlations (table S7) computed with the complete updated data set are not considerably different from those for 2011 to 2015 (table S5), lending credence to the basic model.

Our results indicate that seismic moment release in Oklahoma is strongly correlated with injection proximity to crystalline basement (Figs. 3

and 4A and fig. S7), corroborating previous hypotheses (*30*, *31*) but contrary to observations encompassing the CEUS (*4*). We attribute this disparity to scale dependence (*19*). In Oklahoma, the permeability structure of the Arbuckle Group permits downward fluid migration into crystalline basement (*1*, *9*), causing reactivation of optimally oriented strike-slip faults (*14*, *16*–*18*). On a continental scale, lithologic and structural heterogeneity and concomitant stress-field variations could obscure any association between depth and seismicity, as some lithologies are more susceptible to earthquakes (*9*) (fig. S6); e.g., where fracture flow causes localized pore fluid pressure increases, allowing rupture along critically stressed faults (*9*). Regions of relatively high-permeability basement in other parts of the CEUS will conceivably permit greater pore fluid diffusion and pressure release, inhibiting earthquakes (*9*). Induced seismicity is also controlled by local hydrogeologic conditions (*30*) and the location, scale, and orientation of critically stressed faults—all unique aspects of the geotectonic setting. This highlights the importance of local-scale assessments.

Our Oklahoma assessment exploited a 6-year record of fluid injection to develop a comprehensive understanding of the controls on induced seismicity. The two-stage BN quantifies the joint effects of operational parameters and latent spatial features on seismic moment release, facilitates regular model updating, and offers improved forecast performance compared to

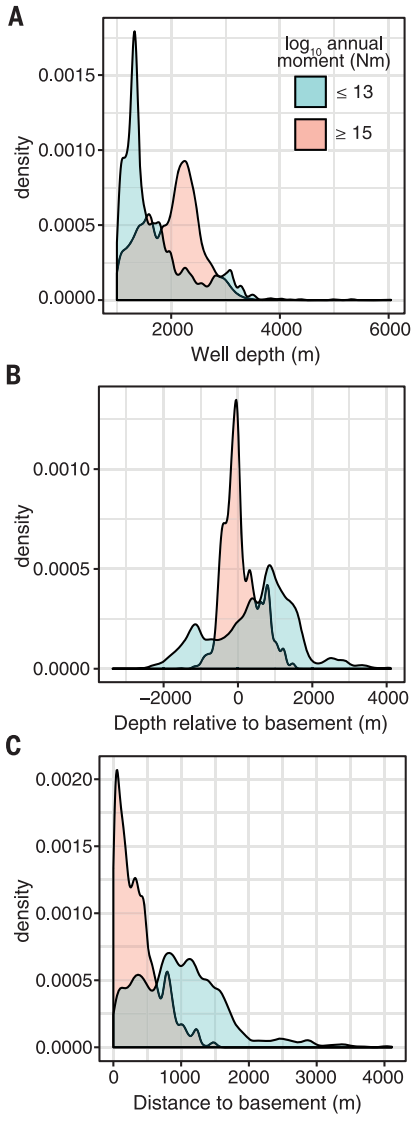


Fig. 3. Probability distributions for well depth and distance to basement using BN inputs for January 2011 to December 2015 (wells ≥1 km deep and injection records >10,000 bbl/month). Probability density for (**A**) well depth below surface; (**B**) depth relative to basement and (**C**) distance to basement; identifying (i) cases with low or no seismic activity (log_{10} moment in 1 year ≤13; shaded blue) and (ii) cases with relatively high seismic activity (log_{10} moment in 1 year ≥15; shaded pink). The distributions in (B) and (C) show that wells drilled closer to the basement are associated with higher seismic moment release. Distributions for injection volume are shown in fig. S7.

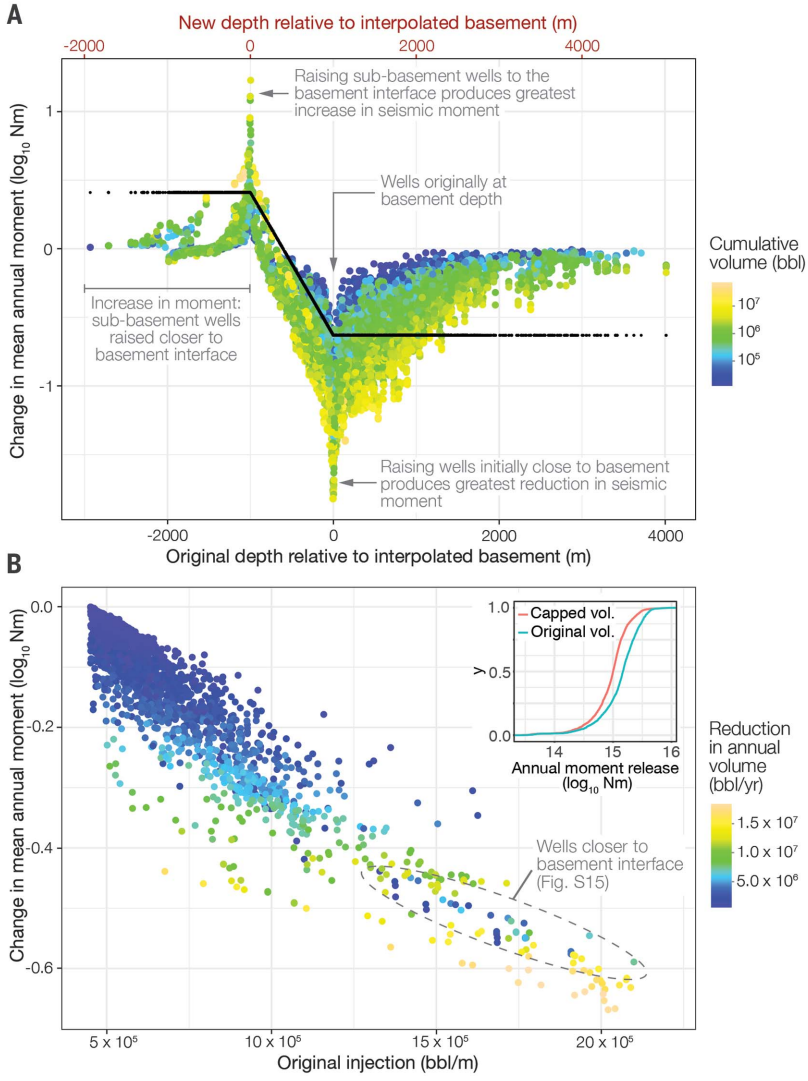


Fig. 4. Simulated impact of raising the injection well level or capping monthly injection volume. (**A**) For a subset of records, well depth was reduced by 1 km and used to generate revised estimates for annual moment release using the BN (colored points) and linear regression models (black points). The plot shows change in annual moment versus original well depth and predicts the greatest reduction in moment release for wells initially closest to the basement interface. The joint effect of cumulative volume is also seen—wells with higher cumulative injections are expected to experience proportionally greater moment reduction. (**B**) Predicted change in mean annual moment resulting from a regulatory limit of 15,000 bbl/day (450,000 bbl/month) versus monthly injection with no cap. This plot shows cases of high injection only (where the limit is applied). Points are colored by total reduction in annual volume to illustrate the importance of maintaining injection limits over time. Inset: CDFs for moment release (BN mean estimates) show the overall effect on all high-volume wells.

traditional linear models. This approach has important implications for statewide assessments and regulatory decision-making under geologic uncertainty [note that the Pawnee (*16*, *18*, *44*), Cushing (*44*), Prague (*11*, *14*), and Fairview (*45*) earthquakes occurred on previously unrecognized faults or splays]. Our framework allows regulatory actions to be evaluated on a rational, quantitative basis in terms of seismic effects. The model could be used to identify evolving vulnerability, complementing zoned seismic hazard forecasts of the U.S. Geological Survey (USGS) (*21*), and provide a basis for targeted management of induced seismicity in Oklahoma and other wastewater disposal regions. We welcome the support of researchers, operators, regulators, and policymakers in updating and extending our model.

REFERENCES AND NOTES

1. C. Langenbruch, M. D. Zoback, *Sci. Adv.* **2**, e1601542 (2016).
2. W. L. Ellsworth, *Science* **341**, 1225942 (2013).
3. K. M. Keranen, M. Weingarten, G. A. Abers, B. A. Bekins, S. Ge, *Science* **345**, 448–451 (2014).
4. M. Weingarten, S. Ge, J. W. Godt, B. A. Bekins, J. L. Rubinstein, *Science* **348**, 1336–1340 (2015).
5. M. D. Petersen *et al*., Incorporating induced seismicity in the 2014 United States national seismic hazard model—results of 2014 workshop and sensitivity studies, Tech. Rep. 20151070, U.S. Geological Survey (2015).
6. M. D. Petersen *et al*., 2016 one-year seismic hazard forecast for the Central and Eastern United States from induced and natural earthquakes, Tech. Rep., U.S. Geological Survey, Reston, VA (2016).
7. F. R. Walsh 3rd, M. D. Zoback, *Sci. Adv.* **1**, e1500195 (2015).
8. P. O. Ogwari, S. P. Horton, *Geofluids* **16**, 954–970 (2016).
9. A. K. Shah, G. R. Keller, *Geophys. Res. Lett.* **44**, 152–161 (2017).
10. N. D. McMahon, R. C. Aster, W. L. Yeck, D. E. McNamara, H. M. Benz, *Geophys. Res. Lett.* **44**, 7149–7158 (2017).
11. K. M. Keranen, H. M. Savage, G. A. Abers, E. S. Cochran, *Geology* **41**, 699–702 (2013).
12. W.-Y. Kim, *J. Geophys. Res. Solid Earth* **118**, 3506–3518 (2013).
13. S. Horton, *Seismol. Res. Lett.* **83**, 250–260 (2012).
14. D. E. McNamara *et al*., *Geophys. Res. Lett.* **42**, 2742–2749 (2015).
15. M. Shirzaei, W. L. Ellsworth, K. F. Tiampo, P. J. González, M. Manga, *Science* **353**, 1416–1419 (2016).
16. E. J. Fielding, S. S. Sangha, D. P. S. Bekaert, S. V. Samsonov, J. C. Chang, *Seismol. Res. Lett.* **88**, 971–982 (2017).
17. M. Schoenball, W. L. Ellsworth, *Seismol. Res. Lett.* **88**, 1252–1258 (2017).
18. X. Chen *et al*., *Sci. Rep.* **7**, 4945 (2017).
19. National Research Council (U.S.), *Induced Seismicity Potential in Energy Technologies* (National Academy of Sciences, Washington, DC, 20001, 2013).
20. Oklahoma Corporation Commission Earthquake Response Summary, 18 Nov 2016; www.occeweb.com/News/2016/11-23-16EARTHQUAKE%20ACTION%20SUMMARY.pdf (2016).
21. M. D. Petersen *et al*., *Seismol. Res. Lett.* **88**, 772–783 (2017).
22. *Sierra Club* v. *Chesapeake Operating LLC et al.*, F. Supp. 3d, WL 1287546 (W.D. Okla. 2017).
23. D. E. McNamara *et al*., *Geophys. Res. Lett.* **42**, 8328–8332 (2015).
24. S. Barrett, The link between hydrofracking, wastewater injection and earthquakes: key issues for re/insurers, Tech. Rep., Swiss Reinsurance (2016). www.swissre.com/library/2016_08_link_hydrofracking_wastewater_injection_earthquakes_key_issues_reinsurers.html.
25. Oklahoma Corporation Commission, Central Oklahoma Regional Volume Reduction Plan and Expansion of Area of Interest; www.occeweb.com/News/2016/03-07-16ADVISORY-AOI,%20VOLUME%20REDUCTION.pdf (2016).
26. M. van der Baan, F. J. Calixto, *Geochem. Geophys. Geosyst.* **18**, 2467–2485 (2017).
27. T. H. W. Goebel, J. I. Walter, K. Murray, E. E. Brodsky, *Sci. Adv.* **3**, e1700441 (2017).
28. A. J. Barbour, J. H. Norbeck, J. L. Rubinstein, *Seismol. Res. Lett.* **88**, 1040–1053 (2017).
29. Oklahoma Corporation Commission directive, http://occeweb.com/News/DIRECTIVE-2.pdf (2015).
30. Y. Zhang *et al*., *Ground Water* **51**, 525–538 (2013).
31. B. S. Currie, M. Brudzinski, R. Skoumal, Geological Society of America Abstracts with Programs (2016), vol. 48.
32. Materials and methods are available as supplementary materials.
33. UNINET software designed by the Risk and Environmental Modeling group, Delft University of Technology, developed by Dan Ababei, Lighttwist Software; www.lighttwist.net/wp/uninet.
34. The R Project for Statistical Computing, https://cran.r-project.org/.
35. Oklahoma Corporation Commission, UIC well data from www.occeweb.com/og/ogdatafiles2.htm.
36. ANSS Comprehensive Earthquake Catalog (ComCat), U.S. Geological Survey, https://earthquake.usgs.gov/data/comcat/; accessed 8 March 2017.
37. A. McGarr, A. J. Barbour, *Geophys. Res. Lett.* **44**, 9330–9336 (2017).
38. QGIS Development Team, QGIS Geographic Information System, https://qgis.org (2016).
39. J. Campbell, J. Weber, Wells drilled to basement in Oklahoma, Tech. Rep. 2006-1, Oklahoma Geological Survey, http://ogs.ou.edu/docs/specialpublications/SP2006-1.pdf (2006).
40. E. K. Franseen, A review of Arbuckle Group strata in Kansas from a sedimentologic perspective: Insights for future research from past and recent studies, Kansas Geological Survey, Open-file Report 1999-49 75, 68 (2000).
41. R. A. Northcutt, J. A. Campbell, in *Geologic Provinces of Oklahoma* (Springer Netherlands, Dordrecht, 1998), pp. 29–37.
42. S. Marsh, A. Holland, Comprehensive fault database and interpretive fault map of Oklahoma, Tech. Rep. OF2-2016, Oklahoma Geological Survey, The University of Oklahoma, Norman, OK 73019 (2016).
43. Kansas Corporation Commission, Second Order Reducing Saltwater Injection Rates; www.kcc.state.ks.us/pi/press/16-11.htm (2016).
44. W. L. Yeck *et al*., *Geophys. Res. Lett.* **44**, 711–717 (2017).
45. W. L. Yeck *et al*., *Geophys. Res. Lett.* **43**, 10,198–10,207 (2016).

ACKNOWLEDGMENTS

T.H. and W.A. were supported in part by the CREDIBLE consortium (UK Natural Environment Research Council grant NE/J 017299/1). W.A. received support in part from Risk Management Solutions (RMS) for a conceptual study associated with induced seismicity attribution. D. Ababei at LightTwist Software provided invaluable assistance with UNINET software and scripting (available at www.lighttwist.net/wp/uninet). The study used UIC injection well data obtained from the Oklahoma Corporation Commission (www.occeweb.com/) and earthquake data from the USGS ANSS Comprehensive Catalog (https://earthquake.usgs.gov/data/comcat/). Input data are provided in the supplementary materials. We acknowledge helpful discussions with R. Muir-Wood (RMS, London) and R. J. Briggs (formerly Praedicat Inc., CA). W.A. and T.H. conceived the research. T.H. performed the modeling and analyzed and refined the BN. W.A. provided seismological interpretation. T.G. performed geologic interpretation and coordinated and wrote the paper. R.C. provided support with statistical analysis and performed additional UNINET tests. All authors contributed to the interpretation, discussion, and writing.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/359/6381/1251/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S7
Data File S1
References (*46*–*54*)

26 August 2017; resubmitted 20 September 2017
Accepted 19 January 2018
Published online 1 February 2018
10.1126/science.aap7911

FISHERIES

Protecting marine mammals, turtles, and birds by rebuilding global fisheries

Matthew G. Burgess,[1,2]*† Grant R. McDermott,[3,1]† Brandon Owashi,[1,2] Lindsey E. Peavey Reeves,[1,4] Tyler Clavelle,[1,2] Daniel Ovando,[1,2] Bryan P. Wallace,[5,6] Rebecca L. Lewison,[7] Steven D. Gaines,[1,2] Christopher Costello[1,2]

Reductions in global fishing pressure are needed to end overfishing of target species and maximize the value of fisheries. We ask whether such reductions would also be sufficient to protect non–target species threatened as bycatch. We compare changes in fishing pressure needed to maximize profits from 4713 target fish stocks—accounting for >75% of global catch—to changes in fishing pressure needed to reverse ongoing declines of 20 marine mammal, sea turtle, and seabird populations threatened as bycatch. We project that maximizing fishery profits would halt or reverse declines of approximately half of these threatened populations. Recovering the other populations would require substantially greater effort reductions or targeting improvements. Improving commercial fishery management could thus yield important collateral benefits for threatened bycatch species globally.

Fisheries employ 260 million people and fish are a primary animal protein source for roughly 40% of the world's population (*1*). Recent studies suggest that more than half of the world's fisheries are overfishing (*2*), and rebuilding these fisheries could increase global fishing yields by ~15% and profits by ~80% (*2*, *3*). Fisheries also affect many protected, non–target species through bycatch (incidental capture), including ecologically important and charismatic megafauna such as marine mammals, sea turtles, seabirds, and sharks (*4*). Some of these bycatch species, such as Mexico's vaquita porpoise (*Phocoena sinus*) and New Zealand's Hector's dolphin subspecies (Māui dolphin, *Cephalorhynchus hectori maui*), face imminent extinction (*5*, *6*). For these reasons, ending overfishing and protecting threatened bycatch species are two of the main goals of modern marine conservation efforts.

At first glance, sustaining high fishery profits and yields can seem in conflict with bycatch species conservation. Unless targeting can become more selective through changing fishing technology or practices, reducing bycatch requires reducing target stock catch. However, because rebuilding overfished target stocks requires reducing fishing effort, bycatch populations should also benefit. Indeed, regions with the most severe bycatch—coastal fisheries of the developing world and, to a lesser extent, high-seas fisheries (*4*)—also experience some of the most severe overfishing (*2*) (Fig. 1 and fig. S1).

We quantify the trade-offs globally between protecting bycatch species and meeting economic fisheries objectives. To do this, we compare estimates of the changes in fishing pressure needed to maximize long-term profits [termed "maximum economic yield" (MEY)] for 4713 fish stocks, accounting for >75% of global catch (*2*), to the changes in bycatch mortality needed to reverse ongoing population declines of 20 populations substantially affected by fisheries bycatch, for which sufficient published information is available to calculate the reductions in mortality needed to prevent further declines (materials and methods and table S1).

Our sample includes 9 of 26 marine mammal populations, 6 of 8 sea turtle populations or species, and 3 of 22 seabird populations that the International Union for Conservation of Nature (IUCN) identifies as threatened, declining, and having bycatch as a primary threat (*7*). We also include the Northwest Atlantic loggerhead turtle (*Caretta caretta*) population, but it is not listed as threatened by the IUCN owing to uncertainty as to whether it remains in decline (*7*) (materials and methods). The IUCN last assessed olive ridley turtle (*Lepidochelys olivacea*) populations jointly (*7*), and we include two of these in our analysis (materials and methods). We restrict our analysis to marine mammals, sea turtles, and seabirds, because they are rarely retained or commercially valuable (*4*). However, future work could use similar methods to consider sharks, rays, and other taxa retained as both target and non–target catch (*8*).

Accounting for multiple uncertainties, we ask how likely it is that solely managing all target fisheries to MEY would reduce bycatch mortality sufficiently to halt each bycatch population's decline. We further ask how much long-term profit would need to be foregone, or how much more selective targeting would need to become, to ensure that each bycatch population's decline was halted. In other words, we assess whether there is currently a trade-off between maximizing long-term profit and halting each bycatch population's decline, and how severe the trade-off is, if one exists. In the supplementary materials, we explore trade-offs relative to maximum long-term catch [termed "maximum sustainable yield" (MSY)] and obtain results similar to those for MEY (figs. S2 to S4).

We assume that each population's annual rate of change (denoted Δ, e.g., $\Delta = -0.05$ year^{-1} implies a 5% annual decline in abundance) can be approximately expressed as (materials and methods)

$$\Delta = \Delta_n - F_e \quad (1)$$

Here, Δ_n denotes the annual rate of change in abundance that would occur if there were no bycatch, and F_e denotes the "effective" annual bycatch mortality rate—the fraction of the population's total reproductive value removed by bycatch annually. Derived from age-structured

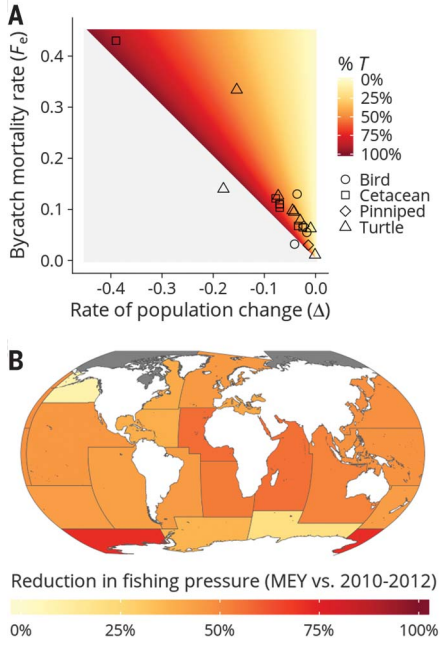


Fig. 1. Reductions in fishing pressure needed to meet profit and bycatch objectives. (**A**) Population decline (Δ) and bycatch mortality (F_e) rates, and reductions in bycatch mortality needed to halt population declines (%*T*), for 20 bycatch populations. (**B**) Projections (*2*) of average reductions in target stock fishing mortality (weighted by 2010 to 2012 expenditure), by FAO Major Fishing Area (except Arctic Sea, gray), under the MEY scenario [same color bar as in (A)].

[1]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106, USA. [2]Marine Science Institute, University of California, Santa Barbara, CA 93106, USA. [3]Department of Economics, University of Oregon, Eugene, OR 97403, USA. [4]Channel Islands National Marine Sanctuary, Santa Barbara, CA 93106, USA. [5]Conservation Science Partners, Inc., Ft. Collins, CO 80524, USA. [6]Nicholas School of Environment, Duke University, Beaufort, NC 28506, USA. [7]Biology Department, San Diego State University, San Diego, CA 92182, USA.
*Corresponding author. Email: mburgess@ucsb.edu
†These authors contributed equally to this work.

population models, reproductive value measures the relative contributions of individuals in each age group to overall population growth [e.g., see (*9*, *10*)]. We use this measure to standardize bycatch of different ages across fisheries, since fisheries primarily causing bycatch of breeding adults tend to have much larger population impacts than fisheries causing bycatch of small juveniles [e.g., (*10*)]. To keep the units of Eq. 1 consistent, we also measure Δ and Δ_n in reproductive-value units where possible, i.e., where a published age-structured assessment is available [e.g., (*10*)]. Otherwise, we assume that abundance and mortality trends measured in individual units reflect trends in reproductive value.

From Eq. 1, we calculate the percentage (denoted $\%T$) by which each bycatch population's mortality rate, F_e, would have to decrease to halt its population decline (i.e., $\Delta = 0$), if all other mortality sources remained constant:

$$0 = \Delta_n - \left(1 - \frac{\%T}{100}\right) F_e \qquad (2A)$$

$$\%T = 100\left(1 - \frac{\Delta_n}{F_e}\right) = 100\left(-\frac{\Delta}{F_e}\right)$$

$$= 100\left(\frac{\Delta}{\Delta - \Delta_n}\right) \qquad (2B)$$

Figure 2 illustrates the steps of our analysis for each bycatch population, using the relatively data-rich Northwest Atlantic loggerhead turtle as an example. Materials and methods and table S1 describe our analysis for all populations. First, we obtain point estimates and approximate uncertainty for two of Δ, Δ_n, and F_e, from the literature. From these, we calculate point estimates and distributions for $\%T$ using Eq. 2B (Figs. 1A and 2B and fig. S2). We also use information from the literature to infer which target fisheries may be contributing to bycatch mortality (Fig. 2, A and C).

We then perform a Monte Carlo simulation that defines 1000 different "states of the world." In each state, we randomly draw a value of $\%T$ from its distribution (Fig. 2B), as well as an allocation of bycatch mortality among target fisheries from the set of identified target fisheries. We weight allocation probabilities by the fisheries' relative efforts, measured by 2010 to 2012 fishing expenditures (Fig. 2C). We assume that bycatch mortality (F_e) responds proportionally to changes in target stock mortality. Thus, the percentage reduction in bycatch mortality in a given state of the world is equal to the average change in sampled target stock fishing mortality at MEY relative to 2010 to 2012 rates (*2*).

In some states of the world, the projected reduction in bycatch mortality at MEY is greater than $\%T$. The bycatch population's decline is thus already halted under economically optimal conditions and current targeting, implying that zero cost or targeting improvement is required. In states of the world where the projected reduction in bycatch mortality is less than $\%T$, we calculate the total cost of reducing bycatch mortality by $\%T$ according to principles of economic efficiency (i.e., additional reductions in target stock mortality beyond MEY are ordered in ascending order of marginal cost). We calculate the required targeting improvement as the additional percentage change in bycatch mortality required beyond MEY. When $\%T \geq 100$, fishing or bycatch must cease entirely, so the required cost or targeting improvement is 100%. Our Monte Carlo analysis thus yields distributions of $\%T$ and expected reductions in bycatch mortality (Fig. 2D), as well as costs (Fig. 2E) and targeting improvements (Fig. 2F) required to halt the decline of each bycatch population.

In 95% of simulated states of the world, halting the declines of 7 to 13 populations (median 10) is fully accomplished by managing target stocks to MEY, or requires only minor loss in total profit (<5%) (Figs. 3 and 4 and figs. S2 and S3). In >50% of states of the world, this includes seven turtles, one pinniped, one cetacean, and two birds (Fig. 3). Required costs are often substantial (>50%) for the remaining populations. Even eliminating bycatch completely is insufficient to halt declines of one turtle and one bird in most states of the world, owing to other mortality sources. Targeting improvements required for recovery are always slightly larger than required profit losses (Figs. 3 and 4 and fig. S4), because long-term profits are insensitive to small deviations from the exactly optimal fishing pressure [(*11*) demonstrates

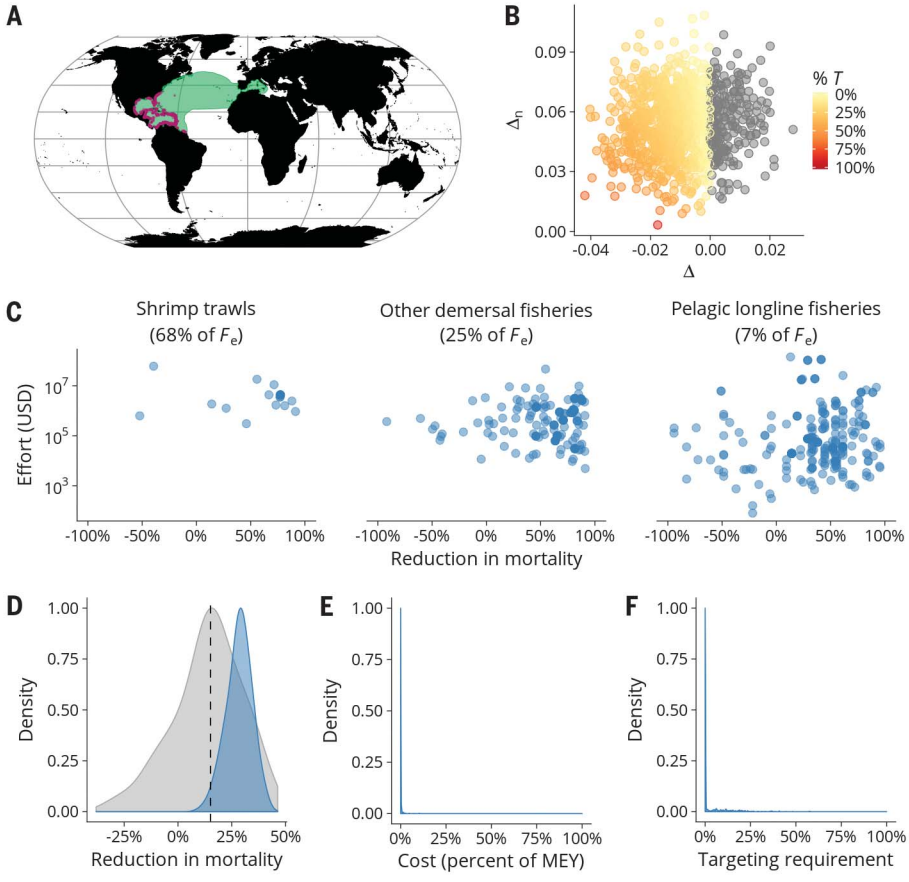


Fig. 2. Analysis. The steps of our analysis are illustrated using the Northwest Atlantic loggerhead turtle population. From the literature, we determine (**A**) its geographic range (green, nesting sites in purple) (*18*, *19*), (**B**) the joint distribution of its rates of abundance change currently (Δ) and without bycatch (Δ_n)—from which we sample the mortality reduction needed ($\%T$) (dots, gray indicates $\%T < 0$)—and (**C**) the target stock groups implicated in bycatch and their relative contributions to mortality. (C) The effort (measured as average 2010 to 2012 fishing expenditures) and projected reduction in mortality under MEY in each target fishery (*2*). From this information, we use Monte Carlo simulation to estimate distributions on (**D**) $\%T$ (gray) and the percentage reduction in bycatch mortality under MEY (blue), (**E**) the fraction of cumulative MEY that would need to be forgone, or (**F**) the improvement in targeting ("targeting requirement") needed to halt the population's decline.

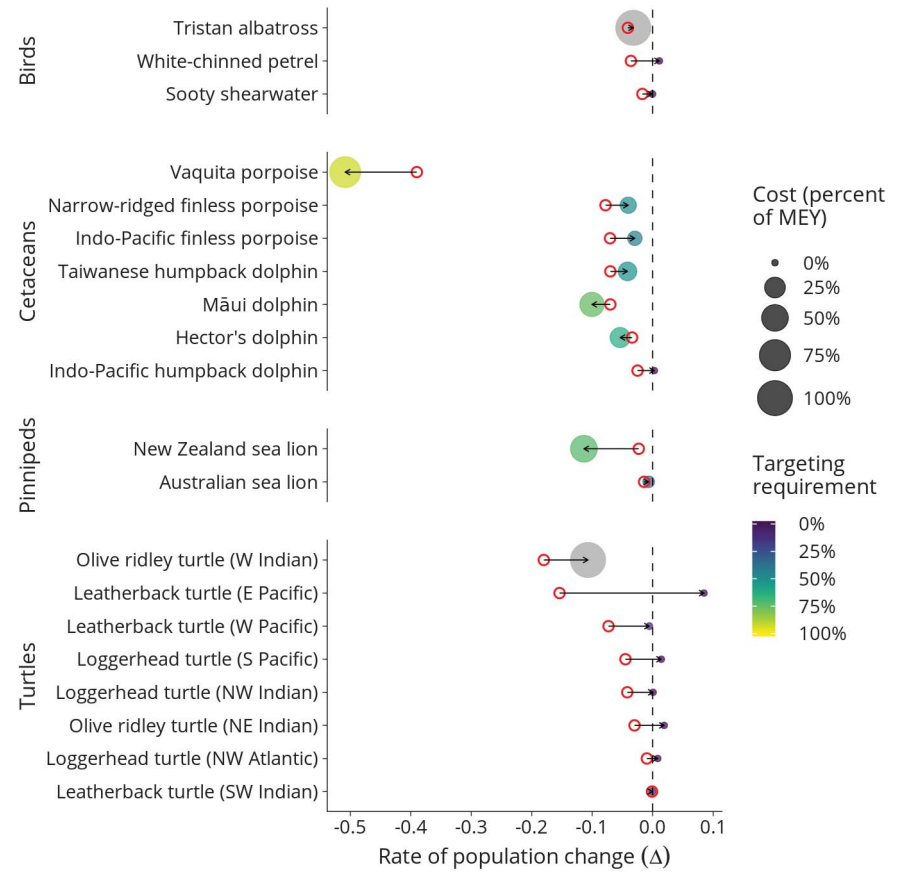


Fig. 3. Trade-offs. For each bycatch population, we compare the median projected rates of population change (Δ) under current conditions (open red circles) and with all target stocks fished at the profit-maximizing rate (denoted F_{MEY}) (filled colored circles). Arrows illustrate the effect of transitioning to MEY. Bycatch populations whose target stocks are currently fished at lower rates than F_{MEY} on average experience greater mortality at MEY (left-facing arrows), and vice versa (right-facing arrows). Sizes and colors of filled circles respectively represent median required costs (as a percentage of MEY) and median targeting requirements (percentage reduction in bycatch mortality, starting from MEY). Gray color indicates that the decline would continue even if bycatch were completely eliminated.

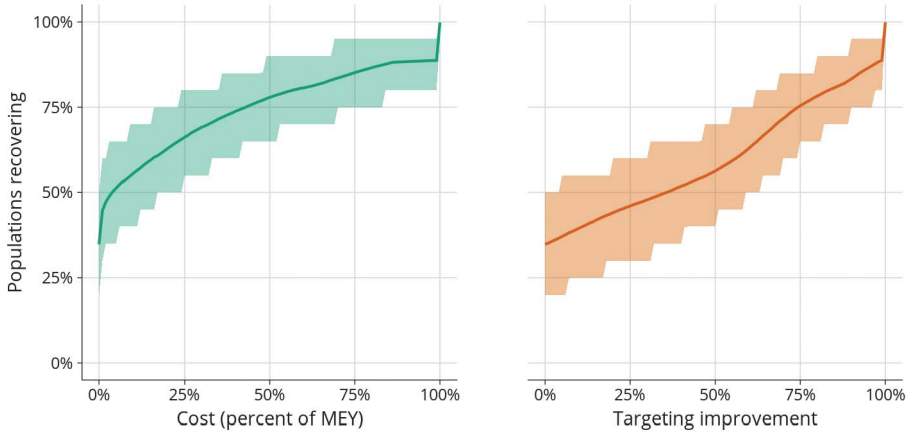


Fig. 4. Summary. At each cost level (% of MEY, green) and targeting improvement level (%, orange), the fraction of populations increasing in abundance is shown. Lines represent means; shaded regions represent 95% of states of the world.

this principle for catch]. Efficiently ordering reductions in fishing pressure among fisheries to minimize costs enhances this insensitivity (fig. S5).

Given the data limitations associated with both fisheries bycatch (*4*) and assessing the status of target fisheries lacking formal stock assessments (*2*), we urge cautious interpretation of our results for any specific bycatch population, some of which have a large uncertainty (figs. S2 to S4). Each population would benefit from a locally tailored follow-up study. However, several broader conclusions are robust to both these uncertainties and a wide range of sensitivity analyses (materials and methods and figs. S6 to S8).

First, our results suggest that recovery of approximately half of the world's marine mammals, turtles, and birds most threatened by fishery bycatch could be achieved as a collateral benefit to ending overfishing of target stocks. Given that achieving MEY and MSY would respectively require 52% and 33% reductions in fishing mortality for the median target stock (*2*), it makes sense that this alone could allow many threatened bycatch populations to recover. Marine turtles and cetaceans in developing-world waters stand to benefit in particular (Fig. 3). These populations are caught in coastal trawl and gillnet fisheries targeting shrimp and finfish (*12*, *13*), which are estimated (*2*) to need the greatest average reductions in fishing effort to achieve MEY (Fig. 1B and fig. S1). However, MEY reference points for shrimp fisheries may need to be refined to account for their highly variable, environmentally driven recruitment (*14*).

Second, we project that recovery of some bycatch populations would require substantial profit losses or targeting improvements. These bycatch populations tend to be caught in fisheries whose target stocks are already sustainably harvested [e.g., the New Zealand sea lion (*Phocarctos hookeri*)], require total or near-total elimination of bycatch to persist (e.g., the vaquita porpoise), or both (e.g., the Māui dolphin). Such bycatch populations should thus receive high priority in efforts to improve fishery targeting. Recent progress in bycatch mitigation efforts suggests that substantial targeting improvements are achievable (*15*). In many cases, non–fishery-related threats to these populations will also need to be addressed.

Ending overfishing can benefit fisheries and fishers. Our results suggest that it can also contribute substantially to reducing global bycatch of threatened species. Of course, ending overfishing is not easy. In many places, it will require new institutions and infrastructure, combined with increases in science and enforcement capacity (*16*). Substantially reducing fishing pressure can create short-term hardship for fishing communities until stocks recover (*2*). Rebuilding target stocks may also have important—sometimes negative—indirect effects on bycatch populations, and vice versa [e.g., via competition for prey (*17*)]. These issues deserve attention in future studies. Nonetheless, our conclusions enhance the motivation for continued global progress in sustainable fisheries reforms.

REFERENCES AND NOTES

1. Food and Agricultural Organization of the United Nations (FAO), *The State of World Fisheries and Aquaculture* (FAO, 2016).
2. C. Costello *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5125–5129 (2016).
3. U. R. Sumaila *et al*., *PLOS ONE* **7**, e40542 (2012).
4. R. L. Lewison *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5271–5276 (2014).
5. B. L. Taylor *et al*., *Cons. Lett*. 10.1111/conl.12331 (2016)
6. C. Pala, *Science* **355**, 559 (2017).
7. International Union for Conservation of Nature (IUCN), *The IUCN Red List of Threatened Species*, 2015; www.iucnredlist.org (accessed 13 December 2017).
8. K. C. James, R. L. Lewison, P. W. Dillingham, K. A. Curtis, J. E. Moore, *Environ. Conserv.* **43**, 3–12 (2016).
9. B. P. Wallace, S. S. Heppell, R. L. Lewison, S. Kelez, L. B. Crowder, *J. Appl. Ecol.* **45**, 1076–1085 (2008).
10. A. B. Bolten *et al*., *Front. Ecol. Environ.* **9**, 295–301 (2011).
11. R. Hilborn, *Mar. Policy* **34**, 193–196 (2010).
12. R. L. Lewison, L. B. Crowder, *Conserv. Biol.* **21**, 79–86 (2007).
13. N. M. Young, S. Iudicello, Worldwide bycatch of cetaceans: An evaluation of the most significant threats to cetaceans, the affected species and the geographic areas of high risk, and the recommended actions from various independent institutions. NOAA Technical Memorandum (NMFS-OPR-36), National Marine Fisheries Service, 2007.
14. C. S. Szuwalski, K. A. Vert-Pre, A. E. Punt, T. A. Branch, R. Hilborn, *Fish Fish.* **16**, 633–648 (2015).
15. S. J. Hall, B. M. Mainprize, *Fish Fish.* **6**, 134–155 (2005).
16. M. C. Melnychuk, E. Peterson, M. Elliott, R. Hilborn, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 178–183 (2017).
17. J. A. Estes, M. Heithaus, D. J. McCauley, D. B. Rasher, B. Worm, *Annu. Rev. Environ. Resour.* **41**, 83–116 (2016).
18. C. Y. Kot *et al*., The State of The World's Sea Turtles online database: Data provided by the SWOT team and hosted on OBIS-SEAMAP (Oceanic Society, IUCN Marine Turtle Specialist Group [MTSG], and Marine Geospatial Ecology Lab, Duke University, 2015); http://seamap.env.duke.edu/swot (accessed 13 December 2017).
19. P. N. Halpin *et al*., *Oceanography (Wash. D.C.)* **22**, 104–115 (2009).

ACKNOWLEDGMENTS

We thank J. Moore, R. Reeves, D. Bradley, and the reviewers for helpful comments and C. Kot for assistance with nesting data. **Funding:** We acknowledge funding from the Waitt Foundation, Ocean Conservancy, the NASA Earth Science Division–Applied Sciences Program (NNH12ZDA001N-COF to R.L.L.), and an NSF Graduate Research Fellowship (to L.E.P.R.). **Author contributions:** M.G.B. and C.C. conceived the study. M.G.B., G.R.M., and C.C. designed the study with input from all authors. G.R.M., M.G.B., and B.O. performed the analysis. M.G.B. wrote the paper with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and code used in our analysis are available from https://doi.org/10.5281/zenodo.1188538. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1255/suppl/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References (*20–110*)

20 July 2017; accepted 26 January 2018
10.1126/science.aao4248

MALARIA

GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing

Michael Filarsky,[1,2] Sabine A. Fraschka,[3] Igor Niederwieser,[1,2] Nicolas M. B. Brancucci,[1,2] Eilidh Carrington,[1,2] Elvira Carrió,[1,2] Suzette Moes,[4] Paul Jenoe,[4] Richárd Bártfai,[3] Till S. Voss[1,2]*

Malaria is caused by *Plasmodium* parasites that proliferate in the bloodstream. During each replication cycle, some parasites differentiate into gametocytes, the only forms able to infect the mosquito vector and transmit malaria. Sexual commitment is triggered by activation of AP2-G, the master transcriptional regulator of gametocytogenesis. Heterochromatin protein 1 (HP1)–dependent silencing of *ap2-g* prevents sexual conversion in proliferating parasites. In this study, we identified *Plasmodium falciparum* gametocyte development 1 (GDV1) as an upstream activator of sexual commitment. We found that GDV1 targeted heterochromatin and triggered HP1 eviction, thus derepressing *ap2-g*. Expression of GDV1 was responsive to environmental triggers of sexual conversion and controlled via a *gdv1* antisense RNA. Hence, GDV1 appears to act as an effector protein that induces sexual differentiation by antagonizing HP1-dependent gene silencing.

Heterochromatin protein 1 (HP1) is a conserved regulator of heterochromatin formation, heritable gene silencing, and variegated gene expression (*1*). In *Plasmodium falciparum*, HP1-dependent clonally variant expression allows parasites to adapt rapidly to environmental challenges encountered during infection (*2*–*4*). For example, immune evasion via antigenic variation of PfEMP1 is the hallmark of *Plasmodium* survival. Other processes, such as expression of red blood cell (RBC) invasion ligands or nutrient transporters, are similarly regulated in this parasite (*4*). Most clonally variant genes cluster in subtelomeric domains, but some also occur in internal heterochromatic regions of chromosomes. In addition, HP1 forms microdomains at some euchromatic genes (*2*). One of these encodes the transcription factor AP2-G, which is required for sexual conversion and differentiation (*2*, *5*–*7*). HP1-dependent regulation of *ap2-g* controls the rate at which parasites commit to sexual differentiation (*7*).

To explore the mechanisms regulating HP1 occupancy in *P. falciparum*, we identified HP1-interacting proteins by liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis of native HP1 complexes that were purified by coimmunoprecipitation (co-IP) from parasites expressing green fluorescent protein (GFP)–tagged HP1 (*7*) (Fig. 1A and table S3). We consistently observed GDV1 among the potential HP1 interaction partners (table S1). GDV1 is a nuclear protein implicated in sexual commitment and early gametocytogenesis, but its exact function remains unknown (*8*). We therefore created a parasite line for the conditional expression of fluorescently labeled ectopic GDV1 (designated GDV1-GFP-DD) (Fig. 1B). Proteins tagged with the destabilization domain (DD) of the immunophilin protein-folding chaperone FK506 binding protein (FKBP) are proteolytically degraded unless cells are cultured in the presence of Shield-1, a small-molecule ligand stabilizer (*9*, *10*). Thus, GDV1-GFP-DD is barely detectable in parasites cultured in the absence of Shield-1 (designated 3D7/GDV1-GFP-DD^{OFF} parasites), but its expression is markedly induced in parasites grown in the presence of Shield-1 (3D7/GDV1-GFP-DD^{ON} parasites) (Fig. 1, B and C). In agreement with the co-IP results, GDV1-GFP-DD colocalized with HP1 at the nuclear periphery (Fig. 1C and fig. S1). Furthermore, we found that recombinant HP1 and GDV1 formed a complex (fig. S1) and that HP1 copurified with GDV1-GFP-DD in reverse co-IPs (Fig. 1D and tables S2 and S4). The chromodomain-helicase-DNA-binding protein 1 (CHD1) and a protein of unknown function (PF3D7_1451200) also consistently copurified with both HP1 and GDV1-GFP-DD (tables S1 and S2). Given that CHD1 plays important roles in cell fate decision and heterochromatin remodeling in other organisms (*11*, *12*) and that GDV1 is implicated in gametocytogenesis (*8*), this putative regulatory complex may function in activating sexual commitment.

Malaria parasites proliferate by iterative rounds of intraerythrocytic replication through schizogony, merozoite release, and RBC reinvasion. The decision to enter gametocytogenesis is made in the cell cycle before sexual differentiation; sexually committed schizonts release merozoites that invade RBCs and differentiate all into either female or male gametocytes (*13*, *14*) (Fig. 2A). To test whether GDV1 triggers sexual commitment, we divided samples of 3D7/GDV1-GFP-DD^{OFF} parasites and cultured them in the absence or presence of Shield-1. After reinvasion, stage I gametocytes were quantified by immunofluorescence assays (IFAs) with antibodies to the gametocyte marker Pfs16 (*15*) (anti-Pfs16). The 3D7/GDV1-GFP-DD^{ON} population displayed a sexual conversion rate of 57.2% (±10.0% SD), compared with 11.0% (±2.4% SD) for 3D7/GDV1-GFP-DD^{OFF}

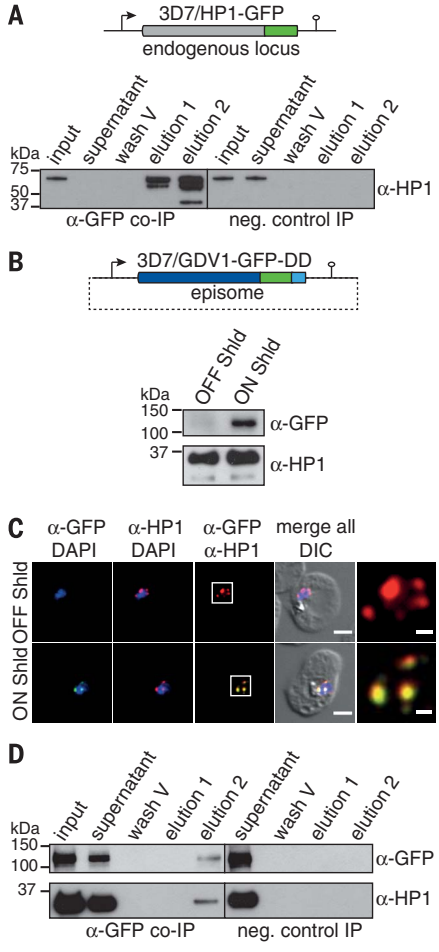


Fig. 1. GDV1 interacts with HP1. (**A**) Endogenous *hp1* locus in 3D7/HP1-GFP parasites and Western blots with antibodies to HP1 (anti-HP1 [α-HP1]) of the anti-HP1-GFP co-IP and negative (neg.) control samples. Results are representative of three biological replicates. α-GFP, antibodies to GFP (anti-GFP). (**B**) GDV1-GFP-DD expression plasmid and anti-GFP Western blots of 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} parasites. Anti-HP1 served as a loading control. (**C**) GDV1-GFP-DD/HP1 colocalization IFAs with 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} trophozoites (24 to 32 hpi). DAPI, 4′,6-diamidino-2-phenylindole; DIC, differential interference contrast; Shld, Shield-1. Scale bars, 2.5 μm (0.5 μm for the magnified views in the rightmost images). Results are representative of three biological replicates. (**D**) Anti-GFP and anti-HP1 Western blots of the anti-GDV1-GFP-DD co-IP and negative control samples. Results are representative of three biological replicates.

[1]Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland. [2]University of Basel, 4003 Basel, Switzerland. [3]Department of Molecular Biology, Radboud University, 6525 GA Nijmegen, Netherlands. [4]Biozentrum, University of Basel, 4056 Basel, Switzerland.
*Corresponding author. Email: till.voss@unibas.ch

parasites, and these gametocytes differentiated normally into both male and female gametocytes and showed a typical female-biased sex ratio (Fig. 2B and fig. S2). Moreover, Shield-1 titration revealed a positive correlation between ectopic GDV1-GFP-DD expression levels and sexual conversion rates (Fig. 2C). To test whether endogenous GDV1 expression levels similarly correlate with gametocyte conversion, we used CRISPR-Cas9–based gene editing to append a triple hemagglutinin (3×HA) tag to the N terminus of GDV1 (yielding 3D7/3×HA-GDV1 parasites) (fig. S3). Endogenous 3×HA-GDV1 colocalized with HP1 as expected (Fig. 2D and fig. S3) but was expressed only in some parasites. We next quantified 3×HA-GDV1 expression under conditions that either suppress or favor sexual conversion. To this end, we made use of the recent discovery of choline as an inhibitor of sexual commitment (*16*). 3D7/3×HA-GDV1 parasites cultured in the presence or absence of 2 mM choline displayed sexual commitment rates of 1.8% (±0.3% SD) or 30.9% (±3.8% SD), respectively (Fig. 2E). Parasites cultured in the absence of choline showed

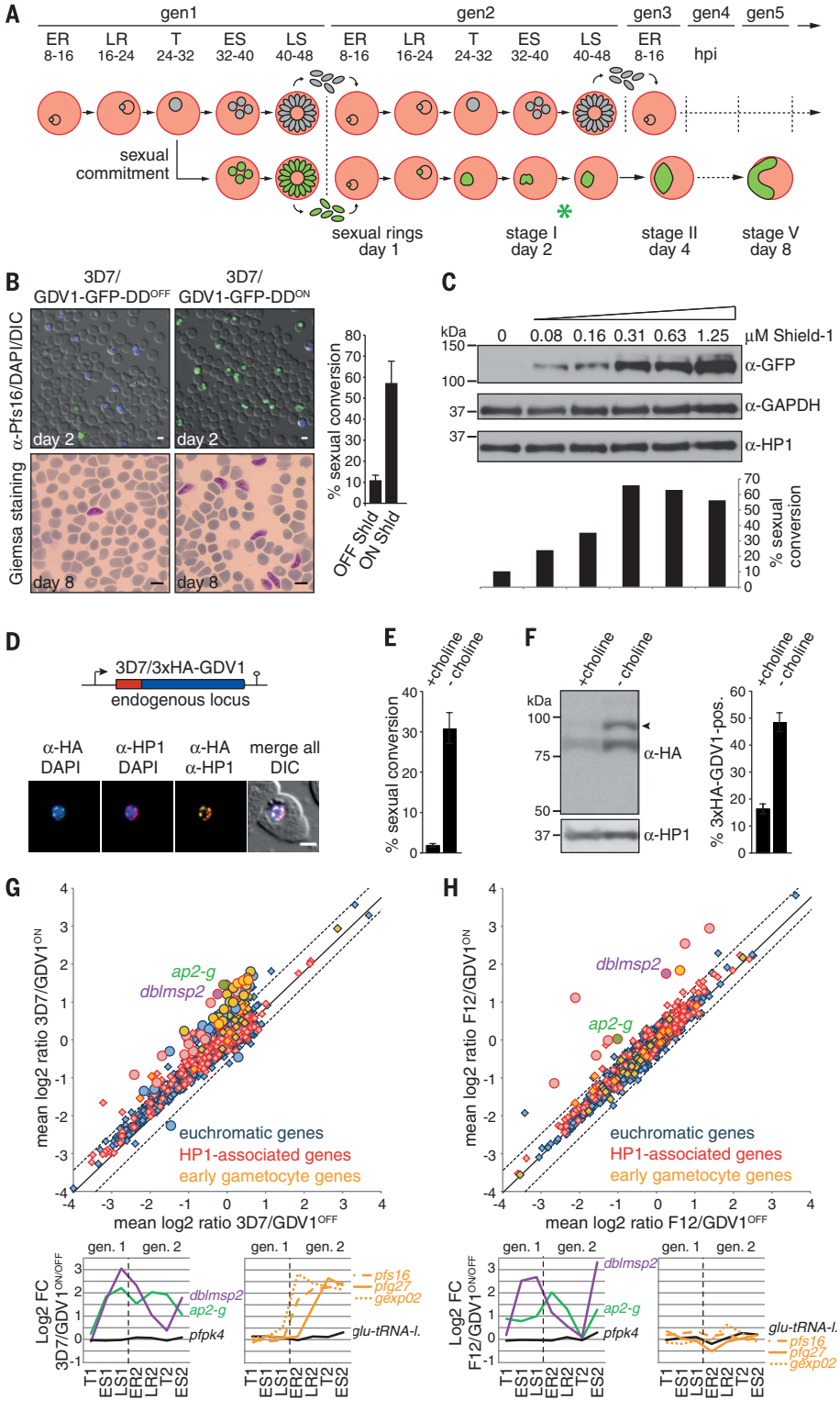


Fig. 2. GDV1 induces sexual commitment and differentiation. (**A**) Schematic illustrating the iterative cycles of schizogony and RBC reinvasion (top) or sexual commitment, RBC reinvasion, and gametocyte differentiation (bottom). ER and LR, early and late ring stages; T, trophozoites; ES and LS, early and late schizonts; gen, generation; *, time point of anti-Pfs16 (α-Pfs16) IFAs. (**B**) Top panels show results from anti-Pfs16 IFAs identifying stage I gametocytes. Quantification of Pfs16-positive parasites is shown at the right [results are the means for three biological replicates (200 infected RBCs counted per sample); error bars indicate SD]. Bottom panels show Giemsa-stained blood smears revealing stage V gametocytes. Scale bars, 5 µm. (**C**) Western blot showing GDV1-GFP-DD expression in the presence of increasing Shield-1 concentrations. Antibodies to glyceraldehyde-3-phosphate dehydrogenase (α-GAPDH) and HP1 served as loading controls. Percentages of Pfs16-positive parasites are shown at the bottom (400 infected RBCs were counted per sample). (**D**) Endogenous *gdv1* locus in 3D7/3×HA-GDV1 parasites and 3×HA-GDV1/HP1 colocalization IFAs in trophozoites (24 to 32 hpi). Scale bar, 2.5 µm. (**E**) Sexual conversion rates in 3D7/3×HA-GDV1 parasites cultured in the presence or absence of choline [results are the means for three biological replicates (>190 infected RBCs counted per sample); error bars indicate SD]. (**F**) Western blot showing 3×HA-GDV1 expression levels in 3D7/3×HA-GDV1 parasites cultured in the presence or absence of choline. The arrowhead indicates the intact full-length protein. Anti-HP1 served as a loading control. Percentages of 3×HA-GDV1-positive (3×HA-GDV1-pos.) parasites in the presence or absence of choline are shown on the right [results are the means for three biological replicates (>100 infected RBCs counted per sample); error bars indicate SD]. (**G** and **H**) Comparison of mean expression levels for all genes in 3D7/GDV1-GFP-DD^{ON} parasites versus 3D7/GDV1-GFP-DD^{OFF} parasites (G) and F12/GDV1-GFP-DD^{ON} parasites versus F12/GDV1-GFP-DD^{OFF} parasites (H). Significantly deregulated genes are indicated by circles [mean fold change (FC) cutoff, >1.5; *q*-value (false discovery rate) cutoff, <0.15]. Known early gametocyte markers (*7*, *8*, *17*) are labeled orange. Line graphs show fold changes in expression across seven consecutive time points. *pfs16*, *pfg27*, and *gexp02*, early gametocyte markers (*15*, *22*, *32*); *pk4* (PF3D7_0628200) and *glu-tRNA-l*. (PF3D7_1331700), control genes (*7*).

markedly increased 3×HA-GDV1 expression levels (Fig. 2F). This increase was accounted for by a higher proportion of 3×HA-GDV1–positive cells [48.6% (±3.4% SD) in the absence of choline compared with 16.4% (±1.8% SD) in the presence of choline] (Fig. 2F) and higher 3×HA-GDV1 expression levels in individual 3×HA-GDV1–positive parasites (fig. S3). Together, these results show that GDV1 activates sexual conversion in a dose-dependent manner and that endogenous GDV1 expression can be induced by environmental signals triggering sexual commitment.

We next performed comparative transcriptome analyses using two-color microarrays. 3D7/GDV1-GFP-DD^{OFF} ring-stage parasites were split and cultured separately in the absence or presence of Shield-1, and total RNA was harvested at seven paired time points spanning the remaining 24 hours of generation 1 [24 to 32 hours post-invasion (hpi), 32 to 40 hpi, and 40 to 48 hpi] and the first 40 hours after reinvasion in generation 2 (8 to 16 hpi, 16 to 24 hpi, 24 to 32 hpi, and 32 to 40 hpi) (Fig. 2A). As expected, GDV1-GFP-DD expression triggered a transcriptional response characteristic of sexual commitment and early differentiation. This effect was evident from the induction of *ap2-g* in generation 1, followed by the activation of early gametocyte markers (*5*, *7*, *8*, *17*) after reinvasion (Fig. 2G, fig. S4, and table S5). In the parasite line F12, a 3D7-derived gametocyte-deficient clone carrying a loss-of-function mutation in *ap2-g* (*5*, *18*), GDV1-GFP-DD expression still activated *ap2-g* but failed to launch a sexual differentiation response (Fig. 2H and table S6). Next to *ap2-g*, only eight other genes, all of which are marked by HP1, were significantly induced in F12/GDV1-GFP-DD^{ON} parasites. This set included *dblmsp2*, which was also induced in 3D7/GDV1-GFP-DD^{ON} parasites (Fig. 2, G and H). Given that DBLMSP2 is a merozoite surface antigen expressed only in a small subpopulation of schizonts (*19*, *20*), the GDV1-dependent activation of the *dblmsp2* locus suggests that DBLMSP2 may be expressed specifically in sexually committed schizonts. In summary, these findings show that GDV1 is an upstream activator of sexual commitment and likely triggers this process by antagonizing HP1-dependent silencing of *ap2-g*.

To test whether GDV1 associates with heterochromatin in vivo, we conducted comparative chromatin immunoprecipitation–sequencing (ChIP-seq) experiments. 3D7/GDV1-GFP-DD^{OFF} parasites were split at 28 to 34 hpi and cultured in parallel in the absence or presence of Shield-1, and paired chromatin samples were harvested 2 (30 to 36 hpi), 6 (34 to 40 hpi), and 10 (38 to 44 hpi) hours after Shield-1 addition. We found that (i) GDV1-GFP-DD associates specifically with heterochromatin throughout the genome (Fig. 3A, fig. S5, and table S7), (ii) GDV1-GFP-DD occupancy was markedly higher in 3D7/GDV1-GFP-DD^{ON} parasites than in 3D7/GDV1-GFP-DD^{OFF} parasites (fig. S5 and table S7), and (iii) GDV1-GFP-DD occupancy was highly correlated with that of HP1 (Fig. 3B). Moreover, GDV1-GFP-DD occupancy peaked 6 hours postinduction and decreased

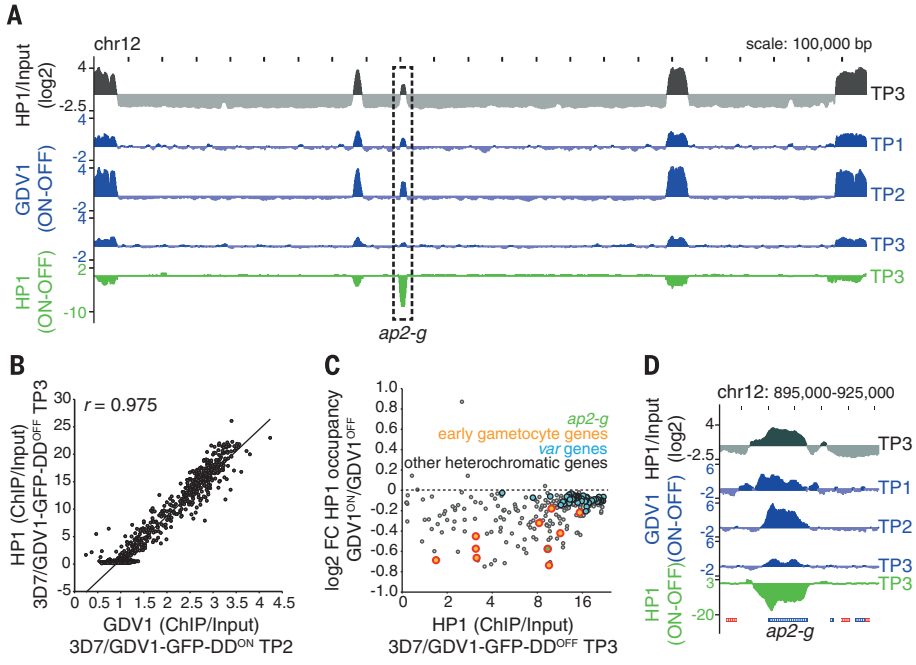


Fig. 3. GDV1 associates with heterochromatin throughout the genome and triggers HP1 removal at *ap2-g*. (**A**) Plots of the ratio of the HP1 level to the input for 3D7/GDV1-GFP-DD^{OFF} schizonts [38 to 44 hpi; time point 3 (TP3)] (gray), ChIP-seq subtraction tracks displaying relative enrichment of GDV1-GFP-DD in 3D7/GDV1-GFP-DD^{ON} schizonts at 2 (30 to 36 hpi; TP1), 6 (34 to 40 hpi; TP2), and 10 (38 to 44 hpi; TP3) hours after Shield-1 addition (blue), and the relative depletion of HP1 in 3D7/GDV1-GFP-DD^{ON} parasites at TP3 (green). chr12, chromosome 12; bp, base pairs. (**B**) Correlation between GDV1-GFP-DD enrichment in 3D7/GDV1-GFP-DD^{ON} schizonts (34 to 40 hpi; TP2) and HP1 enrichment in 3D7/GDV1-GFP-DD^{OFF} schizonts at each coding region. *r*, Pearson correlation coefficient. (**C**) Fold change in HP1 enrichment upon GDV1-GFP-DD overexpression in relation to HP1 enrichment in 3D7/GDV1-GFP-DD^{OFF} schizonts for each heterochromatic gene. (**D**) Zoom-in view of the enrichment and subtraction tracks at the *ap2-g* locus. Numbers above the graph indicate the position in base pairs.

substantially thereafter (Fig. 3A, fig. S5, and table S7). This drop in GDV1-GFP-DD signal coincided with reduced HP1 occupancy over heterochromatic genes in 3D7/GDV1-GFP-DD^{ON} parasites compared with that in 3D7/GDV1-GFP-DD^{OFF} parasites (Fig. 3, A and C, and table S7). Although the vast majority of heterochromatic loci, in particular those displaying high HP1 occupancy, such as the *var* genes encoding PfEMP1, displayed only slightly decreased HP1 levels, some genes exhibited as much as 40% reduction in HP1 occupancy (Fig. 3C and table S7). This group of genes included *ap2-g* and most known HP1-associated early gametocyte markers, including *geco* (*21*), *pfgexp17* (*22*), and *pfg14_748* (*8*, *17*) (Fig. 3, C and D, and table S7). These data are consistent with the microarray results, where GDV1-GFP-DD expression activated *ap2-g* and early gametocyte genes but had no effect on the expression of the bulk of heterochromatic loci, including *var* genes (Fig. 2, G and H). Given the 50 to 60% sexual conversion rate observed for 3D7/GDV1-GFP-DD^{ON} parasites (see above), a 30 to 40% reduction in HP1 occupancy indicates that HP1 may be depleted at these loci specifically in sexually committed parasites, but single-cell approaches are required to confirm this hypothesis. Overall, we suggest that GDV1 destabilizes heterochromatin and thus allows specific transcription factors to activate expression of *ap2-g* and other gametocyte-specific heterochromatic genes, and this process may play an important role in the positive autoregulatory feedback loop proposed to reinforce AP2-G expression in committed parasites (*5*, *6*, *23*). How GDV1 achieves specificity in unlocking specific HP1-associated genes despite binding heterochromatin genome-wide is a challenging question to be addressed in the future.

Because GDV1 activates sexual commitment, the question arises of how parasites limit GDV1 expression to prevent sexual conversion in asexual schizonts. A recent study identified a multiexon-long noncoding *gdv1* antisense RNA (asRNA) that initiates downstream of the *gdv1* locus and overlaps with the ATG start codon of *gdv1* (*24*), which is a hallmark feature of regulatory asRNAs (*25*). To investigate whether the *gdv1* asRNA participates in regulating sexual commitment, we created a *gdv1* asRNA loss-of-function [asRNA knockout (asKO)] mutation in F12 parasites (yielding F12/*gdv1*-asKO parasites) (Fig. 4A and fig. S6). Strand-specific RNA sequencing (RNA-seq) analysis identified a small set of genes that were consistently differentially expressed between F12/*gdv1*-asKO

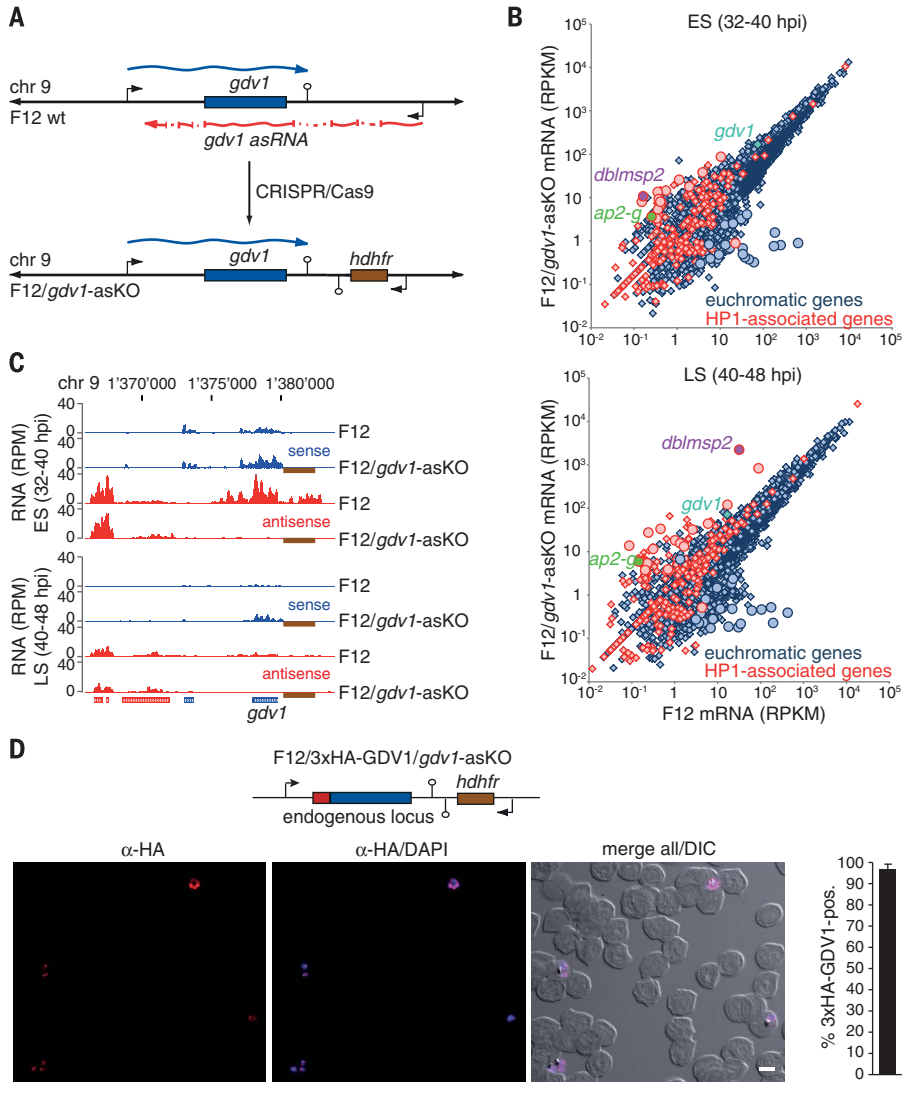


Fig. 4. A *gdv1* asRNA antagonizes GDV1-dependent sexual commitment. (**A**) *gdv1* locus in F12 wild-type (wt) and F12/*gdv1*-asKO parasites. The *gdv1* sense transcript (blue), five-exon *gdv1* asRNA (*24*) (red), and human *dhfr* (*hdhfr*) resistance marker (brown) are highlighted. chr 9, chromosome 9. (**B**) Comparison of gene expression levels in F12 wild-type and F12/*gdv1*-asKO early and late schizonts. Genes deregulated >5-fold at both time points are indicated by circles. RPKM, reads per kilobase million. (**C**) University of California–Santa Cruz Genome Browser screenshots of RNA-seq coverage plots over the *gdv1* locus in F12 wild-type and F12/*gdv1*-asKO early and late schizonts. The human *dhfr* resistance cassette that is present downstream of the *gdv1* locus in F12/*gdv1*-asKO parasites and absent in the 3D7 reference genome is indicated by a brown rectangle. Numbers along the top indicate the position in base pairs. RPM, reads per million. (**D**) Endogenous *gdv1* locus in F12/3×HA-GDV1/*gdv1*-asKO parasites and results from overview IFAs with antibodies to HA (α-HA) in early schizonts (32 to 40 hpi). Scale bar, 5 μm. The percentage of 3×HA-GDV1-positive parasites is shown at the right [results are the mean for three biological replicates (100 infected RBCs counted per sample); error bars indicate SD].

and F12 wild-type parasites (17 genes were up-regulated and 23 genes were down-regulated in the mutants) (Fig. 4B and table S8). Similar to F12 parasites expressing ectopic GDV1-GFP-DD (Fig. 2H), F12/*gdv1*-asKO parasites showed marked induction of *ap2-g*, *dblmsp2*, and two early gametocyte genes (*pfg14_748* and PF3D7_1477400) (*8*, *17*), and all except one up-regulated gene are HP1-associated genes (Fig. 4B, fig. S6, and table S8). *gdv1* sense transcripts were slightly increased in the F12/*gdv1*-asKO population, and *gdv1* antisense transcripts were undetectable, as expected (Fig. 4, B and C; fig. S6; and table S8). These results indicated that the *gdv1* asRNA acts as a negative regulator of GDV1 expression. To confirm this hypothesis, we tagged endogenous GDV1 with HA in these parasites (yielding an F12/3×HA-GDV1/*gdv1*-asKO population) and observed that indeed almost all parasites (96.7% ± 2.5% SD) expressed 3×HA-GDV1 (Fig. 4D and fig. S7). Lastly, we showed that deletion of the *gdv1* asRNA locus in a conditional AP2-G mutant resulted in markedly increased production of gametocytes (fig. S8 and supplementary text). Together, these findings demonstrate a central role for the *gdv1* asRNA in regulating GDV1-dependent activation of sexual commitment. We anticipate that this mechanism likely involves inhibiting GDV1 expression by interference with *gdv1* mRNA transcription, stability, or translation, similar to asRNA-mediated gene regulation in other organisms (*26*).

We identified GDV1-mediated heterochromatin destabilization as an epigenetic control strategy regulating sexual cell fate decision in *P. falciparum*. Our discovery of the *gdv1* asRNA as a negative regulator of sexual commitment is reminiscent of long noncoding RNA–mediated control of gametogenesis in yeasts (*27*, *28*). In *Saccharomyces cerevisiae*, nutritional stress triggers gametogenesis by activating the transcriptional regulator inducer of meiosis 1 (IME1) (*28*). A long noncoding RNA in the *ime1* promoter and antisense transcription of *ime4* are key factors in preventing IME1 expression under noninducing conditions (*29*, *30*). These parallels raise the exciting possibility that evolutionarily divergent unicellular eukaryotes employ conceptually similar regulatory logic to control entry into the sexual phases of their life cycles. All *Plasmodium* species infecting humans possess a GDV1 ortholog, suggesting that GDV1-based regulation of sexual commitment is conserved in all human-infective malaria parasites. In conclusion, our study contributes to understanding of the molecular pathway underlying the formation of malaria transmission stages and provides opportunities for the development of intervention strategies targeting transmission of human malaria.

REFERENCES AND NOTES

1. S. H. Kwon, J. L. Workman, *Mol. Cells* **26**, 217–227 (2008).
2. C. Flueck *et al.*, *PLOS Pathog.* **5**, e1000569 (2009).
3. N. Rovira-Graells *et al.*, *Genome Res.* **22**, 925–938 (2012).
4. T. S. Voss, Z. Bozdech, R. Bártfai, *Curr. Opin. Microbiol.* **20**, 88–95 (2014).
5. B. F. Kafsack *et al.*, *Nature* **507**, 248–252 (2014).
6. A. Sinha *et al.*, *Nature* **507**, 253–257 (2014).
7. N. M. B. Brancucci *et al.*, *Cell Host Microbe* **16**, 165–176 (2014).
8. S. Eksi *et al.*, *PLOS Pathog.* **8**, e1002964 (2012).
9. L. A. Banaszynski, L. C. Chen, L. A. Maynard-Smith, A. G. L. Ooi, T. J. Wandless, *Cell* **126**, 995–1004 (2006).
10. C. M. Armstrong, D. E. Goldberg, *Nat. Methods* **4**, 1007–1009 (2007).
11. A. Gaspar-Maia *et al.*, *Nature* **460**, 863–868 (2009).
12. L. Bugga, I. E. McDaniel, L. Engie, J. A. Armstrong, *PLOS ONE* **8**, e59496 (2013).
13. M. C. Bruce, P. Alano, S. Duthie, R. Carter, *Parasitology* **100**, 191–200 (1990).
14. F. Silvestrini, P. Alano, J. L. Williams, *Parasitology* **121**, 465–471 (2000).
15. M. C. Bruce, R. N. Carter, K. Nakamura, M. Aikawa, R. Carter, *Mol. Biochem. Parasitol.* **65**, 11–22 (1994).
16. N. M. B. Brancucci *et al.*, *Cell* **171**, 1532–1544.e15 (2017).
17. S. Eksi *et al.*, *Mol. Biochem. Parasitol.* **143**, 90–99 (2005).
18. P. Alano *et al.*, *Exp. Parasitol.* **81**, 227–235 (1995).
19. J. A. Pearce, K. Mills, T. Triglia, A. F. Cowman, R. F. Anders, *Mol. Biochem. Parasitol.* **139**, 141–151 (2005).
20. A. Amambua-Ngwa *et al.*, *PLOS Genet.* **8**, e1002992 (2012).
21. B. J. Morahan *et al.*, *Eukaryot. Cell* **10**, 1492–1503 (2011).
22. F. Silvestrini *et al.*, *Mol. Cell. Proteomics* **9**, 1437–1448 (2010).
23. A. Poran *et al.*, *Nature* **551**, 95–99 (2017).
24. K. M. Broadbent *et al.*, *BMC Genomics* **16**, 454 (2015).
25. F. Huber *et al.*, *Cell Rep.* **15**, 2625–2636 (2016).

26. V. Pelechano, L. M. Steinmetz, *Nat. Rev. Genet.* **14**, 880–893 (2013).
27. E. Hiriart, A. Verdel, *Chromosome Res.* **21**, 653–663 (2013).
28. F. J. van Werven, A. Amon, *Philos. Trans. R. Soc. London Ser. B* **366**, 3521–3531 (2011).
29. B. Gelfand *et al.*, *Mol. Cell. Biol.* **31**, 1701–1709 (2011).
30. F. J. van Werven *et al.*, *Cell* **150**, 1170–1181 (2012).
31. R. Edgar, M. Domrachev, A. E. Lash, *Nucleic Acids Res.* **30**, 207–210 (2002).
32. P. Alano, S. Premawansa, M. C. Bruce, R. Carter, *Mol. Biochem. Parasitol.* **46**, 81–88 (1991).

ACKNOWLEDGMENTS

We are grateful to M. van de Vegte-Bolmer and R. Sauerwein for determining gametocyte sex ratios and providing anti-Pfs16, to D. Richard for providing the pL6-3HA_glmS-246 plasmid, and to T. Haefliger for technical assistance. **Funding:** This work was supported by the Swiss National Science Foundation (grant numbers 31003A_143916, 31003A_163258, and BSCGI0_157729), the Fondation Pasteur Suisse, and the Netherlands Organization for Scientific Research (NWO-Vidi 864.11.007). S.A.F. is supported by a Ph.D. fellowship from the European Community's Seventh Framework Program (grant no. FP7/2007–2013) under grant agreement no. 242095 and no. ParaMet 290080. **Author contributions:** M.F. designed and performed experiments, analyzed data, prepared illustrations, and wrote the paper. S.A.F. performed and analyzed ChIP-seq and RNA-seq experiments. I.N. designed and cloned CRISPR-Cas9 mother plasmids and performed experiments involving recombinant proteins. N.M.B.B. performed experiments related to the 3D7/3xHA-GDV1 and F12/3xHA-GDV1/*gdv1*-asKO parasites. E. Carrington performed and analyzed quantitative reverse transcription polymerase chain reaction experiments. E. Carrió performed experiments involving 3D7/AP2-G-GFP-DDglmS parasites. S.M. performed LC-MS/MS experiments. P.J. provided conceptual advice. P.J., R.B., and T.S.V. provided resources. R.B. designed, supervised, and analyzed experiments. T.S.V. conceived the study; designed, supervised, and analyzed experiments; and wrote the paper. All authors contributed to editing of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code to understand and assess the conclusions of this research are available in the main text and supplementary materials and via the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) (*31*) under accession numbers GSE95549 (microarray data) and GSE94901 (ChIP-seq and RNA-seq data).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1259/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S9
References (*33*–*58*)

24 May 2017; accepted 17 January 2018
10.1126/science.aan6042

COGNITIVE SCIENCE

Precursors of logical reasoning in preverbal human infants

Nicoló Cesana-Arlotti,[1,2]* Ana Martín,[1]* Ernő Téglás,[2] Liza Vorobyova,[1,2] Ryszard Cetnarski,[1,3] Luca L. Bonatti[4,1]*

Infants are able to entertain hypotheses about complex events and to modify them rationally when faced with inconsistent evidence. These capacities suggest that infants can use elementary logical representations to frame and prune hypotheses. By presenting scenes containing ambiguities about the identity of an object, here we show that 12- and 19-month-old infants look longer at outcomes that are inconsistent with a logical inference necessary to resolve such ambiguities. At the moment of a potential deduction, infants' pupils dilated, and their eyes moved toward the ambiguous object when inferences could be computed, in contrast to transparent scenes not requiring inferences to identify the object. These oculomotor markers resembled those of adults inspecting similar scenes, suggesting that intuitive and stable logical structures involved in the interpretation of dynamic scenes may be part of the fabric of the human mind.

Fifty years ago, Piaget argued that logic in the mind is the culmination of a long developmental process, extending into adolescence. Forty years ago, Fodor answered that if learning implies testing hypotheses, then learners must possess the representational resources to formulate them, including logical primitives: rule-like combinatorial concepts embedded in a compositional system of representation, or a language of thought (*1*).

After four decades, we still lack insight into the nature and development of the logical representations, if any, that structure infants' thinking and problem-solving. Partly, this profound lack of knowledge stems from the widespread belief that infant cognition relies on independent modules, functioning early and efficiently, but not supported or connected by general reasoning (*2*). Partly, it stems from the assumption that although logical representations are involved in processing language, and hence are present in organisms that master a natural language, it is difficult, perhaps impossible (*3*), to identify them in nonverbal organisms (*4*). Acquiring language does improve cognition, perhaps also by creating novel logical representations (*5*). However, none of these considerations weakens the real force of Fodor's argument, although its premises need to be reappraised. Although infants possess learning mechanisms that do not require hypotheses [e.g., bottom-up tracking of statistical regularities (*6*)], flexible and productive hypothesis testing does begin in infancy, with a vengeance. Infants can generate hypotheses about uncertain future events (*7*), flexibly adapting them to novel, albeit subtle, elements of a situation (*8*, *9*). They measure the evidence in support (*10*) and test alternative hypotheses when violations occur (*11*, *12*). Such abilities extend far beyond precompiled mechanisms for domain-specific responses, demonstrating a high degree of rationality in several domains. One prominent account of them depicts infants as precocious Bayesian reasoners. However, most Bayesian theories require a logical scaffolding to formulate, test, and modify hypotheses (*8*, *13*–*15*). Thus, characterizing the basic logical representations available to preverbal infants for formulating hypotheses remains fundamental to understanding the very nature of knowledge acquisition (*16*).

Here we begin investigating the developmental precursors of such scaffolding, looking for behavioral correlates of one simple logical representation and rule: disjunction (either A or B) and disjunctive syllogism (not A, therefore B). Although elementary, this schema grounds one crucial hypothesis-testing strategy: Sherlock Holmes–like case-by-case analysis of different possibilities, excluding alternatives until the culprit is found. Attempts to find clear evidence of disjunctive syllogism in nonhuman animals have so far been inconclusive (*4*, *17*). A related reasoning pattern has been studied in toddlers' and preschoolers' word-learning strategies (*18*, *19*), but it is unknown whether it is within the conceptual repertoire of preverbal infants. We first investigate whether infants can frame disjunctive hypotheses and make inferences by logically eliminating alternatives, testing their reactions to outcomes that violate conclusions of this deductive process. Then, we identify markers of inferential activity by examining the dynamics of oculomotor responses during inference making. Last, we explore stability across development by comparing the oculomotor responses of infants, toddlers, and adults passively looking at nonverbal scenes that potentially involve logical inferences.

We studied 12- and 19-month-old infants, two ages at the onset of speech production and language learning but that precede the development of extensive language knowledge. We presented infants with scenes injected with ambiguity about the identity of an object, which could be resolved through disjunctive syllogism. In experiments 1 and 2, two objects different in shape, texture, color, and category, but with identical top parts (say, a dinosaur and a flower), enter a virtual theater (Fig. 1 and fig. S1). An occluder hides them, and a cup scoops one of them from behind it, with only the top part visible. Thus, infants cannot know the identity of the scooped object and may establish a disjunctive representation. Then, the occluder moves downward, revealing one object—say, the dinosaur. We call this moment the "potential deduction phase," in which infants have evidence to disambiguate the identity of the scooped object by disjunctive syllogism. Last, in the "outcome phase," the dinosaur leaves the stage, and the cup reveals the second object. Half of the time, the revealed object is consistent with the conclusion suggested by the logical inference (it is the flower), whereas the other half, it is inconsistent (it is the dinosaur). We recorded looking time during the outcome phase in a violation of expectation (VOE) paradigm. Both 12- and 19-month-olds looked longer at the inconsistent outcome, suggesting that they may have derived the identity of the object in the cup through logical inference and were surprised when this conclusion was violated, as revealed by mean looking times, M [experiment 1, 19-month-olds (n = 24), $M_{\text{consistent}}$ = 7.7 s, $M_{\text{inconsistent}}$ = 10.5 s, $F_{1,23}$ = 5.79, P = 0.025; experiment 2, 12-month-olds (n = 24), $M_{\text{consistent}}$ = 6.2 s, $M_{\text{inconsistent}}$ = 7.6 s, $F_{1,23}$ = 5.19, P = 0.032; repeated measures analyses of variance (ANOVAs)].

In experiments 1 and 2, in an inconsistent outcome, an object appears twice successively: The occluder lowers, revealing the dinosaur, which exits the stage, and then the cup reveals a dinosaur again (Fig. 1A, vi-2). Conversely, in a consistent outcome, two different objects appear successively: After the dinosaur exits, the cup reveals the flower. Thus, the infants may have reacted not to a logical inconsistency but to the surface aspects of the final sequence, when the same object appeared twice in succession (Fig. 1A, vi-1). In experiments 3 and 4, the logical status of the final object sequence reverses (Fig. 2A). The movies are identical to those of experiments 1 and 2 until the potential deduction phase. There, the occluder never lowers; one object (e.g., a snake) exits from its side, remaining visible for about 1.5 s, then returns behind it. In the outcome phase, the cup never reveals its content. Instead, another object exits the occluder: sometimes the snake again, and sometimes the other object in the pair (in this example, a ball; Fig. 2A, vi). The former outcome is consistent with the logical inference; however, unlike experiments 1 and 2, one single object is seen twice in succession. The latter outcome is

[1]Universitat Pompeu Fabra, Center for Brain and Cognition, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain. [2]Cognitive Development Centre, Central European University, H-1015 Budapest, Hungary. [3]Nencki Institute of Experimental Biology, Polish Academy of Sciences, Ludwika Pasteura 3, 02-093 Warszawa, Poland. [4]ICREA (Catalan Institution for Research and Advanced Studies), Pg. Lluís Companys 23, 08010 Barcelona, Spain.
*Corresponding author. Email: nicolocesanaarlotti@gmail.com (N.C.-A.); ana.martin@upf.edu (A.M.); lucabonatti@mac.com (L.L.B.)

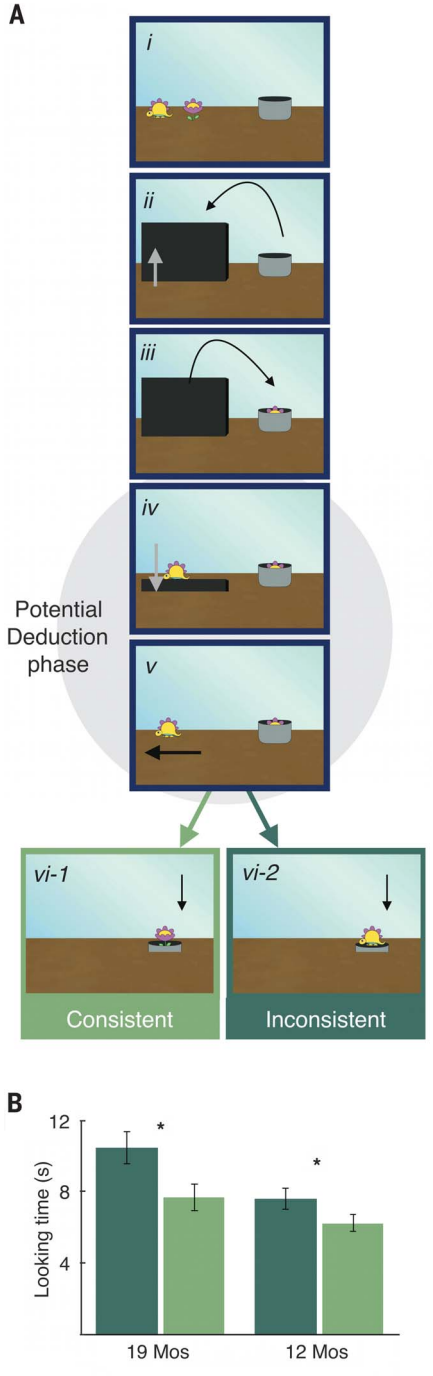
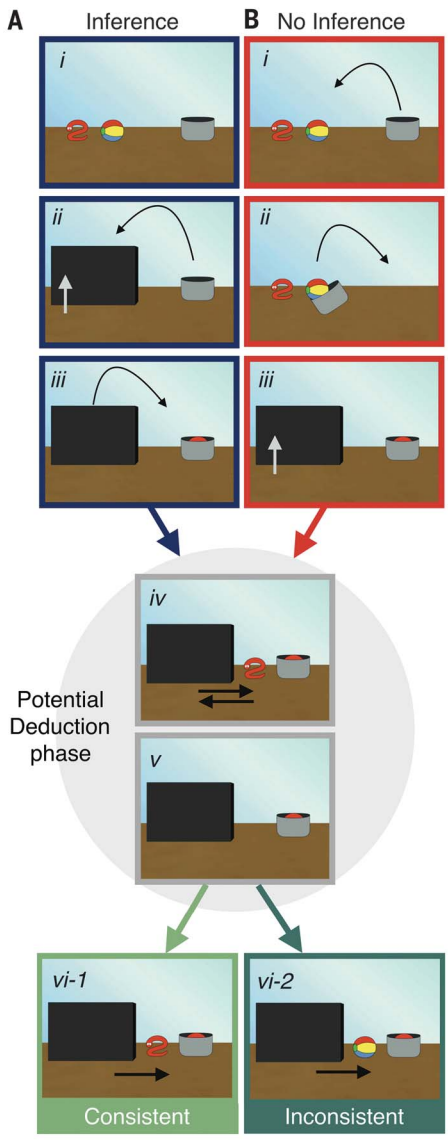


Fig. 1. Infants look longer at outcomes that are inconsistent with a logical deduction. (**A**) (i) Two objects with an identical upper part enter the theater. (ii and iii) An occluder hides them, and a cup scoops one from behind, only the top part of which is visible (dinosaur or flower?). (iv and v) The occluder lowers, allowing the observer to infer the identity of the object in the cup (not dinosaur, therefore flower). (vi) The cup reveals its content, which is either consistent (flower) or inconsistent (dinosaur) with the inference. (**B**) Mean (± SEM) time spent looking at the outcomes (in seconds). Both 19- and 12-month-olds looked longer at the inconsistent outcome. *$P < 0.05$. Mos, months.



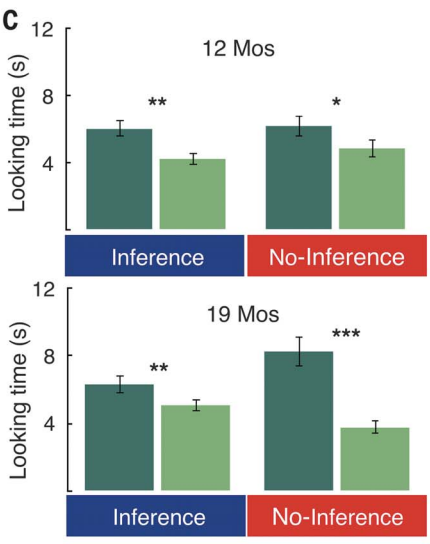


Fig. 2. Infants' logical reasoning does not depend on how the scene is physically realized. (**A**) (i to iii) Inference condition: The identity of the object in the cup cannot be determined before the potential deduction phase. (**B**) (i to iii) No-inference condition: The cup scoops the object in full view, so its identity is known. (iv and v) Potential deduction phase [common to (A) and (B)]: Only in the inference condition is a deduction needed to determine the cup content. Note that the physical realization of this phase is very different from that of experiments 1 and 2. (vi-1 and vi-2) Outcome phase [common to (A) and (B)]: An object exits the occluder, yielding a consistent or inconsistent outcome. (**C**) Mean (± SEM) time spent looking at the outcomes (in seconds). Both 12- and 19-month-olds looked longer at the inconsistent outcome. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

inconsistent, but two different objects are seen once in succession. If infants respond to the surface aspects of the final sequence, they should disregard the logical consistency of the outcome and look longer when the single object appears twice, as in experiments 1 and 2. If, instead, their behavior is guided by a logical inference, they should look longer when the outcome is inconsistent with it even if the final sequence reverses that of experiments 1 and 2. Both 19 and 12-month-olds looked longer at the inconsistent outcome (Fig. 2C), suggesting that they reacted to the logical gist of a scene [experiment 3, 19-month-olds (n = 24), $M_{consistent}$ = 4.9 s, $M_{inconsistent}$ = 6.2 s, $F_{1,23}$ = 8.5, P = 0.008; experiment 4, 12-month-olds (n = 24), $M_{consistent}$ = 4.2 s, $M_{inconsistent}$ = 6.1 s, $F_{1,23}$ = 11, P = 0.003; repeated measures ANOVAs]. These results also control for other nonlogical explanations, such as an object's magical disappearance in the inconsistent outcome of experiments 1 and 2 (no such disappearance occurred in experiments 3 and 4) or its greater featural variability in experiments 3 and 4 (reversed in experiments 1 and 2).

VOE only measures a response post hoc, after a conclusion has been reached (*20*). Adults reasoning with language make disjunctive inferences as early as they have the relevant evidence (*21*, *22*). The data reported so far do not characterize the unfolding of an inference in the infant mind. To explore this, we analyzed oculomotor responses during the potential deduction phase. We created novel scenes identical to experiments 3 and 4 in the potential deduction and outcome phases, but requiring no inference to identify the object in the cup (experiments 5 and 6; Fig. 2B). We did this by showing the cup scooping one object in full view before occlusion. Thus, unlike experiments 3 and 4, in experiments 5 and 6, infants already know which object is in the cup before the potential deduction phase. As expected at these ages (*23*), infants looked longer at an outcome inconsistent with the identity of the (known) object in the cup [experiment 5, 19-month-olds

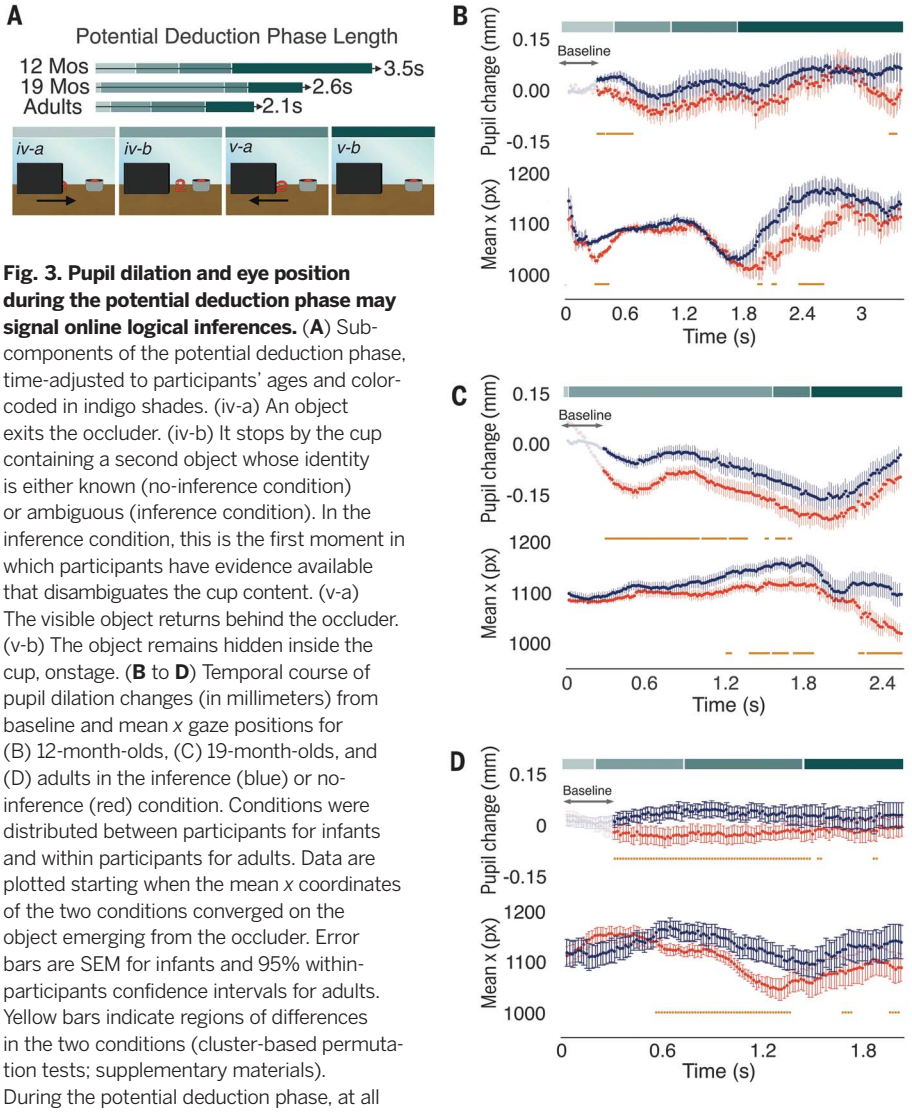


Fig. 3. Pupil dilation and eye position during the potential deduction phase may signal online logical inferences. (**A**) Subcomponents of the potential deduction phase, time-adjusted to participants' ages and color-coded in indigo shades. (iv-a) An object exits the occluder. (iv-b) It stops by the cup containing a second object whose identity is either known (no-inference condition) or ambiguous (inference condition). In the inference condition, this is the first moment in which participants have evidence available that disambiguates the cup content. (v-a) The visible object returns behind the occluder. (v-b) The object remains hidden inside the cup, onstage. (**B** to **D**) Temporal course of pupil dilation changes (in millimeters) from baseline and mean *x* gaze positions for (B) 12-month-olds, (C) 19-month-olds, and (D) adults in the inference (blue) or no-inference (red) condition. Conditions were distributed between participants for infants and within participants for adults. Data are plotted starting when the mean *x* coordinates of the two conditions converged on the object emerging from the occluder. Error bars are SEM for infants and 95% within-participants confidence intervals for adults. Yellow bars indicate regions of differences in the two conditions (cluster-based permutation tests; supplementary materials). During the potential deduction phase, at all ages, participants who had to perform a logical inference to identify the cup content had pupils that were more dilated and gazed more toward the cup.

(n = 24), $M_{consistent}$ = 3.8 s, $M_{inconsistent}$ = 8.3 s, $F_{1,23}$ = 26.1, P = 0.0001; experiment 6, 12-month-olds (n =24), $M_{consistent}$ = 4.9 s, $M_{inconsistent}$ = 6.2 s, $F_{1,23}$ = 4.9, P = 0.037; repeated measures ANOVAs]. But our focus here is the temporal course of oculomotor responses during the potential deduction phase, which we expected to be modulated by the need for an inference. Cluster-based permutation tests (*24*, *25*) revealed that at several points during the potential deduction phase, the infants' pupils dilated more when the scene licensed an inference than when it did not, suggesting increased cognitive activity possibly due to inference-making. By the end of this phase, infants also displaced their eyes toward the cup more markedly (Fig. 3, B and C, and supplementary materials) and switched their focus from the visible object to the cup in more trials ($M_{inference}$ = 71%, $M_{no\_inference}$ = 50%, $F_{1,88}$ = 10.4, P = 0.002, two-way ANOVA; fig. S2 and supplementary materials) when a deduction was needed than when it was not.

Only when the potential deduction phase afforded an inference did higher pupil dilation and visible object-to-cup shifts contribute to predicting success at identifying inconsistencies in the later outcome phase. No such predictive relation occurred absent the need for an inference (fig. S3 and supplementary materials). The fact that this relation occurred only when an inference may have been involved suggests that oculomotor markers in the potential deduction phase are not simply due to memory of past event structures but are tied to some kind of mental inference about the identity of the object in the cup, drawn soon after infants acquire the disambiguating evidence.

Thus, these oculomotor markers suggest that preverbal infants efficiently deploy logical procedures to process the components of an unfolding scene. To assess developmental stability, we inspected adults' oculomotor responses during the same potential deduction phase. Adults (experiment 7, n = 30) saw 96 scenes patterned upon those of experiments 1 to 6. Like infants, during the potential deduction phase, adults' pupils dilated more, and their eyes tended to look more toward the cup, when the scene licensed an inference. Again, this occurred regardless of its physical realization. Adults and infants differed only in the speed of such markers, but not qualitatively (Fig. 3D, figs. S4 and S5, and supplementary materials).

Our data document the early presence of primitive logical abilities. Without instructions or tasks, infants spontaneously reason logically while a scene unfolds. Specific behavioral markers can be used to study the precise temporal course of their reasoning process. Because such markers already appear at ages when language development has barely begun, our data suggest that precursors of logical reasoning are independent of language acquisition. Their stability across ages and spontaneous deployment suggest that some form of elementary logical reasoning may be a primitive property of the logical circuitry in the human brain (*26*). Explaining our data without invoking deductive inferences has a cost. Bayesian iterative models, which evaluate the most likely of the alternatives first and cycle through them when the first choice is discarded, could mimic deductive syllogism without assuming a logical inference in the potential deduction phase. However, they require that infants represent the space of alternatives (which is equivalent to implementing a disjunctive representation), assign ordered priors to the alternatives, and assess alternative evaluations iteratively. A logical inference requires fewer assumptions. Thus, although not incompatible with Bayesian reasoning, the hypothesis that infants perform a logical inference in the potential deduction phase is a more parsimonious explanation of our results.

In spite of the stability that we document, children begin mastering a verbally expressed disjunction late in development (*17*). However, a dissociation between spontaneous inferential abilities in nonverbal contexts and their explicit verbal counterparts need not imply the lack of a concept. Instead, it indicates that mapping the spontaneous logical structures of thought onto their verbal counterparts is an extremely intricate process. A deceivingly simple word such as "or" has a very complex semantics (*27*). Unambiguous evidence for its meaning is hard to come by, a difficulty that affects the acquisition of even simpler abstract words (*28*). Thus, a consequence of our research is that much work is still needed to understand how the proper alignment between language and thought occurs.

This empirical evidence is directly relevant to the old, yet still fundamental questions debated between Fodor and Piaget. Logical representations that are crucial components of infants' natural hypothesis-testing attitude are available when infants start projecting and testing hypotheses about the world. Such representations may consist of nonlinguistic but fully language-like structures, or they may piggyback on sophisticated

object representations that can track object identities in ambiguous situations. Although further research is needed to clarify their nature, our data suggest that intuitive and stable logical structures involved in the interpretation of dynamic scenes may be essential parts of the fabric of the mind. This does not imply that all logical reasoning is spontaneous or innate, just as spontaneous and innate elementary numerical abilities do not imply that all mathematical knowledge is innate. Reasoning occurs in many different forms and at many different levels of our mental processes, and the gulf separating infant thinking from adult explicit logical reasoning is large. However, the development of reasoning abilities builds on a natural logical foundation, whose profile we are beginning to uncover.

REFERENCES AND NOTES

1. M. Piattelli-Palmarini, *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky* (Harvard Univ. Press, 1980).
2. E. S. Spelke, in *Language in Mind: Advances in the Study of Language and Thought*, D. Gentner, S. Goldin-Meadow, Eds. (MIT Press, 2003), pp. 277–311.
3. W. V. O. Quine, *Word and Object* (MIT Press, 1960).
4. D. C. Penn, K. J. Holyoak, D. J. Povinelli, *Behav. Brain Sci.* **31**, 109–130 (2008).
5. S. Carey, *The Origin of Concepts* (Oxford Univ. Press, 2009).
6. J. R. Saffran, R. N. Aslin, E. L. Newport, *Science* **274**, 1926–1928 (1996).
7. E. Téglás, V. Girotto, M. Gonzalez, L. L. Bonatti, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19156–19159 (2007).
8. E. Téglás *et al.*, *Science* **332**, 1054–1059 (2011).
9. E. Téglás, A. Ibanez-Lillo, A. Costa, L. L. Bonatti, *Dev. Sci.* **18**, 183–193 (2015).
10. H. Gweon, J. B. Tenenbaum, L. E. Schulz, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9066–9071 (2010).
11. A. E. Stahl, L. Feigenson, *Science* **348**, 91–94 (2015).
12. H. Gweon, L. Schulz, *Science* **332**, 1524 (2011).
13. A. Gopnik, *Child Dev. Perspect.* **5**, 161–163 (2011).
14. S. T. Piantadosi, J. B. Tenenbaum, N. D. Goodman, *Psychol. Rev.* **123**, 392–424 (2016).
15. S. J. Gershman, E. J. Horvitz, J. B. Tenenbaum, *Science* **349**, 273–278 (2015).
16. S. Carey, in *The Conceptual Mind: New Directions in the Study of Concepts*, E. Margolis, S. Laurence, Eds. (MIT Press, 2015), pp. 415–454.
17. S. Mody, S. Carey, *Cognition* **154**, 40–48 (2016).
18. J. Halberda, *Cognition* **87**, B23–B34 (2003).
19. J. Halberda, *Cognit. Psychol.* **53**, 310–344 (2006).
20. E. Téglás, L. L. Bonatti, *Cognition* **157**, 227–236 (2016).
21. R. B. Lea, D. P. O'Brien, S. M. Fisch, I. A. Noveck, M. D. S. Braine, *J. Mem. Lang.* **29**, 361–387 (1990).
22. C. Reverberi, D. Pischedda, M. Burigo, P. Cherubini, *Acta Psychol.* **139**, 244–253 (2012).
23. L. Bonatti, E. Frot, R. Zangl, J. Mehler, *Cognit. Psychol.* **44**, 388–426 (2002).
24. E. Maris, R. Oostenveld, *J. Neurosci. Methods* **164**, 177–190 (2007).
25. J. R. Hochmann, L. Papeo, *Psychol. Sci.* **25**, 2038–2046 (2014).
26. C. Reverberi *et al.*, *Neuroimage* **59**, 1752–1764 (2012).
27. G. Chierchia, *Logic in Grammar: Polarity, Free Choice, and Intervention* (MIT Press, 2013).
28. A. Papafragou, K. Cassidy, L. Gleitman, *Cognition* **105**, 125–165 (2007).

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Jerry Alan Fodor. We thank J. R. Hochmann, L. Papeo, L. Filippin, E. Parise, R. Feiman, S. Carey, C. Reverberi, J. Gervain, A. Endress, L. Schulz, and K. Clifton for discussions and suggestions. **Funding:** This research was supported by MINECO [Ministerio de Economía, Industria y Competitividad (Spain)] grant PSI2016-75910-P to L.L.B. and European Research Council starting grant (639840 PreLog) to E.T. **Author contributions:** N.C.-A. and L.L.B. designed the research and the infant experiments. A.M. designed the adult experiment. N.C.-A., A.M., L.V., and R.C. built the experimental material and performed the research. N.C.-A., A.M., and L.L.B. analyzed the data and wrote the paper. All authors contributed to the conception of the research and approved the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data are available at https://osf.io/8aueq/?view_only=3a58b6d3794d4bc8a0b4e257893cbda4.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1263/suppl/DC1
Materials and Methods
Figs. S1 to S6
Table S1
Reference (*29*)
Movies S1 to S9

3 August 2017; accepted 26 December 2017
10.1126/science.aao3539

NITROGEN CYCLE

Factoring stream turbulence into global assessments of nitrogen pollution

Stanley B. Grant,[1,2]* Morvarid Azizian,[2] Perran Cook,[3] Fulvio Boano,[4] Megan A. Rippy[1]

The discharge of excess nitrogen to streams and rivers poses an existential threat to both humans and ecosystems. A seminal study of headwater streams across the United States concluded that in-stream removal of nitrate is controlled primarily by stream chemistry and biology. Reanalysis of these data reveals that stream turbulence (in particular, turbulent mass transfer across the concentration boundary layer) imposes a previously unrecognized upper limit on the rate at which nitrate is removed from streams. The upper limit closely approximates measured nitrate removal rates in streams with low concentrations of this pollutant, a discovery that should inform stream restoration designs and efforts to assess the effects of nitrogen pollution on receiving water quality and the global nitrogen cycle.

Over the past century, humans have substantially increased nitrogen loading to streams and rivers, primarily from the overapplication of fertilizer for food production. The environmental consequences of this nitrogen pollution are evident in both developed and developing countries and include eutrophication of inland and coastal waters, ocean acidification, and greenhouse gas generation (*1*–*3*). Thousands of stream, river, lake, groundwater, and coastal sites across the United States are classified as impaired for nitrogen by the U.S. Environmental Protection Agency (*4*). In a recent assessment of critical Earth systems required for the continued development of human societies, nitrogen pollution was identified as one of only three planetary boundaries (along with phosphorous pollution and loss of genetic diversity) that have already been crossed (*5*). According to the U.S. National Academy of Engineering, restoring balance to the nitrogen cycle is one of the 14 "Grand Challenges" facing engineers in the 21st century (*6*).

Streams have a natural capacity to remove dissolved inorganic nitrogen (DIN, which includes nitrate, nitrite, and ammonium) through a coupling of physical transport processes and biologically mediated reactions in streambed sediments (Fig. 1A). DIN is assimilated by autotrophs growing at the sediment-water interface (benthic algal layer) and heterotrophic microbial populations in the hyporheic zone (*7*), a region of the streambed where hydrologic flow paths begin and end in the stream (*8*). As DIN travels through the hyporheic zone, it undergoes a variety of microbially mediated redox reactions, including oxidation of ammonium to nitrate (nitrification) and reduction of nitrate to nitrite, nitrous oxide, and dinitrogen (denitrification). Of these reactions, only denitrification permanently removes nitrogen from the stream through the evasion of nitrous oxide or dinitrogen gas. The production of nitrous oxide by streams is responsible for ~10% of global anthropogenic emissions of this potent greenhouse gas (*9*), of which headwater streams may account for a disproportionate fraction (*2*). Of the DIN that is assimilated, a fraction is stored (for >1 year) as particulate nitrogen in streambed sediments or in adjacent riparian vegetation (*10*), whereas the rest is remineralized and released back to the stream.

The local efficiency with which DIN is removed from a stream can be quantified by one of several nutrient-spiraling metrics (*11*). In our study, we focused on nitrate (because of its mobility, recalcitrance, and environmental effects) and quantified its removal with the nitrate uptake velocity $v_f \geq 0$ (units of meters per second), defined as the flux of nitrate into the streambed divided by the concentration of nitrate in the overlying water column.

The second Lotic Intersite Nitrogen eXperiment (LINX II), which was conducted over 5 years from 2001 to 2006, remains one of the most comprehensive studies of nitrate uptake in headwater streams to date (*7*, *9*, *12*, *13*). LINX II included ^{15}N-labeled nitrate addition experiments in 72 streams across eight regions of the United States, collectively representing eight different biomes (temperate rain forest, chaparral, northern mixed forest, deciduous forest, montane coniferous forest, temperate grassland, shrub desert, and tropical forest) and three different land-use types (reference streams, urban streams, and agriculture streams). On the basis of regression and structural equation modeling of these data, LINX II researchers concluded that the nitrate uptake velocity is controlled primarily by stream chemistry (ambient concentrations of nitrate and ammonium) and biology (gross primary production and ecosystem respiration) and only weakly by stream physics (residence time in the hyporheic zone).

Evaluations of physical controls on nitrate uptake in streams have focused on hyporheic exchange (circulation of water through the hyporheic zone) quantified on the basis of transient storage analysis of conservative tracer injection experiments (*14*) or physical models of water pumping through streambed sediments by static and dynamic pressure variations (*2*, *8*, *15*). Missing from these previous assessments is turbulent mass transport across the concentration boundary layer (CBL) above the streambed. This transport mechanism is a key control on the delivery of oxygen to fine-grained (nonpermeable) sediments (*16*), although its role in mass transfer to coarser (permeable) sediments (like most of the headwater streams included in the LINX II study) is not clear (*17*).

Given the CBL's position between the stream and streambed (Fig. 1A), we hypothesized that nitrate uptake by permeable streambeds might be "bottlenecked" by turbulent transport across the CBL. In that event, the uptake velocity can be expressed as the product of a mass transfer coefficient k_m that depends solely on stream physics (the velocity with which mass is "squeezed" across the CBL by turbulence, units of meters per second) and an efficiency α that captures the coupled hydrogeology and biogeochemistry of nitrate uptake in the benthic algal layer and hyporheic zone (the fraction of nitrate delivered to the streambed that is removed by assimilation and denitrification, unitless) (*18*)

$$v_f = \alpha k_m,\ v_f \geq 0,\ 0 \leq \alpha \leq 1,\ k_m \geq 0 \quad (1A)$$

$$\alpha = 1 - \frac{1}{\psi + 1},\ 0 \leq \psi < \infty \quad (1B)$$

$$\psi = \frac{v_{bed}}{k_m} \quad (1C)$$

Conceptually, the mass transfer coefficient k_m represents the potential (mass transfer–limited) uptake velocity of a stream, whereas the efficiency α indicates the fraction of that potential realized in practice. The efficiency depends on a dimensionless number ψ, which represents the balance of nitrate uptake in the streambed (v_{bed}, units of meters per second) and turbulent mass transfer across the CBL. Because efficiency α varies from 0 ($\psi \to 0$) to 1 ($\psi \to \infty$), if our hypothesis is correct the uptake velocity should always be less than or equal to the mass transfer coefficient: $v_f \leq k_m$ (see Eq. 1A).

As a test of our hypothesis, we estimated values of the mass transfer coefficient at all LINX II sites where uptake velocities were reported for both assimilation and denitrification [total uptake ($v_{f,tot}$), units of meters per second] and denitrification alone [denitrification uptake ($v_{f,den}$), units of meters

[1]Department of Civil and Environmental Engineering, Henry Samueli School of Engineering, University of California, Irvine, CA 92697, USA. [2]Department of Chemical Engineering and Materials Science, Henry Samueli School of Engineering, University of California, Irvine, CA 92697, USA. [3]Water Studies Centre, School of Chemistry, Monash University, Clayton, Victoria 3800, Australia. [4]Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino 10129, Italy.
*Corresponding author. Email: sbgrant@uci.edu

per second] (69 and 49 of the 72 LINX II sites, respectively) (*7*, *12*, *13*). Site-specific values of the transfer coefficient k_m were estimated from surface renewal theory, which assumes that mass transport across the CBL occurs by sweep and ejection events associated with coherent turbulence in the stream, together with molecular diffusion of mass into the streambed (*19*). This theory predicts that k_m can be calculated from routinely measured features of a stream, including slope (S, unitless) and depth (h, units of meters), together with temperature-corrected values for the kinematic viscosity of water (υ, units of square meters per second) and the molecular diffusion coefficient of nitrate in water (D_m, units of square meters per second)

$$k_m = 0.17 u_* \mathrm{Sc}^{-2/3} \quad (2A)$$

$$\mathrm{Sc} = \upsilon / D_m \quad (2B)$$

$$u_* = \sqrt{ghS} \quad (2C)$$

The Schmidt number (Sc, unitless) represents the relative importance of molecular diffusion of momentum and mass, the shear velocity (u_*, units of meters per second) is a measure of stream turbulence, and g = 9.81 m s^{-2} is the acceleration of gravity. Very similar formulae for calculating the mass transfer coefficient (Eq. 2A) are obtained for different conceptual models of the sediment-water interface (e.g., rough versus smooth) [reviewed in (*17*)].

With few exceptions and consistent with our hypothesis, the LINX II total and denitrification uptake velocities conform to the inequality $v_f \le k_m$ (Fig. 1, B and C). The implied removal efficiencies (computed from the ratio $\alpha = v_f/k_m$) span approximately three ($10^{-4} < \alpha_{den} < 0.1$) and four ($10^{-4} < \alpha_{tot} \le 1$) orders of magnitude for denitrification and total uptake, respectively (Fig. 1D). The reduced range for α_{den} probably reflects the restrictive nature of denitrification, which requires nitrate to be transported into the streambed (e.g., by hyporheic exchange) and the presence of anoxic conditions and organic carbon, both of which may be rate-limiting in some streams (*10*, *12*, *13*). For the few sites that do not conform to the inequality $v_f \le k_m$, the total uptake velocity exceeds the mass transfer coefficient by factor of 2 or less, well within the uncertainty of the methods used to estimate the mass transfer coefficients (*17*) and uptake velocities (*12*).

Removal efficiencies calculated from the LINX II data do not exhibit a consistent relationship to catchment land use (Fig. 1D), but they are negatively correlated with stream nitrate concentration (Fig. 2). In one of the most notable findings to come out of the LINX II study, a similar negative correlation was observed between nitrate uptake velocity and stream nitrate concentration (*12*). As noted by Mulholland *et al.*, increasing nitrate load to a stream could therefore reduce the nitrate uptake velocity and elicit "a disproportionate increase in the fraction of nitrate that is exported to receiving waters" (*12*). Our hypothesis provides a mechanistic explanation for this key LINX II finding: Uptake velocities are highest in streams with low nitrate concentration because, under such conditions, all nitrate transported to the streambed by turbulence is removed by assimilation and denitrification ($\alpha_{tot} \approx 1$ when [NO_3^-] < 10^{-3} mol m^{-3}) (Fig. 2A), and the nitrate uptake velocity is limited by mass transfer from the stream to the streambed ($v_{f,tot} \approx k_m$). With increasing nitrate concentration, a smaller fraction of nitrate transported to the streambed is removed (α_{tot} declines with increasing [NO_3^-], presumably because sediment-associated autotrophic and heterotrophic organisms are progressively growth-limited by something other than nitrate) and nitrate uptake in the streambed is inefficient ($v_{f,tot} << k_m$). Denitrification efficiencies α_{den} calculated from the LINX II data set follow a similar trend (compare panels A and B in Fig. 2). Across all stream sites sampled in the LINX II study, the denitrification efficiency is a roughly constant fraction of the total efficiency ($\alpha_{den} \approx 0.14\alpha_{tot}$) (*20*).

Our hypothesis also implies a simple scaling relationship for the fraction of nitrate removed ($0 \le f \le 1$) over a stream reach of length L (units of meters) (*21*)

$$f = 1 - \exp\left[-0.17\alpha\sqrt{\frac{f_D}{8}}\left(\frac{L}{h}\right)\mathrm{Sc}^{-2/3}\right] \quad (3)$$

If the goal is to enhance potential nitrate removal by manipulating stream physics (e.g.,

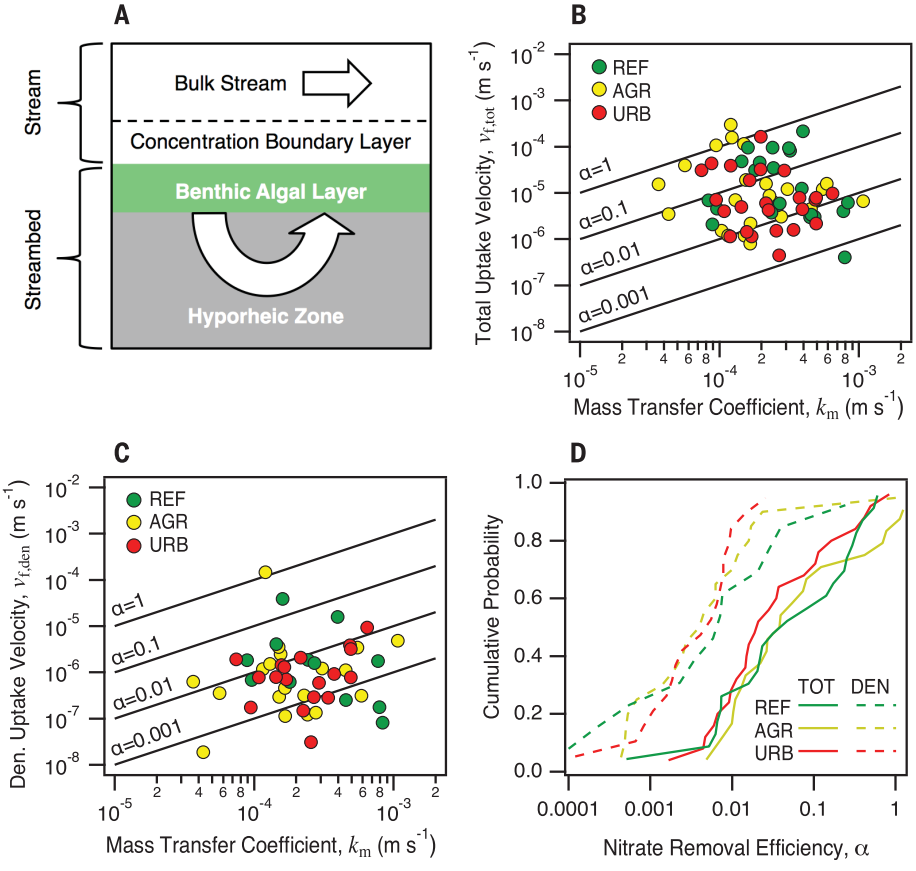


Fig. 1. Stream turbulence imposes an upper limit on nitrate uptake by assimilation and denitrification. (**A**) Conceptual model of how nitrate is transported from the bulk stream, across the concentration boundary layer, and into the streambed where it is assimilated and denitrified in the benthic algal layer and hyporheic zone. (**B**) Total uptake velocities (accounting for both nitrate assimilation and denitrification) measured during the LINX II field campaign, plotted against mass transfer coefficients calculated from Eq. 2A. Colors denote surrounding land use [reference (REF), agriculture (AGR), or urban (URB)]. (**C**) Same as (B), except denitrification uptake velocities are plotted on the vertical axis. (**D**) Empirical cumulative distributions of total (solid curves) and denitrification (dashed curves) efficiencies by land-use type. Efficiencies were calculated from the ratio of measured uptake velocities and site-specific values of the mass transfer coefficient calculated from Eq. 2A.

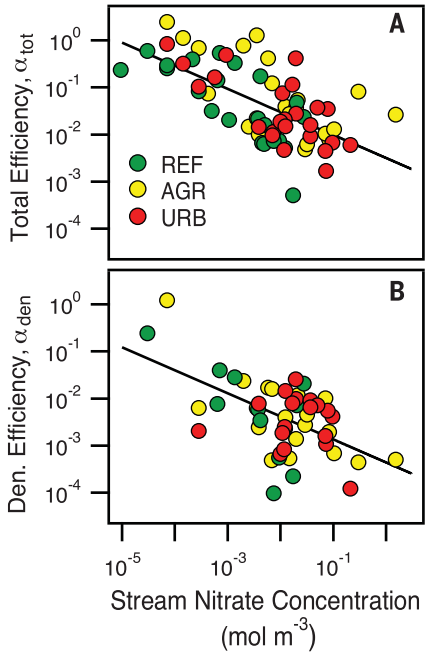


Fig. 2. The turbulence-imposed upper limit on nitrate uptake is observed in streams with low nitrate concentration. (**A**) The fraction of nitrate removed in the streambed by assimilation and denitrification is negatively correlated with stream nitrate concentration (coefficient of determination $r^2 = 0.41$, $P < 0.01$) and approaches 100% ($\alpha_{tot} = 1$) when nitrate concentrations are low ($[NO_3^-] < 10^{-3}$ mol m^{-3}). (**B**) The fraction of nitrate removed in the streambed by denitrification is also negatively correlated with stream nitrate concentration ($r^2 = 0.32$, $P < 0.01$). Lines represent least-squares linear regressions of log-transformed efficiency against log-transformed nitrate concentration: $\log_{10}\alpha = a + b\log_{10}[NO_3^-]$, where the constants are $a = -2.5 \pm 0.18$ and $b = -0.49 \pm 0.07$ for α_{tot} and $a = -3.36 \pm 0.22$ and $b = -0.49 \pm 0.11$ for α_{den}.

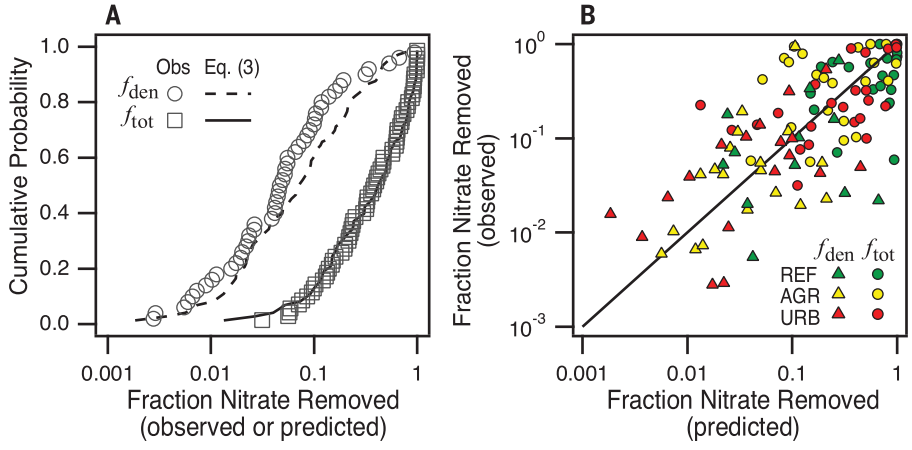


Fig. 3. A test of the scaling law derived in this study. (**A**) Empirical cumulative distributions of the observed (symbols) and predicted (curves) fraction of nitrate removed at LINX II sites by both denitrification and assimilation (f_{tot}) or denitrification alone (f_{den}). Predicted values of f_{tot} and f_{den} were calculated from Eq. 3 after substituting the linear regression models for α_{tot} and α_{den} (Fig. 2) and site-specific values of the shear velocity, stream velocity, reach length, average depth, and stream nitrate concentration (LINX II data tabulated in the supplementary materials). (**B**) Same data as in (A), but plotted so that the observed and predicted values of f_{tot} and f_{den} can be compared on a site-by-site basis. The diagonal line represents a one-to-one relationship.

through stream restoration), Eq. 3 indicates that the Darcy-Weisbach friction factor $f_D = 8u_*^2/U^2$ [where U (units of meters per second) is the average velocity of the stream] and the length-to-depth ratio L/h should be maximized, for instance, using conventional hydraulic relationships (*22*). When stream nitrate concentrations are low (i.e., $[NO_3^-] < 10^{-3}$ mol m^{-3}) nitrate removal is mass transfer–limited and therefore the removal efficiencies can be approximated by the following fixed constants: $\alpha_{tot} \approx 1$ and $\alpha_{den} \approx 0.14$ (Fig. 2) (*20*). For stream nitrate concentrations above this threshold, the results in Fig. 2 imply that nitrate uptake is rate-limited by nitrogen cycling and transport within the streambed rather than by turbulent transport of nitrogen from the stream to the streambed. Under these conditions, several options are available for estimating α_{tot} and α_{den}. The simplest involves substituting into Eq. 3 the linear correlations between log-transformed efficiency and log-transformed nitrate (see lines in Fig. 2). When applied to the entire LINX II data set, this approach closely reproduces empirical distributions of nitrate removal by assimilation and denitrification (f_{tot}) but overestimates nitrate removal by denitrification alone (f_{den}) (Fig. 3A). This method also performs poorly when evaluated on a site-by-site basis (Fig. 3B, Nash-Sutcliff efficiency $E = -0.3$ and 0.0 for f_{tot} and f_{den}, respectively, where $E = 1$ is a perfect model fit and $E < 0$ is worse than the mean), suggesting that much scope exists for model improvement when $\alpha << 1$. One promising approach along these lines involves coupling surface renewal theory for turbulent mass transport above the streambed with process-based models of nitrogen cycling and transport in the benthic algal layer and hyporheic zone (*18*, *23*, *24*). By incorporating Eq. 3 into stream network models [such as the one recently prepared for the Mississippi River basin (*25*, *26*)], the resulting estimates for α can be scaled up to assess the fate and transport of nitrogen pollution at reach, catchment, continental, and global scales.

REFERENCES AND NOTES

1. J. N. Galloway *et al.*, *Biogeochem.* **70**, 153–226 (2004).
2. A. Marzadri, M. M. Dee, D. Tonina, A. Bellin, J. L. Tank, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4330–4335 (2017).
3. W.-J. Cai *et al.*, *Nat. Geosci.* **4**, 766–770 (2011).
4. U.S. Environmental Protection Agency (EPA), Water Quality Assessment and Total Maximum Daily Loads Information (ATTAINS); https://catalog.data.gov/dataset/water-quality-assessment-and-total-maximum-daily-loads-information-attains-1bfd6.
5. W. Steffen *et al.*, *Science* **347**, 1259855 (2015).
6. National Academy of Engineering, 14 Grand Challenges for Engineering in the 21st Century: Manage the Nitrogen Cycle; www.engineeringchallenges.org/challenges/nitrogen.aspx.
7. R. O. Hall Jr. *et al.*, *Limnol. Oceanogr.* **54**, 653–665 (2009).
8. F. Boano *et al.*, *Rev. Geophys.* **52**, 603–679 (2014).
9. J. J. Beaulieu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 214–219 (2011).
10. R. O. Hall, M. A. Baker, C. D. Arp, B. J. Koch, *Limnol. Oceanogr.* **54**, 2128–2142 (2009).
11. S. H. Ensign, M. W. Doyle, *J. Geophys. Res.* **111**, G04009 (2006).
12. P. J. Mulholland *et al.*, *Nature* **452**, 202–205 (2008).
13. P. J. Mulholland *et al.*, *Limnol. Oceanogr.* **54**, 666–680 (2009).
14. K. E. Bencala, R. A. Walters, *Water Resour. Res.* **19**, 718–724 (1983).
15. A. Marzadri, D. Tonina, A. Bellin, J. L. Tank, *Geophys. Res. Lett.* **41**, 5484–5491 (2014).
16. M. Hondzo, *Water Resour. Res.* **34**, 3525–3533 (1998).
17. S. B. Grant, I. Marusic, *Environ. Sci. Technol.* **45**, 7107–7113 (2011).
18. See supplementary materials for details.
19. B. L. O'Connor, M. Hondzo, *Limnol. Oceanogr.* **53**, 566–578 (2008).
20. The linear regressions presented in Fig. 2 (see legend) can be expressed as $\alpha_{den} = 10^{-3.36}[NO_3^-]^{-0.49}$ and $\alpha_{tot} = 10^{-2.5}[NO_3^-]^{-0.49}$. The claim $\alpha_{den}/\alpha_{tot} \approx 0.14$ follows directly from taking the ratio of these two power laws.
21. We derived Eq. 3 by performing mass balance over a stream reach, assuming steady uniform flow: $f = 1 - \exp(-v_t/H_L)$, where $H_L = Uh/L$ is the hydraulic loading rate of the stream. Equation 3 follows by substituting Eqs. 1A and 2A.
22. R. Ferguson, *Water Resour. Res.* **43**, W05427 (2007).
23. J. D. Gomez-Velez, J. W. Harvey, *Geophys. Res. Lett.* **41**, 6403–6412 (2014).
24. M. Azizian *et al.*, *Water Resour. Res.* **53**, 3941–3967 (2017).
25. B. A. Kiel, M. B. Cardenas, *Nat. Geosci.* **7**, 413–417 (2014).
26. J. D. Gomez-Velez, J. W. Harvey, M. B. Cardenas, B. Kiel, *Nat. Geosci.* **8**, 941–945 (2015).

ACKNOWLEDGMENTS

We thank M. Gooseff and A. Mehring for valuable feedback and the LINX II researchers for data access. **Funding:** Financial support was provided by the U.S. NSF Partnerships for International Research and Education (grant OISE-1243543) and the University of California Office of the President Multicampus Research Program Initiatives (award MRP-17-455083). **Author contributions:** S.B.G. conceived of the study and drafted the article; M.A. curated and analyzed the LINX II data set; F.B. and P.C. contributed text on hyporheic exchange and nitrogen cycling, respectively; and M.A.R. helped frame the article. All authors provided edits. **Competing interests:** None declared. **Data and materials availability:** The supplementary materials include a derivation of Eq. 1, data reduction methods, an example of how the theory presented here can be coupled to process-based models of nitrogen cycling and transport in the hyporheic zone, and a compilation of the LINX II data used in this study.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1266/suppl/DC1
Materials and Methods
Supplementary Text
Table S1
References (*27*–*32*)

29 August 2017; accepted 31 January 2018
10.1126/science.aap8074

NEURODEVELOPMENT

Astrocyte-derived interleukin-33 promotes microglial synapse engulfment and neural circuit development

Ilia D. Vainchtein,[1]* Gregory Chin,[1]* Frances S. Cho,[5,7] Kevin W. Kelley,[1] John G. Miller,[1] Elliott C. Chien,[1] Shane A. Liddelow,[8]† Phi T. Nguyen,[1,6] Hiromi Nakao-Inoue,[1] Leah C. Dorman,[1,5] Omar Akil,[3] Satoru Joshita,[9,10] Ben A. Barres,[8] Jeanne T. Paz,[4,7] Ari B. Molofsky,[2]‡ Anna V. Molofsky[1]‡

Neuronal synapse formation and remodeling are essential to central nervous system (CNS) development and are dysfunctional in neurodevelopmental diseases. Innate immune signals regulate tissue remodeling in the periphery, but how this affects CNS synapses is largely unknown. Here, we show that the interleukin-1 family cytokine interleukin-33 (IL-33) is produced by developing astrocytes and is developmentally required for normal synapse numbers and neural circuit function in the spinal cord and thalamus. We find that IL-33 signals primarily to microglia under physiologic conditions, that it promotes microglial synapse engulfment, and that it can drive microglial-dependent synapse depletion in vivo. These data reveal a cytokine-mediated mechanism required to maintain synapse homeostasis during CNS development.

Neuronal synapse formation depends on a complex interplay between neurons and their glial support cells. Astrocytes provide structural, metabolic, and trophic support for neurons (*1*, *2*). Gray-matter astrocytes are in intimate contact with neuronal synapses and are poised to sense local neuronal cues. In contrast, microglia are the primary immune cells of the central nervous system (CNS) parenchyma. Microglia regulate multiple phases of developmental circuit refinement (*3*, *4*), both inducing synapse formation (*5*, *6*) and promoting synapse engulfment (*7*, *8*), in part via complement, an effector arm of the innate immune system (*9*). Excess complement activity has been implicated in schizophrenia, a neurodevelopmental disorder that includes cortical gray matter thinning and synapse loss (*10*), suggesting that microglial synapse engulfment may have broad implications for neuropsychiatric disease.

[1]Department of Psychiatry/Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. [2]Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, USA. [3]Department of Otolaryngology, University of California, San Francisco, San Francisco, CA, USA. [4]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [5]Neuroscience Graduate Program, University of California, San Francisco, San Francisco, CA, USA. [6]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA. [7]Gladstone Institute of Neurological Disease, San Francisco, CA 94158, USA. [8]Department of Neurobiology, Stanford University, Palo Alto, CA, USA. [9]Department of Medicine, Division of Gastroenterology and Hepatology, Shinshu University School of Medicine, Matsumoto, Japan. [10]Research Center for Next Generation Medicine, Shinshu University, Matsumoto, Japan.
*These authors contributed equally to this work.
†Present address: Neuroscience Institute and Department of Neuroscience and Physiology, New York University, Langone Medical School, New York, NY, USA.
‡Corresponding author. Email: anna.molofsky@ucsf.edu (A.V.M.); ari.molofsky@ucsf.edu (A.B.M.)

Despite the emerging roles of astrocytes and microglia in neuronal synapse formation and remodeling, how they coordinate synaptic homeostasis in vivo remains obscure. Interleukin-33 (IL-33) is an IL-1 family member with well-described roles as a cellular alarmin released from nuclear stores after tissue damage, including in spinal cord injury (*11*, *12*), stroke (*13*), and Alzheimer's disease (*14*). Whereas many cytokines are primarily defined by their roles in inflammation and disease (e.g., IL-1, tumor necrosis factor–α, or IL-6), IL-33 also promotes homeostatic tissue development and remodeling (*15*). The CNS undergoes extensive synapse remodeling during postnatal brain development, but a role for IL-33 or other stromal-derived cytokines is unknown. Here, we report that IL-33 is produced postnatally by synapse-associated astrocytes, is required for synaptic development in the thalamus and spinal cord, and signals to microglia to promote increased synaptic engulfment. These findings reveal a physiologic requirement for cytokine-mediated immune signaling in brain development.

We previously developed methods to identify functionally heterogeneous astrocytes by expression profiling of distinct CNS regions (*16*). In an RNA-sequencing screen of developing forebrain astrocytes (P9) (flow sorted using an *Aldh1l1*^{eGFP} reporter), we identified the cytokine IL-33 as a candidate that is both astrocyte-enriched and heterogeneously expressed by astrocytes throughout the CNS (fig. S1, A to C). We confirmed astrocyte-specific developmental expression of IL-33 in spinal cord and thalamus using a nuclear-localized IL-33 reporter (*Il33*$^{mCherry/+}$) (Fig. 1A) and validated these findings with flow cytometry and protein immunostaining (fig. S2). By adulthood, a subset of oligodendrocytes also colabeled with IL-33 (fig. S2, C to E), consistent with previous reports (*11*). Thus, astrocytes are the primary source of IL-33 during postnatal synapse maturation.

Although most IL-33–positive cells were astrocytes, not all developing astrocytes expressed IL-33, and this number increased in the early postnatal period (fig. S3) (*17*). In fact, IL-33 was detected only in gray matter, where most synapses are located (Fig. 1B and figs. S2H and S3D). In the thalamus, which receives regionally distinct sensory synaptic inputs, IL-33 expression in the visual nucleus (dLGN) increased coincident with eye opening [postnatal days 12 to 14 (P12 to P14)] (Fig. 1, C and D). Removal of afferent sensory synapses by enucleation at birth prevented this developmental increase in IL-33 expression (Fig. 1, E and F), whereas dark rearing, in which synapse maturation is largely preserved (*18*), had no effect. Molecular profiling of IL-33–positive astrocytes in both thalamus and spinal cord (Fig. 1, G and H) revealed a negative correlation with white-matter astrocyte markers (*Gfap* and *Vimentin*), enrichment for genes involved in astrocyte synaptic functions (connexin-30/*Gjb6*) (*19*), and enrichment in G protein–coupled and neurotransmitter receptors (e.g., *Adora2b* and *Adra2a*) (tables S1 to S3 and data S1). Together, these data demonstrate that IL-33 expression is correlated with synaptic maturation and marks a subset of astrocytes potentially sensitive to synaptic cues, raising the question of whether IL-33 plays a role in synapse development.

To determine whether IL-33 regulates neural circuit development and function, we examined the effect of IL-33 deletion on synapse numbers and circuit activity. In the thalamus, a region with high IL-33 expression, an intrathalamic circuit between the ventrobasal nucleus (VB) and the reticular nucleus of the thalamus (RT) displays spontaneous oscillatory activity that can also be evoked by stimulating the internal capsule that contains cortical afferents (*20*, *21*). We quantified this oscillatory activity in slices from young adult mice (P30 to P40), which revealed enhanced evoked activity in response to stimulation (Fig. 2, A and B, and fig. S4, A and B), as well as elevated spontaneous firing in the absence of IL-33 (Fig. 2C and fig. S4C). This increase could result at least in part from enhanced numbers of glutamatergic synapses. To investigate this hypothesis, we performed whole-cell patch-clamp recordings of VB neurons to quantify miniature excitatory postsynaptic currents (mEPSCs) (Fig. 2D). We found that the frequency of mEPSCs was enhanced in VB neurons from IL-33–deficient mice, whereas the amplitude and the kinetics were unchanged (fig. S4D). Together, these results suggest that IL-33 deficiency leads to excess excitatory synapses and a hyperexcitable intrathalamic circuit.

In the spinal cord, α-motor neurons (α-MN) are the primary outputs of the sensorimotor circuit and receive inputs from excitatory (VGLUT2$^+$) and inhibitory (VGAT$^+$) interneurons (Fig. 2E) (*22*). We conditionally deleted IL-33 from astrocytes (*hGFAPcre*) (*16*) (fig. S5A) and found increased numbers of excitatory and inhibitory

inputs onto α-MN at P30; global deletion of *Il1rl1* (*ST2*) (Fig. 2, F to I) or *Il33* (fig. S5, B and C) phenocopied this finding. Neuronal soma size, interneuron numbers, and oligodendrocyte numbers were unchanged (fig. S5, D to F). However, by adulthood, IL-33 deficiency led to increased gray-matter expression of glial fibrillary acidic protein (GFAP) (fig. S5, G and H), a marker of tissue stress. We also found that $Il33^{-/-}$ animals had deficits in acoustic startle response, a sensorimotor reflex mediated by motor neurons in the brainstem and spinal cord (Fig. 2, J and K) (*23*). Auditory acuity and gross motor performance were normal (fig. S5, I and J). Taken together, these data demonstrate that IL-33 is required for normal synapse numbers and circuit function in the thalamus and spinal cord.

To determine the cellular targets of IL-33 signaling, we first quantified expression of its obligate co-receptor IL1RL1 (ST2) (*15*). We detected *Il1rl1* in microglia by RNA sequencing (7.1 ± 2.1 fragments per kilobase of transcript per million mapped reads) and by quantitative polymerase chain reaction (qPCR), in contrast to astrocytes, neurons, or the lineage-negative fraction (Fig. 3A). The transcriptome of acutely isolated microglia from $Il33^{-/-}$ animals revealed 483 significantly altered transcripts, including reduced expression of nuclear factor κB (NF-κB) targets (e.g., *Tnf*, *Nfkbia*, *Nfkbiz*, and *Tnfaip3*) (Fig. 3, B and C; fig. S6A; and data S2), consistent with diminished NF-κB signaling (*24*). The transcriptome of $Il33^{-/-}$ astrocytes was unchanged (fig. S6B), arguing against cell-autonomous roles of IL-33 in this context. These data demonstrate physiologic signaling by IL-33 to microglia during brain development, raising the question of whether it promotes physiologic microglial functions.

Given the increased synapse numbers in IL-33–deficient animals, we investigated whether IL-33 is required for microglial synapse engulfment. We detected engulfed PSD-95^{+} synaptic puncta within spinal cord microglia throughout development, as in other CNS regions (*7*, *8*), and found decreased engulfment in microglia from $Il33^{-/-}$ animals (P15) (Fig. 3D). This was further validated by dye labeling of spinal cord motor neurons, which revealed fewer dye-filled microglia in $Il33^{-/-}$ (fig. S7). Conversely, local injection of IL-33 increased PSD-95 within microglia in both spinal cord (Fig. 3E) and thalamus (fig. S8, A and B) and altered markers consistent

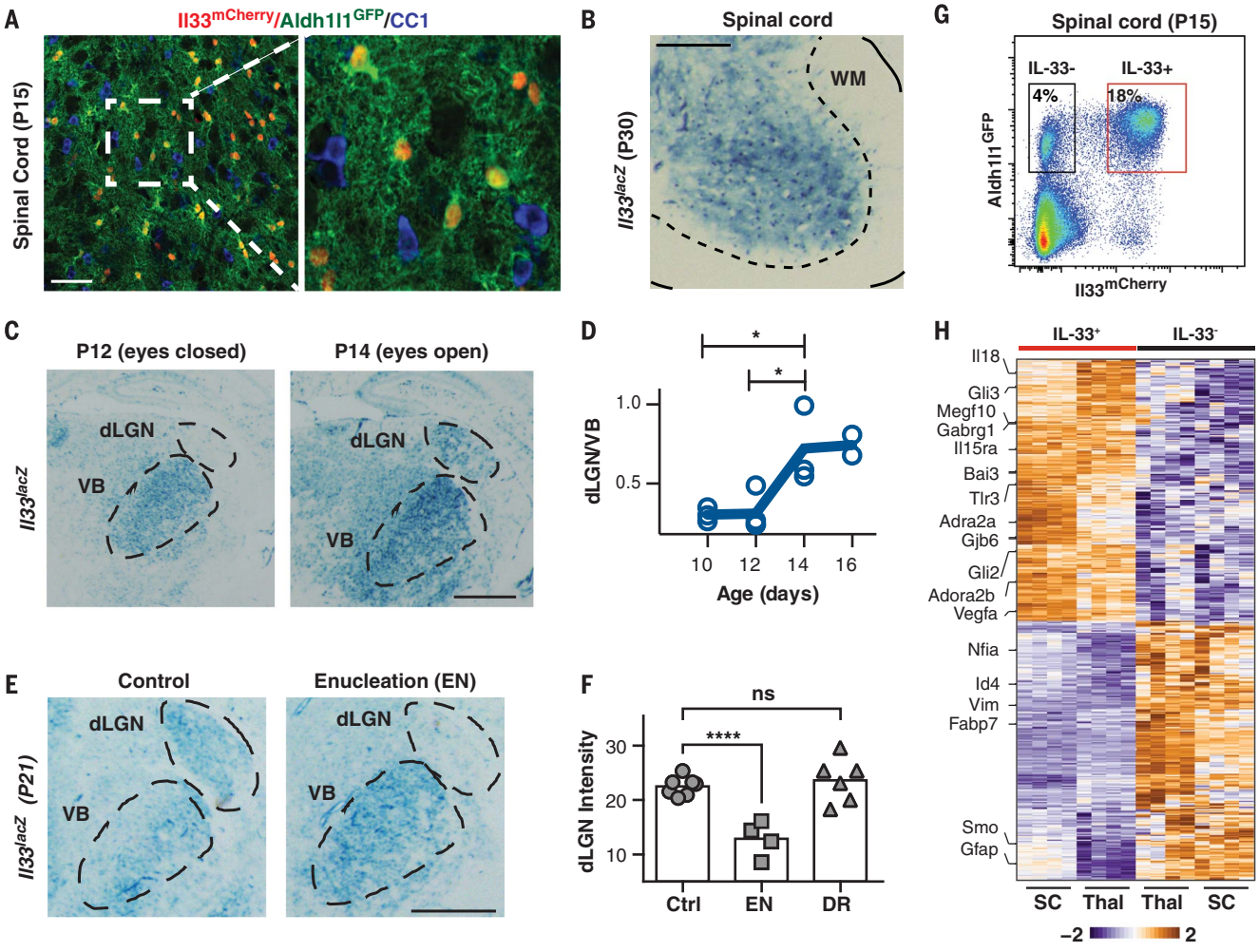


Fig. 1. IL-33 is developmentally induced in synapse-associated astrocytes. (**A**) Representative image of Il33^{mCherry} with *Aldh1l1*^{GFP} astrocytes and oligodendrocyte marker CC1 in spinal cord ventral horn (scale bar, 50 μm). (**B**) Gray matter restricted expression of *Il33*^{lacZ} in the spinal cord at P30 (scale bar, 0.5 mm). (**C** and **D**) *Il33*^{lacZ} increases in the visual thalamus (dLGN) during eye opening, normalized to sensorimotor thalamus (VB) (scale bar, 0.5 mm). (**E**) Representative images of *Il33*^{lacZ} in P21 thalamus in littermate controls and after perinatal enucleation (scale bar, 0.5 mm). (**F**) *Il33*^{lacZ} mean pixel intensity in dLGN. (**G**) Representative flow plot of spinal cord from *Il33*^{mCherry}/*Aldh1l1*^{GFP} mice at P15 with sorting gates indicated. (**H**) Heat map of the top 444 differentially expressed genes in Il33-mCherry^{+} versus mCherry^{-} astrocytes in spinal cord and thalamus (fold change > 2; adjusted *P* value < 0.05), select candidates highlighted. One-way analysis of variance (ANOVA) with Tukey's post hoc comparison or Student's *t* test. All points represent independent biological replicates. *P < 0.05, ****P < 0.0001.

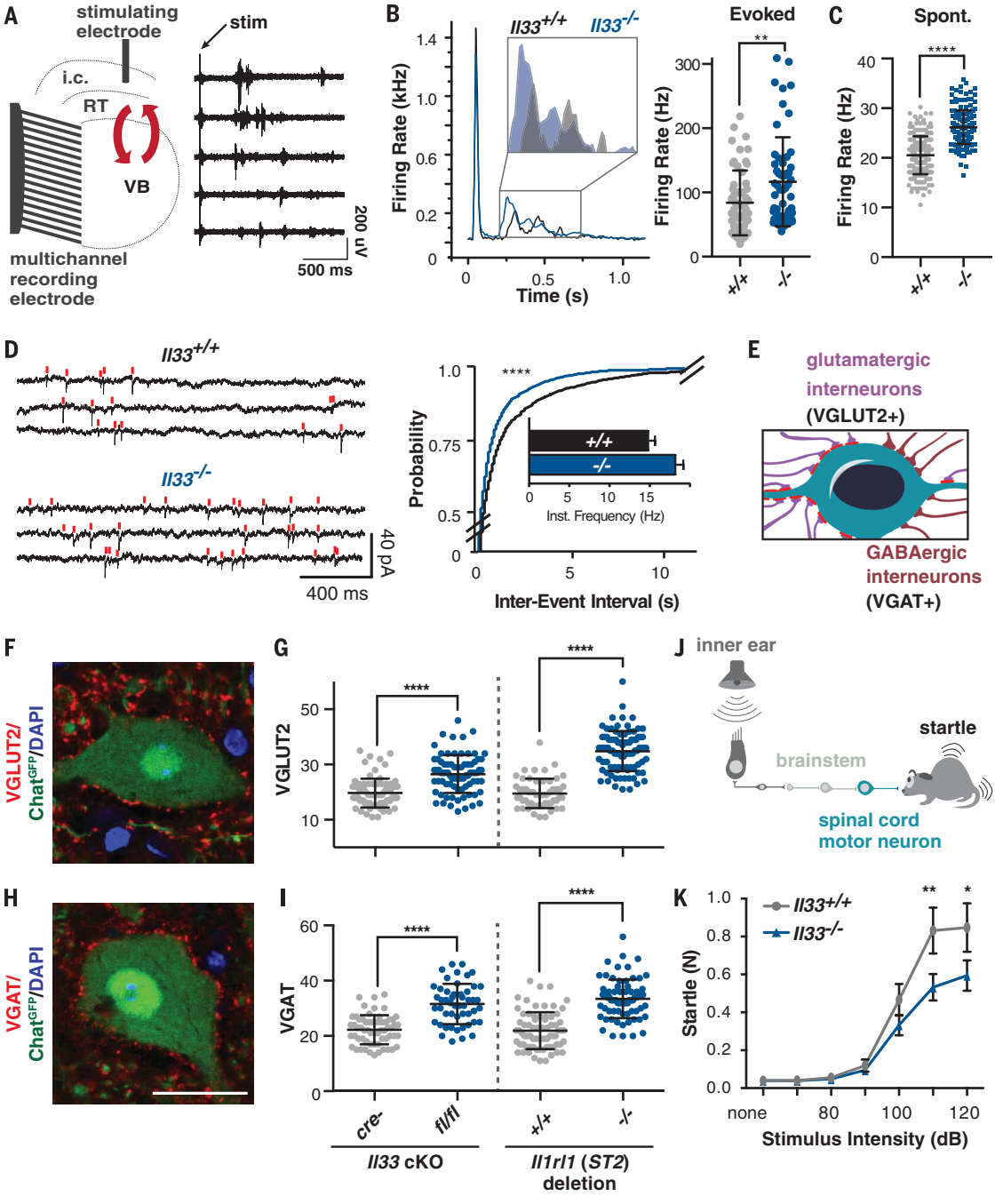


Fig. 2. IL-33 deficiency leads to excess synapses and abnormal thalamic and sensorimotor circuit function. (**A**) Schematic of extracellular recording setup to measure circuit activity between ventrobasal (VB) and reticular thalamic (RT) nuclei, with representative recording showing activity in five channels after stimulation of the internal capsule (i.c.) that contains cortical afferents. Red arrows indicate reciprocal VB-RT connections. (**B**) Average traces and quantification of mean firing rates reveal higher evoked firing in *Il33*$^{-/-}$. (**C**) Quantification of mean firing rates in the absence of stimulation reveals increased spontaneous firing in *Il33*$^{-/-}$. (**D**) Representative traces and quantification of intracellular patch-clamp recordings from neurons in the VB show increased miniature excitatory postsynaptic currents (mEPSCs) in *Il33*$^{-/-}$. (**E**) Schematic of motor neuron synaptic afferents. (**F** and **G**) Representative image and quantification of excitatory inputs per motor neuron after conditional deletion of *Il33* (*hGFAPcre*) or global deletion of *Il1rl1*. (**H** and **I**) Inhibitory (VGAT^{+}) inputs in the same mice (scale bar, 25 μm). (**J**) Schematic of startle pathway. (**K**) Impaired sensorimotor startle in *Il33*$^{-/-}$ animals. Data in (B) and (C) from wild type: n = 6 to 8 slices, 2 mice. Knockout (KO): n = 14 to 15 slices, 3 mice; points are individual recordings. Data in (B) analyzed by Mann-Whitney and (C) with Student's t test. Data in (D) from n = 23 to 25 cells and 3 to 4 mice per group, analyzed by Kolmogorov-Smirnov test. Data in (F) to (I) from n = 3 animals, >75 neurons per genotype, Student's t test; points are individual neurons. (K) is n = 12/group, two-way ANOVA with Sidak's multiple comparisons. *P < 0.05, **P < 0.01, ****P < 0.0001. (B) to (I) are mean ± SD; (K) is mean ± SEM.

with microglial activation, including IL1RL1-dependent down-regulation of P2Y12 (fig. S9, A to C) (*25*). In vitro, IL-33 promoted synaptosome engulfment by purified microglia, whereas the canonical IL-1 family member IL-1β had no effect (fig. S9, D and E). In vivo, injection of IL-33 into the developing spinal cord led to twofold depletion of excitatory synapses (colocalized VGLUT2/PSD-95), whereas conditional deletion of IL1RL1 from microglia partly reversed this effect (*Cx3cr1cre:Il1rl1*$^{fl/fl}$) (Fig. 3, F and G). In comparison, global loss of *Il1rl1* completely reversed IL-33–dependent synapse depletion in spinal cord (fig. S10) and thalamus (fig. S8, C and D), suggesting that nonmicroglial sources of ST2 could also contribute. These data indicate that IL-33 regulates synapse numbers in vivo at least in part via IL1RL1 receptor–mediated signaling in microglia.

Our data reveal a mechanism of astrocyte-microglial communication that is required for synapse homeostasis during CNS development. We propose that astrocyte-derived IL-33 serves as a rheostat, helping to tune microglial synapse

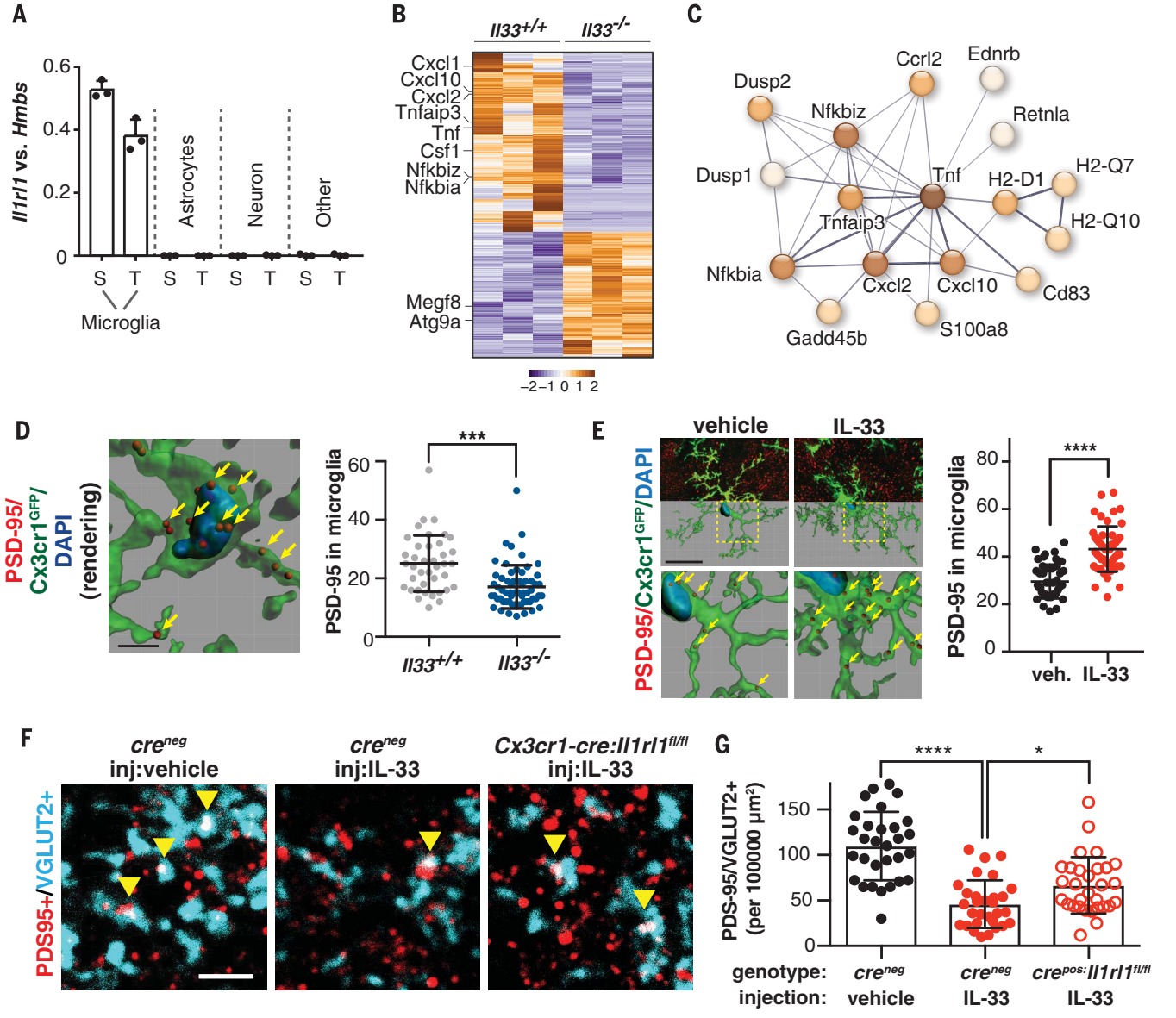


Fig. 3. IL-33 drives microglial synapse engulfment during development. (**A**) Expression of *Il1rl1* by qPCR of flow-sorted populations (S, spinal cord; T, thalamus). (**B**) A total of 484 differentially expressed genes in spinal cord microglia at adjusted *P* value < 0.05. (**C**) Functionally associated gene clustering (STRING) identifies immune genes enriched in wild-type versus *Il33*$^{-/-}$ microglia. (**D**) PSD-95 puncta within microglia (yellow arrows) after IL-33 deletion (scale bar, 4 µM). (**E**) Representative image and quantification of engulfed PSD-95 in vehicle or IL-33 injected spinal cord (scale bar, 20 µm). (**F** and **G**) Colocalized pre- and postsynaptic puncta in spinal cord ventral horn at P14 (yellow arrows) after IL-33 injection into control mice or in littermates with conditional deletion of *Il1rl1* (*Cx3cr1cre*) (scale bar, 3 µm). Points in (A) represent mice; in (D) to (G), individual microglia from *n* = 3 to 5 animals per group; in (G), images from *n* = 5 mice per group. In (D) and (E), Student's *t* test and in (G), a one-way ANOVA with Tukey's post hoc comparison. **P* < 0.05, ****P* < 0.001, *****P* < 0.0001; all data are mean ± SD.

engulfment during neural circuit maturation and remodeling (fig. S11). Key unanswered questions include the nature of the cues that induce astrocyte *Il33* expression, the mechanism of IL-33 release, and the signals downstream of IL-33 that promote microglial function. These data also raise the broader question of how this process affects neural circuit function. Synapses are the most tightly regulated variable in the developing CNS (*26*) and are a primary locus of dysfunction in neurodevelopmental diseases. *Il33* is one of five genes that molecularly distinguish astrocytes from neural progenitors in the developing human forebrain (*27*), suggesting possibly conserved roles in the human CNS. Defining whether signals like IL-33 are permissive or instructive, promiscuous or synapse specific, is a first step toward understanding how neural circuits remodel during development and under stress.

REFERENCES AND NOTES

1. A. V. Molofsky *et al.*, *Genes Dev.* **26**, 891–907 (2012).
2. L. E. Clarke, B. A. Barres, *Nat. Rev. Neurosci.* **14**, 311–321 (2013).
3. R. M. Ransohoff, A. E. Cardona, *Nature* **468**, 253–262 (2010).
4. M. W. Salter, S. Beggs, *Cell* **158**, 15–24 (2014).

5. A. Miyamoto *et al*., *Nat. Commun.* **7**, 12540 (2016).
6. C. N. Parkhurst *et al*., *Cell* **155**, 1596–1609 (2013).
7. R. C. Paolicelli *et al*., *Science* **333**, 1456–1458 (2011).
8. D. P. Schafer *et al*., *Neuron* **74**, 691–705 (2012).
9. B. Stevens *et al*., *Cell* **131**, 1164–1178 (2007).
10. A. Sekar *et al*., *Nature* **530**, 177–183 (2016).
11. S. P. Gadani, J. T. Walsh, I. Smirnov, J. Zheng, J. Kipnis, *Neuron* **85**, 703–709 (2015).
12. Y. Pomeshchik *et al*., *Brain Behav. Immun.* **44**, 68–81 (2015).
13. Y. Luo *et al*., *Brain Res.* **1597**, 86–94 (2015).
14. A. K. Y. Fu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2705–E2713 (2016).
15. A. B. Molofsky, A. K. Savage, R. M. Locksley, *Immunity* **42**, 1005–1019 (2015).
16. A. V. Molofsky *et al*., *Nature* **509**, 189–194 (2014).
17. G. Wicher, E. Husic, G. Nilsson, K. Forsberg-Nilsson, *Neurosci. Lett.* **555**, 171–176 (2013).
18. B. M. Hooks, C. Chen, *Neuron* **52**, 281–291 (2006).
19. U. Pannasch *et al*., *Nat. Neurosci.* **17**, 549–558 (2014).
20. J. T. Paz, C. A. Christian, I. Parada, D. A. Prince, J. R. Huguenard, *J. Neurosci.* **30**, 5465–5479 (2010).
21. H. Lui *et al*., *Cell* **165**, 921–935 (2016).
22. S. Arber, *Neuron* **74**, 975–989 (2012).
23. M. Koch, *Prog. Neurobiol.* **59**, 107–128 (1999).
24. J. Ruland, *Nat. Immunol.* **12**, 709–714 (2011).
25. S. E. Haynes *et al*., *Nat. Neurosci.* **9**, 1512–1519 (2006).
26. R. Chovatiya, R. Medzhitov, *Mol. Cell* **54**, 281–288 (2014).
27. A. A. Pollen *et al*., *Cell* **163**, 55–67 (2015).

ACKNOWLEDGMENTS

We are grateful to the Molofsky laboratories, the Poskanzer laboratory, D. H. Rowitch, R. M. Locksley, and J. R. Chan for helpful comments on the manuscript. Thanks to the J. Huguenard laboratory for the thalamic network analysis code, J. P. Girard for *Il33*^{lacZ} mice, R. T. Lee for *Il33*$^{fl/fl}$ and *Il1rl1*$^{fl/fl}$, M. Colonna and the Mucosal Immunology Studies Team (MIST) for *Il33*$^{H2B\text{-}mCherry}$, D. Julius for the P2Y12 antibody, and the Gladstone Genomics and Behavioral Cores (P30NS065780) for technical contributions. **Funding:** A.V.M. is supported by a Pew Scholars Award, NIMH (K08MH104417), the Brain and Behavior Research Foundation, and the Burroughs Wellcome Fund. A.B.M. is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K08DK101604) and the Larry L. Hillblom Foundation. J.T.P. is supported by the National Institute of Neurological Disorders and Stroke (R01NS096369). S.A.L. is supported by the Australian National Health and Medical Research Council (GNT1052961) and the Glenn Foundation Glenn Award. F.S.C. (NSF 1144247) and P.T.N. (NSF 1650113) were supported by Graduate Student Fellowships from the National Science Foundation. **Author contributions:** I.D.V., G.C., A.V.M., and J.G.M. designed, performed, and analyzed most experiments. E.C.C., H.N.-I., P.T.N., and L.C.D. contributed to experiments and data analysis. F.S.C. and J.T.P. designed, performed, and analyzed the electrophysiology experiments. K.W.K. and I.D.V. designed and performed bioinformatics analyses. S.A.L. performed and analyzed culture experiments under the supervision of B.A.B. O.A. performed and analyzed auditory testing. S.J. generated *Il33*$^{H2B\text{-}mCherry}$ mice. A.V.M. and A.B.M. designed experiments and wrote the manuscript, together with I.D.V., G.C., and other authors. **Competing interests:** None declared. **Data and materials availability:** Supplementary materials contain additional data. All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. RNA-sequencing data are available through GEO no. GSE109354.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1269/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S3
Data S1 and S2
References (*28*–*49*)

8 November 2016; resubmitted 4 December 2017
Accepted 17 January 2018
Published online 1 February 2018
10.1126/science.aal3589

CIRCADIAN RHYTHMS

Rev-erbα dynamically modulates chromatin looping to control circadian gene transcription

Yong Hoon Kim,[1,2,3]* Sajid A. Marhon,[1,2,3]* Yuxiang Zhang,[1,2] David J. Steger,[1,2] Kyoung-Jae Won,[1,2,3]† Mitchell A. Lazar[1,2,3]†

Mammalian physiology exhibits 24-hour cyclicity due to circadian rhythms of gene expression controlled by transcription factors that constitute molecular clocks. Core clock transcription factors bind to the genome at enhancer sequences to regulate circadian gene expression, but not all binding sites are equally functional. We found that in mice, circadian gene expression in the liver is controlled by rhythmic chromatin interactions between enhancers and promoters. Rev-erbα, a core repressive transcription factor of the clock, opposes functional loop formation between Rev-erbα–regulated enhancers and circadian target gene promoters by recruitment of the NCoR-HDAC3 co-repressor complex, histone deacetylation, and eviction of the elongation factor BRD4 and the looping factor MED1. Thus, a repressive arm of the molecular clock operates by rhythmically modulating chromatin loops to control circadian gene transcription.

Circadian rhythms of mammalian physiology are orchestrated by core clock transcription factors (TFs) functioning as activators or repressors in interlocking feedback loops that drive daily oscillations of gene expression (*1*, *2*). These TFs generate circadian rhythms of histone modification at circadian enhancers (*3*–*6*). Enhancer activities are mediated by looping to promoters within insulated topologically associating domains (TADs) (*7*–*9*). Previous studies with cultured cells have described circadian regulation of higher-order chromatin organization (*10*), long-range interchromosomal interactions (*11*), and short-range chromatin loops (*12*) at specific loci, but this has not been examined at a genome-wide level in native tissues under normal physiology.

We performed in situ Hi-C (*13*, *14*) on C57BL/6J mouse livers harvested 12 hours apart, at zeitgeber time 22 (ZT22, 5 a.m.) and ZT10 (5 p.m.), to examine whether chromatin interactions change in a circadian manner. Hi-C identified megabase-size TADs whose boundaries were highly similar to those identified in mouse embryonic stem cells (fig. S1A). This was expected because TADs are largely conserved among different tissues (*7*), and indeed the overall TAD organization was also highly similar between ZT22 and ZT10 (fig. S1B). Within each TAD we also observed submegabase structures (sub-TADs) of different lengths (fig. S1C) that were flanked by CTCF and cohesin (RAD21) and internally occupied by Mediator (MED1) (fig. S1D), as described previously (*14*–*17*). Globally, genomic occupancy of CTCF and cohesin at ZT10 versus ZT22 was very similar at sub-TAD boundaries (fig. S1, E and F).

Because TFs bind at enhancers to regulate genes confined within sub-TAD boundaries (*16*), we next searched for "intra-TAD" interactions occurring within sub-TADs. Of 6510 intra-TAD interactions, only 349 were ZT22-specific, whereas 527 were ZT10-specific (Wilcoxon signed-rank test, $P < 0.001$). For example, the circadian *Npas2* gene exhibited increased intra-TAD interactions at ZT22, including extensive looping between the transcriptional start site and four noncoding regions (E1 to E4) (Fig. 1A). These noncoding regions

[1]Institute for Diabetes, Obesity, and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA. [2]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA. [3]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: wonk@pennmedicine.upenn.edu (K.-J.W.); lazar@pennmedicine.upenn.edu (M.A.L.)

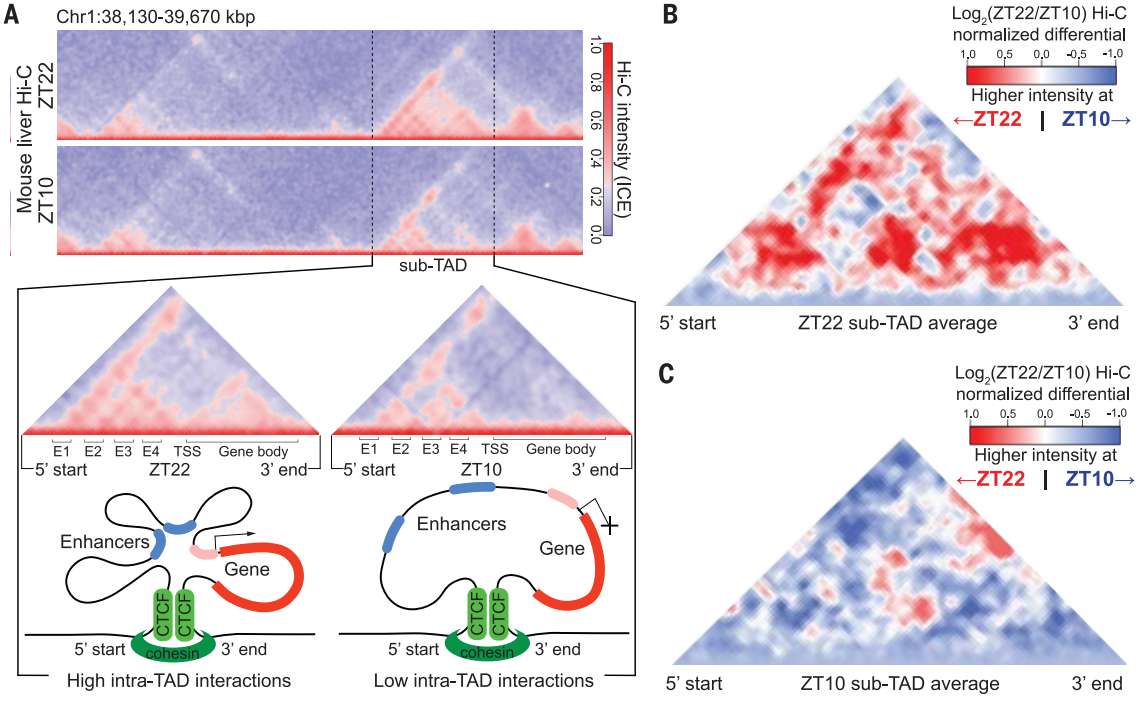


Fig. 1. Circadian sub-TADs undergo rhythmic intra-TAD compaction within stable boundaries. (**A**) Heat maps of ZT22 and ZT10 Hi-C demonstrating circadian intra-TAD interactions within sub-TAD boundaries (dashed lines), as represented by Hi-C intensity normalized by iterative correction and eigenvector decomposition (ICE). The transcriptional start site of the *Npas2* gene (TSS) forms rhythmic intra-TAD loops with upstream enhancers (E1 to E4) as well as with the gene body, as illustrated by schematics below. (**B** and **C**) ZT22 sub-TAD (B) and ZT10 sub-TAD (C) averaged differential changes in intra-TAD interactions visualized as $\log_2$ ratio within size-normalized sub-TAD 5′ and 3′ boundaries (red, higher interaction ratio at ZT22; blue, higher interaction ratio at ZT10).

mapped to regions of divergent transcriptional activity characteristic of enhancer RNA (eRNA) that were similarly regulated (fig. S2A). Intra-TAD interaction was also observed within the gene body, as previously reported at actively transcribed genes (*18*, *19*), and this was enhanced at ZT22 as well (Fig. 1A).

Genome-wide, we identified hundreds of sub-TADs containing circadian genes whose expression has been shown to peak between ZT21 and ZT24 ("ZT22 sub-TADs") or between ZT9 and ZT12 ("ZT10 sub-TADs") (*6*). The overall structure of these sub-TADs (fig. S2B) and the binding of CTCF and RAD21 at their boundaries (fig. S2, C and D) changed very little between ZT22 and ZT10 (fig. S2, B to D). However, ZT22 and ZT10 sub-TADs exhibited greater intra-TAD interactions corresponding to their transcriptional activities (Fig. 1, B and C), whereas noncircadian sub-TADs did not (fig. S3A). Similar conclusions were obtained when the circadian windows were adjusted by 1 hour (fig. S3, B and C). Interactions within gene bodies were also circadian (fig. S3, D to H).

The circadian *Cry1* locus is located within a ZT22 sub-TAD, and Hi-C revealed an interaction between the gene promoter and an intronic enhancer that was increased at ZT22 (fig. S4, A and B), as best visualized by differential analysis of the data (Fig. 2A). The nuclear receptor Rev-erbα, a repressive component of the mammalian clock whose expression peaks at ZT10 to confer circadian expression of genes in the opposite phase (*6*, *20*, *21*), binds at this intronic enhancer (Fig. 2A). The enhancer-promoter (E-P) loop identified in ZT22 Hi-C was confirmed by chromatin conformation capture (3C) experiments (Fig. 2B). Tandem 3C and Rev-erbα chromatin immunoprecipitation (ChIP) at six time points throughout the day revealed that this E-P loop was indeed circadian and in phase with *Cry1* mRNA expression, with a peak that was antiphase to Rev-erbα binding (Fig. 2C). At ZT10, both looping from the Rev-erbα site to the *Cry1* promoter (Fig. 2D) and *Cry1* gene expression (Fig. 2E) were enhanced by knockout (KO) of Rev-erbα, which attenuated the rhythmicity of these parameters over the course of 24 hours (fig. S4C). Reciprocally, ectopic expression of Rev-erbα in liver was sufficient to reduce looping as well as mRNA expression at ZT22 (Fig. 2, F and G), consistent with an active role of Rev-erbα in opposing loop formation. We next performed Hi-C on livers from mice genetically lacking Rev-erbα and harvested at ZT10, which confirmed the enhanced E-P looping at the *Cry1* locus (fig. S4D). Moreover, throughout the genome, intra-TAD interactions that were normally favored at ZT22 were increased in the genetic absence of Rev-erbα at ZT10 (Fig. 2H).

These findings suggested that the ability of Rev-erbα to oppose loop formation is a critical feature at binding sites from which Rev-erbα actively represses transcription. To explore the relationship between regulation of looping and functional repression, we identified Rev-erbα binding sites that directly loop to promoters at ZT22 in the physiological absence of Rev-erbα. E-P loops at Rev-erbα binding sites were defined as "engaged" when transcription of the gene body was repressed at ZT10, or "passive" if not repressed (Fig. 3, A and B). Engaged sites were highly correlated

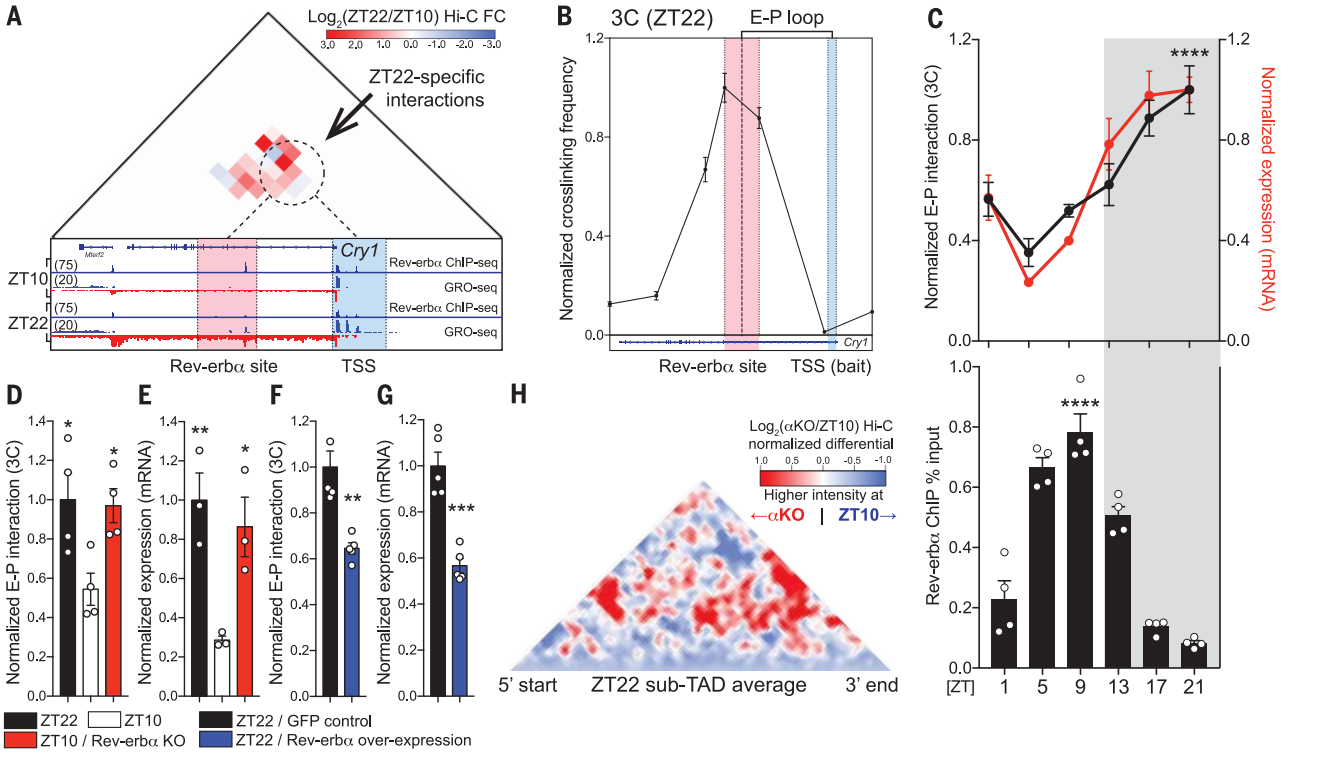


Fig. 2. Rev-erbα causally opposes enhancer-promoter loop formation. (**A**) Differential Hi-C analysis at the *Cry1* locus revealing ZT22-specific interactions, represented as $\log_2$ ratio (ZT22 Hi-C/ZT10 Hi-C). ZT22-specific interactions (dashed circle) occur between a region around the intronic Rev-erbα site (red) and the *Cry1* TSS (blue). Global run-on sequencing (GRO-seq) demonstrates circadian nascent transcription as well as the presence of bidirectional eRNA at the Rev-erbα site at ZT22. (**B**) 3C validation of enhancer-promoter loop (E-P loop) identified at ZT22 between the Rev-erbα site (red) and TSS (blue) (n = 5, mean ± SEM). (**C**) Circadian plot demonstrating *Cry1* E-P loop, mRNA expression, and Rev-erbα ChIP ± SEM [n = 4 or 5, P values shown for 3C and ChIP peaks compared to troughs, one-way analysis of variance (ANOVA) followed by multiple-comparisons correction with the Tukey method]. (**D** and **E**) E-P loop (D) and mRNA expression (E) of *Cry1* at ZT22 (black), ZT10 (white), and ZT10 Rev-erbα knockout (KO) (red), represented as mean ± SEM (n = 4, one-way ANOVA followed by Dunnett multiple-comparisons test). (**F** and **G**) E-P loop (F) and mRNA expression (G) of *Cry1* at ZT22 with control green fluorescent protein (GFP) expression (black) versus Rev-erbα overexpression (blue) expressed as mean ± SEM (n = 5, two-tailed Student t test). (**H**) Same analysis as in Fig. 1B, but comparing ZT10 Rev-erbα KO (αKO) to wild-type ZT10 at ZT22 sub-TADs (red, higher interaction ratio at ZT10 αKO; blue, higher interaction ratio at wild-type ZT10). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

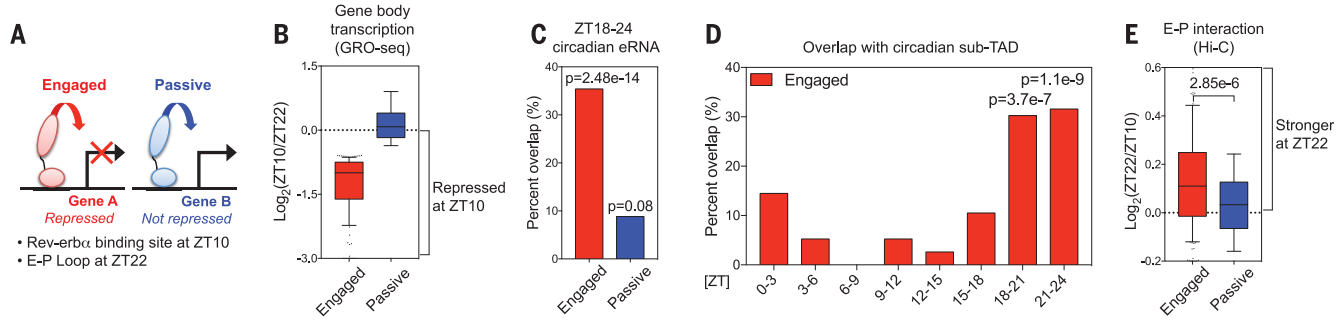


Fig. 3. Rev-erbα attenuates enhancer-promoter looping at functional binding sites. (**A**) Rev-erbα sites at E-P loops were classified as engaged when looped to genes whose transcription was repressed at ZT10 relative to ZT22, and otherwise classified as passive. (**B**) Gene body transcription change between ZT22 and ZT10 at engaged versus passive sites. Engaged target genes were defined as those displaying gene body transcription relative change of ≥1.5 between ZT22 and ZT10. (**C**) Engaged Rev-erbα sites were highly correlated with ZT18-24 circadian eRNAs (*6*) (within ±2 kbp, one-tailed hypergeometric tests). (**D**) Engaged Rev-erbα sites were confined within sub-TADs that contain circadian genes peaking at ZT18-24 (one-tailed hypergeometric tests). (**E**) E-P loops between engaged Rev-erbα binding sites and target gene promoters were stronger at ZT22 than at ZT10 (Mann-Whitney test). For box plots in (B) and (E), whiskers denote the 10th and 90th percentiles.

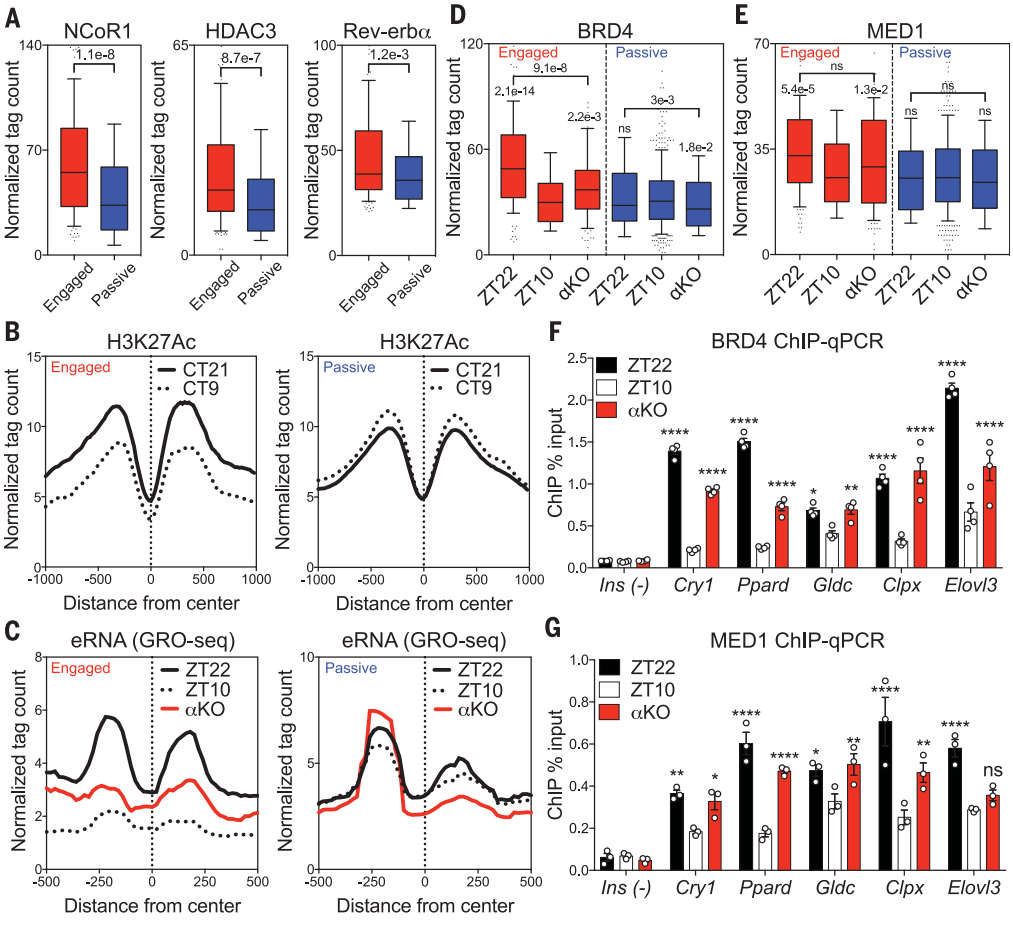


Fig. 4. Functional Rev-erbα binding evicts BRD4 and MED1 from sites of looping. (**A**) Greater recruitment of NCoR1 and HDAC3 at engaged Rev-erbα sites associated with a slight average increase in Rev-erbα binding (Mann-Whitney tests). (**B**) Circadian deacetylation of H3K27Ac at circadian time 21 (CT21) and CT9 at engaged versus passive sites [data sets from (*5*)]. (**C**) Circadian eRNA transcription between ZT22 and ZT10, with increased transcription at ZT10 in αKO at engaged sites [data sets from (*6*)]. (**D** and **E**) Circadian eviction of BRD4 (D) and MED1 (E) between ZT22 and ZT10, with enhanced binding at ZT10 in αKO at engaged sites (Dunn multiple-comparisons tests after one-way ANOVA/Friedman test). (**F** and **G**) ChIP-qPCR validation of BRD4 (F) and MED1 (G) eviction at ZT10 and enhanced binding at ZT10 in αKO at engaged sites (*Ins* as a negative control, $n = 3$ or 4, two-way ANOVA followed by Dunnett multiple-comparisons test). For box plots in (A), (D), and (E), whiskers denote the 10th and 90th percentiles. *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$; ns, not significant.

with circadian eRNAs whose activity peaked around ZT18 to ZT24 (ZT18-24) antiphase to Rev-erbα binding (Fig. 3C), demonstrating that engaged sites are direct links between circadian eRNAs and genes repressed by Rev-erbα. De novo motif analysis at engaged sites revealed enrichment of DNA motifs known to be bound by Rev-erbα directly (RORE and RevDR2) (*22*) or indirectly (HNF6) (*21*) (fig. S5A), whereas these motifs were not enriched at passive sites (fig. S5B). Although the number of sub-TADs containing circadian genes from each phase was similar (fig. S5C), only engaged Rev-erbα sites were enriched in ZT18-24 circadian sub-TADs (Fig. 3D and fig. S5D). Most important, E-P interactions at engaged Rev-erbα binding sites were strengthened at ZT22 relative to ZT10 (Fig. 3E). Together, these findings show that the ability of Rev-erbα to oppose E-P loops within sub-TADs is a likely determinant of active repression by Rev-erbα.

We next addressed the mechanism by which Rev-erbα binding controls circadian E-P interactions. The repressive action of Rev-erbα is often mediated by recruitment of a co-repressor complex containing NCoR (nuclear receptor co-repressor) and HDAC3 (histone deacetylase 3) (*23*), leading to circadian histone deacetylation associated with repressed enhancers (*3*). Indeed, recruitment of NCoR and HDAC3 was greater at engaged Rev-erbα binding sites, with a modest average increase in Rev-erbα binding relative to passive sites (Fig. 4A). Consistent with this finding, the previously demonstrated circadian acetylation of histone H3 Lys^{27} (H3K27Ac) in mouse liver (*5*) was exaggerated at engaged Rev-erbα binding sites but

was nearly absent at passive sites (Fig. 4B). eRNA transcription also demonstrated an enhanced circadian rhythm at engaged sites, which was abrogated in livers lacking Rev-erbα (Fig. 4C); this finding shows that Rev-erbα was required for the epigenomic rhythms that occurred selectively at engaged sites.

The transcriptional regulator BRD4 acts as a reader of acetylated histone (*24*–*26*) and forms a functional transcriptional complex with the well-established looping factor MED1 (*27*–*30*). In agreement with the changes in H3K27Ac at engaged sites, BRD4 binding was greater at ZT22 than at ZT10, and this difference was attenuated in the genetic absence of Rev-erbα (Fig. 4D). Similarly, MED1 was also evicted at engaged sites but not in livers lacking Rev-erbα (Fig. 4E). The opposition of Rev-erbα to the circadian binding of BRD4 and MED1 was confirmed by ChIP–quantitative polymerase chain reaction (qPCR) in both the genetic absence (Fig. 4, F and G) and ectopic expression of Rev-erbα (fig. S6, A to C). Focusing on Rev-erbα binding sites where BRD4 is evicted at ZT10 (fig. S6D) revealed the concurrent eviction of MED1 (fig. S6E) and also independently predicted functional sites with circadian eRNA transcription (fig. S6F). However, the binding of CTCF and RAD21 was low and not circadian at Rev-erbα binding sites (fig. S6, G to J).

Our findings demonstrate genome-wide organizational plasticity at the level of sub-TADs that occurs in a circadian manner as a component of normal mammalian physiology. The mechanisms by which Rev-erbα functionally opposes E-P loop formation, leading to circadian repression of gene transcription within sub-TADs, are likely applicable to other transcriptional repressors whose function in controlling chromatin architecture is currently not as well defined as for transcriptional activators.

REFERENCES AND NOTES

1. J. Bass, M. A. Lazar, *Science* **354**, 994–999 (2016).
2. J. S. Takahashi, *Nat. Rev. Genet.* **18**, 164–179 (2017).
3. D. Feng *et al.*, *Science* **331**, 1315–1319 (2011).
4. N. Koike *et al.*, *Science* **338**, 349–354 (2012).
5. C. Vollmers *et al.*, *Cell Metab.* **16**, 833–845 (2012).
6. B. Fang *et al.*, *Cell* **159**, 1140–1152 (2014).
7. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
8. E. P. Nora *et al.*, *Nature* **485**, 381–385 (2012).
9. T. Sexton *et al.*, *Cell* **148**, 458–472 (2012).
10. H. Zhao *et al.*, *Mol. Cell* **59**, 984–997 (2015).
11. L. Aguilar-Arnal *et al.*, *Nat. Struct. Mol. Biol.* **20**, 1206–1213 (2013).
12. Y. Xu *et al.*, *PLOS Genet.* **12**, e1005992 (2016).
13. E. Lieberman-Aiden *et al.*, *Science* **326**, 289–293 (2009).
14. S. S. P. Rao *et al.*, *Cell* **159**, 1665–1680 (2014).
15. J. E. Phillips-Cremins *et al.*, *Cell* **153**, 1281–1295 (2013).
16. J. M. Dowen *et al.*, *Cell* **159**, 374–387 (2014).
17. R. Siersbæk *et al.*, *Mol. Cell* **66**, 420–435.e5 (2017).
18. S. M. Tan-Wong *et al.*, *Science* **338**, 671–675 (2012).
19. A. R. Grosso, S. F. de Almeida, J. Braga, M. Carmo-Fonseca, *Genome Res.* **22**, 1447–1456 (2012).
20. N. Preitner *et al.*, *Cell* **110**, 251–260 (2002).
21. Y. Zhang *et al.*, *Science* **348**, 1488–1492 (2015).
22. H. P. Harding, M. A. Lazar, *Mol. Cell. Biol.* **15**, 4791–4802 (1995).
23. L. Yin, M. A. Lazar, *Mol. Endocrinol.* **19**, 1452–1459 (2005).
24. A. Dey, F. Chitsaz, A. Abbasi, T. Misteli, K. Ozato, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8758–8763 (2003).
25. J. S. Roe, F. Mercan, K. Rivera, D. J. Pappin, C. R. Vakoc, *Mol. Cell* **58**, 1028–1039 (2015).
26. M. K. Jang *et al.*, *Mol. Cell* **19**, 523–534 (2005).
27. Y. W. Jiang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8538–8543 (1998).
28. J. Lovén *et al.*, *Cell* **153**, 320–334 (2013).
29. S. Y. Wu, C. M. Chiang, *J. Biol. Chem.* **282**, 13141–13145 (2007).
30. A. S. Bhagwat *et al.*, *Cell Rep.* **15**, 519–530 (2016).

ACKNOWLEDGMENTS

We thank M. Adlanmerini, P. Dierickx, B. Fang, D. Guan, R. Papazyan, R. Soccio, Y. Aubert, and the Lazar laboratory for helpful discussions; J. Phillips-Cremins (University of Pennsylvania) and P. Huang and C. Edwards (Children's Hospital of Philadelphia) for technical advice; the Functional Genomics Core of the Penn Diabetes Research Center and the Penn Epigenetics Institute for next-generation sequencing; and the Viral Vector Core of the Penn Diabetes Research Center for recombinant virus preparation. ChIP-seq and Hi-C data have been deposited in the Gene Expression Omnibus (GSE104129). Supported by the JPB Foundation (M.A.L.) and by NIH grants R01 DK45586 and P30 DK19525 (M.A.L.), R01 DK106027 (K.-J.W.), and T32 GM007170, T32 GM008216, and F30 DK112507 (Y.H.K.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1274/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S5
References (*31*–*35*)

18 August 2017; accepted 17 January 2018
Published online 8 February 2018
10.1126/science.aao6891

AGING

Lysosome activation clears aggregates and enhances quiescent neural stem cell activation during aging

Dena S. Leeman,[1,2] Katja Hebestreit,[1]* Tyson Ruetz,[1] Ashley E. Webb,[1]† Andrew McKay,[1,3] Elizabeth A. Pollina,[1,2]‡ Ben W. Dulken,[1,4] Xiaoai Zhao,[1] Robin W. Yeo,[1] Theodore T. Ho,[5] Salah Mahmoudi,[1] Keerthana Devarajan,[1] Emmanuelle Passegué,[5]§ Thomas A. Rando,[6,7] Judith Frydman,[1,8] Anne Brunet[1,2,7]||

In the adult brain, the neural stem cell (NSC) pool comprises quiescent and activated populations with distinct roles. Transcriptomic analysis revealed that quiescent and activated NSCs exhibited differences in their protein homeostasis network. Whereas activated NSCs had active proteasomes, quiescent NSCs contained large lysosomes. Quiescent NSCs from young mice accumulated protein aggregates, and many of these aggregates were stored in large lysosomes. Perturbation of lysosomal activity in quiescent NSCs affected protein-aggregate accumulation and the ability of quiescent NSCs to activate. During aging, quiescent NSCs displayed defects in their lysosomes, increased accumulation of protein aggregates, and reduced ability to activate. Enhancement of the lysosome pathway in old quiescent NSCs cleared protein aggregates and ameliorated the ability of quiescent NSCs to activate, allowing them to regain a more youthful state.

Preservation of a pristine proteome is critical for maintaining cell function over long periods of time, and decline in proteome health is associated with aging and neurodegenerative disease (*1*, *2*). Three primary mechanisms are used to maintain protein homeostasis (proteostasis): molecular chaperones, the proteasome proteolytic system, and the lysosome-autophagy proteolytic system (*2*, *3*). Proteostasis is particularly important for stem cell proliferation and differentiation (*4*). However, somatic stem cell pools contain not only proliferating (activated) stem cells but also quiescent stem cells. How these different types of adult stem cells maintain protein homeostasis and deal with protein aggregation is largely unknown.

Protein aggregates are particularly relevant in the brain, where their accumulation is associated with neurodegenerative disease (*1*). The adult brain contains pools of regenerative neural stem cells (NSCs) that can generate new neurons (neurogenesis) (*5*, *6*). In one of the neurogenic niches—the subventricular zone (SVZ)—a subset of quiescent astrocytes are the source of quiescent NSCs (qNSCs), which in turn give rise to activated NSCs (aNSCs) and neural progenitor cells (NPCs) (*5*, *7*). These NPCs generate new neurons, oligodendrocytes, and astrocytes, contributing to the maintenance of olfactory function, as well as repair after brain injury (*5*, *8*). The ability of qNSCs to activate declines with age (*9*–*11*). Consistently, aging is associated with a decline in neurogenesis and a reduction in olfactory discrimination and memory (*6*, *9*). Understanding the age-dependent changes in proteostasis in the NSC pool could help identify ways to enhance the function of old NSCs and maintain brain health.

To gain an unbiased overview of the proteostasis network in different cell types in a regenerative brain region during aging, we used transcriptomic profiling. We aged cohorts of transgenic mice expressing green fluorescent protein (GFP) under the control of the human glial fibrillary acidic protein (*GFAP*) promoter, which drives expression in astrocytes and NSCs (*7*). Using fluorescence-activated cell sorting (FACS) with a combination of GFP and cell surface markers, we simultaneously purified five cell types from the SVZ of paired young (3- to 4-month-old) and old (19- to 22-month-old) mice: endothelial cells (which provide key environmental signals to NSCs), SVZ astrocytes (hereafter, astrocytes), qNSCs, aNSCs, and NPCs (Fig. 1A and fig. S1A) (*7*). Each population expressed well-established markers (fig. S1B) and showed expected cell-cycle characteristics (fig. S1C). We used RNA-sequencing (RNA-seq) to globally profile the transcriptomes of four biological replicates per cell type and per age (tables S1 and S2 and fig. S2, A to C). The transcriptomes of endothelial cells, quiescent cells (astrocytes and qNSCs), and activated cells (aNSCs and NPCs) could be distinguished by principal component analysis (PCA) (Fig. 1B) and hierarchical clustering (fig. S2D). By contrast, young and old samples from the same cell type were not easily separated when all cell types were considered together (Fig. 1B and fig. S2D), though age-dependent differences were revealed when cell types were analyzed individually (see below, Fig. 4 and fig. S8).

Analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that are up-regulated or down-regulated between qNSCs and aNSCs showed differences in several homeostatic and metabolic pathways (e.g., lipid metabolism, DNA repair, cell cycle) (*7*, *12*–*15*) as well as strong differences in the protein homeostasis network (Fig. 1C; figs. S3, A to D, and S4A; and tables S3 to S6). aNSCs exhibited increased expression of proteasome-associated genes (Fig. 1D, fig. S4B, and tables S5 and S6) (*7*, *15*). By contrast, qNSCs had increased expression of lysosome-associated genes (Fig. 1D, fig. S4B, and tables S5 and S6). A motif for the transcription factor EB (TFEB), a master regulator of lysosomes and autophagy (*16*, *17*), was enriched in lysosomal genes of qNSCs (fig. S4C). More generally, TFEB target genes were more highly expressed in qNSCs than in aNSCs (fig. S4C) (*17*), and TFEB was among the top five most significant upstream regulators of qNSC gene expression (fig. S4D). qNSCs and aNSCs also expressed different types of genes that encode protein chaperones (Fig. 1D, fig. S4E, and table S6): endoplasmic reticulum (ER) unfolded-protein response genes were more highly expressed in qNSCs, whereas the TCP-1 ring complex (TRiC) and prefoldin complex genes were more highly expressed in aNSCs (fig. S4E). Similar differences in branches of the proteostasis network were observed between the transcriptomes of quiescent and activated stem cells of other tissues (muscle and hematopoietic) (fig. S4F). Thus, quiescent and activated stem cells mobilize different branches of the protein quality-control network.

We experimentally assessed the status of the lysosomes and proteasomes in qNSCs and aNSCs. Staining for the lysosomal membrane protein LAMP-1 revealed that qNSCs freshly isolated from the adult brain had more LAMP-1 staining than did their activated counterparts (Fig. 1E). This was confirmed by staining for LAMP-1 in whole mounts of the SVZ (fig. S5A). Staining for the lysosomal membrane protein LAMP-2 in a primary culture system for NSC quiescence and activation (*18*) showed that qNSCs not only had more lysosomes but also larger lysosomes than aNSCs (Fig. 1F and fig. S5B). The large qNSC lysosomes were labeled with LysoTracker, a fluorescent probe specific to acidic organelles (Fig. 1G), suggesting that these lysosomes had the acidity required for proteolytic activity. Lysosomes can fuse with autophagosomes, thereby

[1]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [2]Cancer Biology Program, Stanford University, Stanford, CA 94305, USA. [3]Biology Graduate Program, Stanford University, Stanford, CA 94305, USA. [4]Stanford Medical Scientist Training Program, Stanford University, Stanford, CA 94305, USA. [5]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Department of Medicine, Division of Hematology and Oncology, University of California, San Francisco, San Francisco, CA 94143, USA. [6]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA. [7]Glenn Center for the Biology of Aging at Stanford University, Stanford, CA 94305, USA. [8]Department of Biology, Stanford University, Stanford, CA 94305, USA.
*Present address: Verge Genomics, San Francisco, CA 94103, USA. †Present address: Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI 02903, USA. ‡Present address: Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA. §Present address: Department of Genetics and Development, Columbia University, New York, NY 10032, USA.
||Corresponding author. Email: anne.brunet@stanford.edu

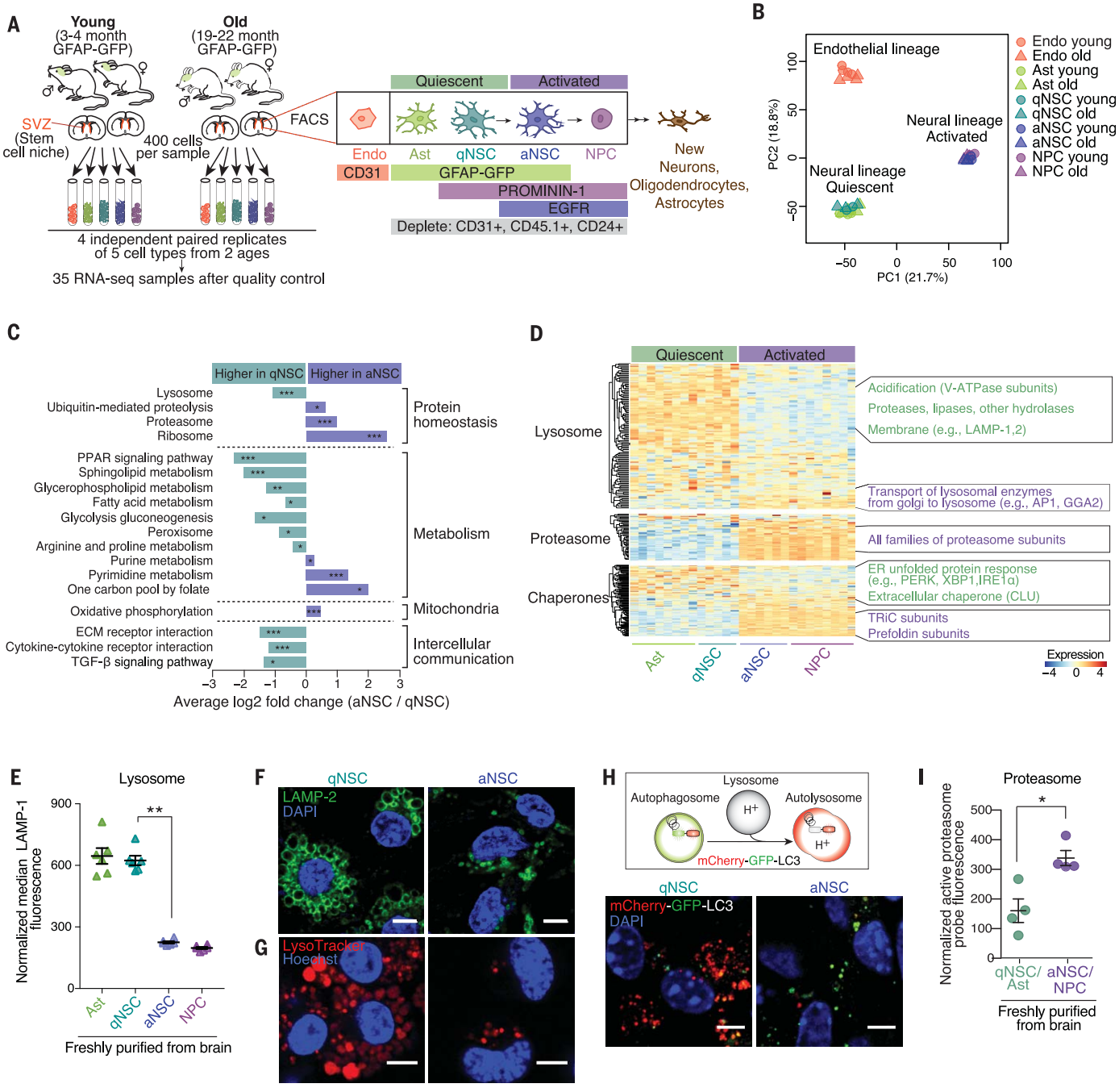


Fig. 1. qNSCs exhibit a protein quality-control signature distinct from their activated progeny, with enlarged lysosomes. (**A**) Experimental design. Endo, endothelial cell; Ast, astrocyte; EGFR, epidermal growth factor receptor. (**B**) PCA of five cell types and two ages, performed on variance stabilizing transformation (VST)–normalized read counts of all detected genes. PC, principal component. (**C**) Selected KEGG pathways significantly upregulated in qNSCs or aNSCs. *P* values determined by bootstrap sampling, with Benjamini-Hochberg correction. *False discovery rate (FDR)–adjusted $P \leq 0.05$; **FDR-adjusted $P \leq 0.001$; ***FDR-adjusted $P \leq 0.0001$. See also fig. S3 and table S5. (**D**) Heatmap showing expression levels of all genes in pathways from the three main branches of proteostasis: proteasome, lysosome, and chaperones. VST-normalized read counts, scaled row-wise. V-ATPase, vacuolar H^+–adenosine triphosphatase. See also table S6. (**E**) Median fluorescence of LAMP-1 normalized to cell size, as measured by flow cytometry. Mean ± SEM of values from six different mice (3 to 4 months old) analyzed on a single day. Each point represents FACS-sorted cells from a single mouse. *P* value determined by one-sided Wilcoxon rank sum test. **$P \leq 0.01$. (**F** and **G**) Representative immunofluorescence images of primary cultures of qNSCs and aNSCs from 3-month-old mice stained with (F) LAMP-2 (green, lysosomal membranes) and 4′,6-diamidino-2-phenylindole (DAPI) (blue, nuclei) or (G) LysoTracker (red, acidic organelles) and Hoechst (blue, nuclei). Scale bars, 5 μm. See also figs. S5B and S9B. (**H**) Top: In cells expressing mCherry-GFP-LC3, autophagosomes display both GFP and mCherry fluorescence (yellow-green), whereas autolysosomes display only mCherry fluorescence (red) because GFP is denatured by the acidity of the lysosome. Bottom: Representative images of GFP and mCherry fluorescence in primary cultures of qNSCs and aNSCs from 3-month-old mice. Scale bars, 5 μm. See also fig. S5C. (**I**) Fluorescence level of Me4BodipyFL-Ahx3Leu3VS, a probe that covalently binds active proteasomes, in populations directly isolated from the brain. Mean ± SEM of values from four different mice (3 to 4 months old) sorted and analyzed on two different days. Each point represents cells from a single mouse. *P* value determined by one-sided Wilcoxon rank sum test. *$P \leq 0.05$.

forming autolysosomes that, in turn, degrade autophagosome contents (*3*). Using a system in which LC3 (a marker of autophagosomes) is fused to both GFP and mCherry (Fig. 1H) (*19*), we found that qNSCs contained numerous large lysosomes that had fused with autophagosomes to form autolysosomes but had not yet degraded their contents (Fig. 1H and fig. S5C). Consistently, qNSCs freshly isolated from the SVZ showed less accumulation of LC3 than did aNSCs in response to inhibitors of lysosomal acidification (bafilomycin A and chloroquine) (fig. S5D), indicative of slower degradation of autophagosomes by qNSC lysosomes. Thus, although the lysosomes of qNSCs are acidified, they degrade their contents relatively slowly at steady state.

We then tested whether proteasome activity was different between qNSCs and aNSCs. As predicted by our RNA-seq data, qNSCs directly isolated from the brain had reduced staining for catalytically active proteasomes (Fig. 1I) (*20*) and conjugated ubiquitin (fig. S5E) compared to that of aNSCs. Furthermore, lysates from primary cultured qNSCs had less proteasome activity than did lysates of aNSCs (fig. S5F). Thus, qNSCs and aNSCs use different branches of the proteostasis network, with qNSCs exhibiting large lysosomes and aNSCs displaying active proteasomes.

The differences between protein quality-control pathways in qNSCs and aNSCs raise the question of each cell type's ability to regulate proteome health (fig. S6A). We stained young adult qNSCs and aNSCs freshly purified from the brain and in primary culture with Proteostat dye, which becomes highly fluorescent upon binding to the amyloid-type β-sheet tertiary structure of protein aggregates (*21*). qNSCs, even from young adults, exhibited increased Proteostat staining compared to that of aNSCs (Fig. 2, A and B). The accumulation of protein aggregates in young qNSCs was surprising, given that cellular quiescence is accompanied by reduced protein synthesis (*14*, *15*), and indeed, freshly isolated qNSCs had less protein synthesis than did aNSCs (fig. S6B). A large proportion of protein aggregates in qNSCs were contained within lysosomes (Fig. 2A), consistent with the observation that qNSC lysosomes degrade the contents of autophagosomes more slowly.

Biochemical fractionation of lysates from qNSCs and aNSCs to enrich for protein aggregates confirmed that qNSCs contained significantly more insoluble protein aggregates than did aNSCs (Fig. 2C and fig. S6, C to E). The presence of proteins at a variety of molecular weights in the qNSC insoluble fraction suggests that qNSC aggregates contain a wide array of proteins (Fig. 2C). Additionally, accumulation of qNSC insoluble proteins at the top of the SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gel is consistent with the presence of highly ordered amyloid-like aggregates (*22*) (Fig. 2C and fig. S6E). Thus, qNSCs contain more protein aggregates than their activated counterparts, and this aggregation appears to affect many proteins. The accumulation of protein aggregates in healthy young adult qNSCs raises the possibility that these aggregates play a physiological role.

The primary function of qNSCs is to generate aNSCs in response to growth factor signals. We tested whether manipulations that regulate lysosomal activity could affect protein aggregates in qNSCs and, in turn, influence activation of qNSCs in response to activation signals [epidermal growth factor (EGF) and basic fibroblast growth factor (bFGF)]. Treatment with bafilomycin A, which blocks lysosomal acidification, led to increased accumulation of protein aggregates (Fig. 3, A and B, and fig. S7A) and to decreased activation in response to growth factors (Fig. 3C). To induce lysosomal degradation, we deprived cells of nutrients by incubating them in Hank's balanced salt solution (HBSS) (*23*). Nutrient deprivation of qNSCs led to a significant decrease

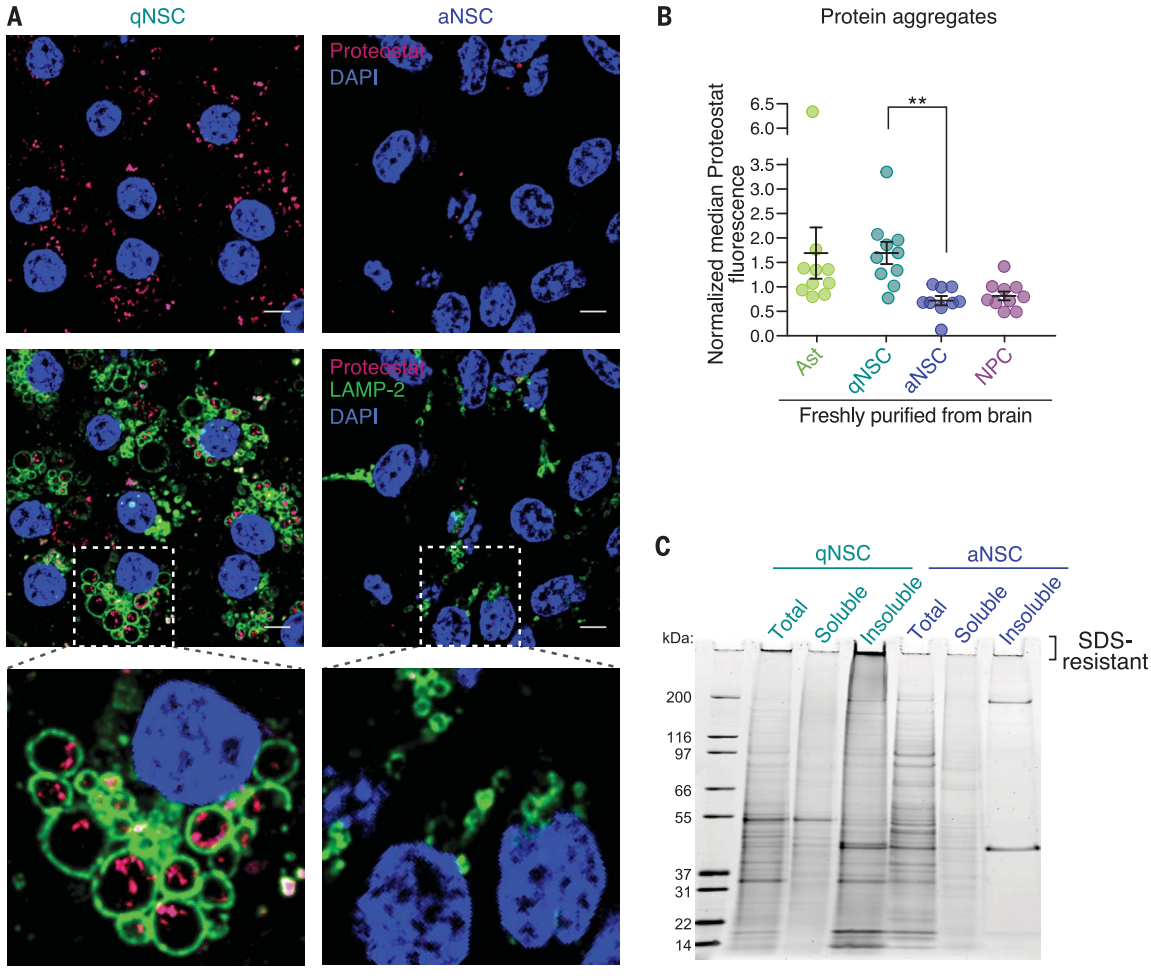


Fig. 2. qNSCs have insoluble protein aggregates localized in large lysosomes. (**A**) Representative immunofluorescence images of qNSC and aNSC primary cultures from 3-month-old mice stained with Proteostat (magenta, protein aggregates), LAMP-2 (green, lysosomal membranes), and DAPI (blue, nuclei). Scale bars, 5 μm. (**B**) Median fluorescence of Proteostat normalized to cell size and to the median value on the day of measurement (to combine independent experiments). Mean ± SEM of values from 10 different mice (3 to 4 months old) analyzed by flow cytometry on four different days. Each point represents FACS-sorted cells from a single mouse. *P* value determined by two-sided Wilcoxon rank sum test. $^{**}P \leq 0.01$. (**C**) Representative SDS-PAGE gel stained with SYPRO Ruby for detection of protein in the total, soluble, and insoluble fractions of qNSCs and aNSCs. See also fig. S6, C to E.

in lysosomal staining (fig. S7B) and in the amount of protein aggregates (Fig. 3, A and D, and fig. S7, C and D). A brief pulse of nutrient deprivation in qNSCs enhanced their subsequent activation by growth factors, and this was abolished by bafilomycin A (Fig. 3E). Enhanced qNSC activation did not result from direct effects of nutrient deprivation on proliferation: Nutrient deprivation did not affect proliferation of qNSCs in the absence of growth factors, and indeed, it decreased proliferation of aNSCs (fig. S7E). To activate lysosome function more selectively, we used expression of TFEB—a master regulator of lysosome biogenesis and autophagy (*16*, *17*). Brief expression of an activated form of TFEB in qNSCs (*24*) (fig. S7F) led to decreased abundance of protein aggregates in qNSCs (Fig. 3F) and to enhanced qNSC activation in response to growth factors (Fig. 3G). Thus, activating the lysosome-autophagy pathway in qNSCs cleared protein aggregates and improved qNSC activation in response to growth factors.

The collapse of protein homeostasis is a key feature of aging in differentiated cells (*1*). To understand how aging affects protein homeostasis in stem cells, we analyzed our RNA-seq data sets for changes in global transcriptional profiles as well as specific transcriptional signatures in qNSCs and aNSCs freshly purified from old (19- to 22-month-old) versus young (3- to 4-month-old) mice (Fig. 1A). Quiescent cell types (qNSCs and astrocytes) displayed more and larger-magnitude transcriptional changes with age than did their activated counterparts (aNSCs and NPCs) (Fig. 4A, fig. S8A, and tables S7 and S8). This was not caused by technical variability between biological replicates (fig. S8B) and was also observed with a different transcript quantification and analysis method (fig. S8, C and D). Consistently, the transcriptome of young and old qNSCs and astrocytes could be distinguished by PCA, whereas that of young and old aNSCs and NPCs could not (fig. S8, E and F). The age-associated differences in qNSCs could reflect changes that arise from a heterogeneous population of qNSCs (*5*, *11*, *15*). The expression of genes associated with lysosome function, but not proteasome function, was significantly changed with age in qNSCs but not in aNSCs (fig. S8G). Thus, qNSCs undergo more transcriptional changes with age than do aNSCs and NPCs, and the expression of lysosome genes changes with age.

We experimentally examined the status of lysosomes and proteasomes in old qNSCs and aNSCs. Old qNSCs in their in vivo niches or freshly purified from the brain exhibited less lysosomal staining (LAMP-1) than young qNSCs (Fig. 4, B and C), and this decrease was significantly greater than that observed with age in aNSCs (Fig. 4C). This reduction in lysosomal staining was also observed in primary qNSC cultures from old mice (fig. S9, A and B). The age-dependent decrease in lysosome levels could result in fewer lysosomes available to fuse with autophagosomes (and/or endosomes). Indeed, using old GFP-LC3 reporter mice, we identified a distinct

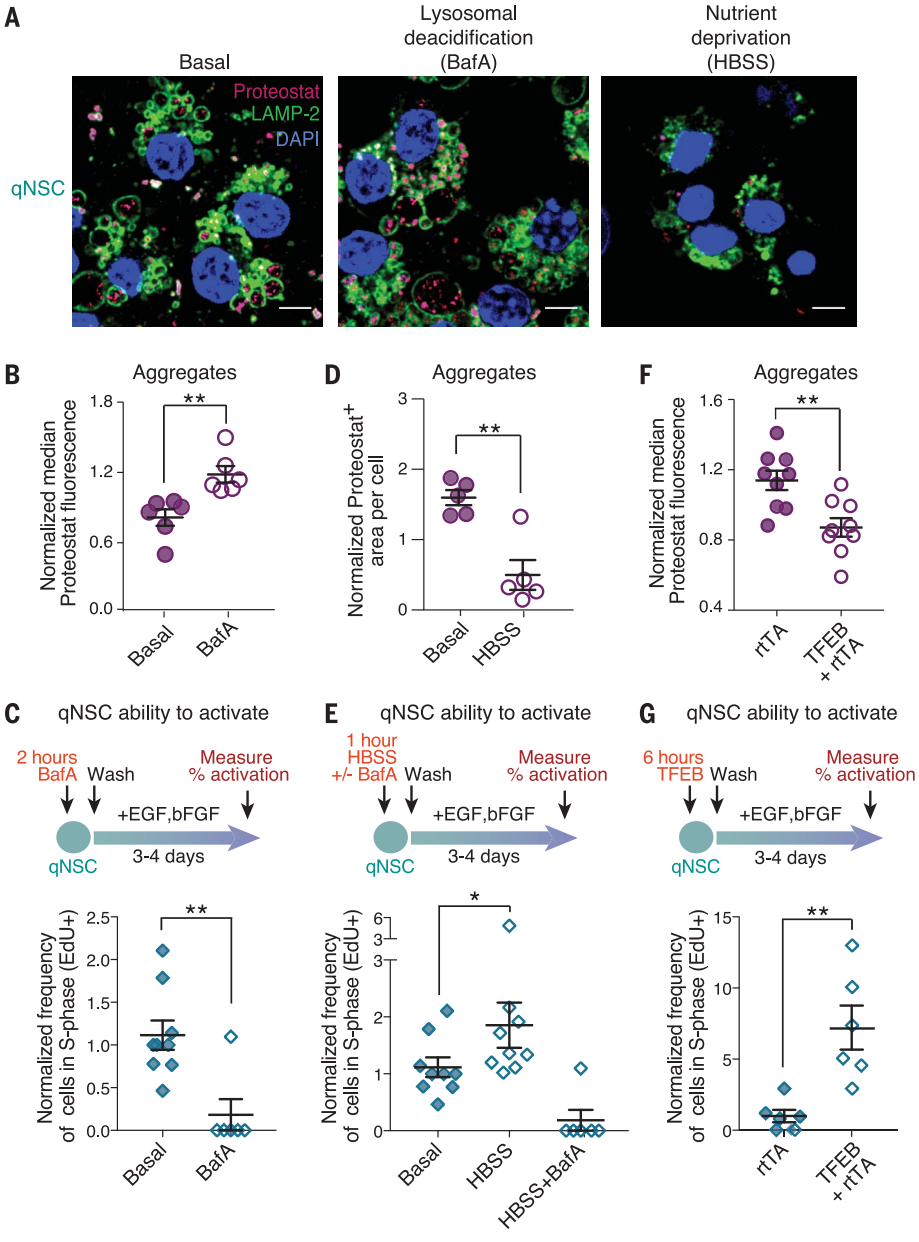


Fig. 3. Modulation of lysosomal activity in qNSCs affects protein aggregates and the transition from quiescence to activation. (**A**) Representative immunofluorescence images of primary cultures of qNSCs from 3-month-old mice after 18 hours of treatment with 50 nM bafilomycin A (BafA, a lysosomal V-ATPase inhibitor) or 3 hours of nutrient deprivation in HBSS. Cells are stained with Proteostat (magenta, protein aggregates), LAMP-2 (green, lysosomal membranes), and DAPI (blue, nuclei). Scale bars, 5 μm. Basal, quiescence media. (**B**) Median Proteostat fluorescence normalized to the median value per replicate to combine samples prepared at different times. Mean ± SEM of values from six biological replicates analyzed by flow cytometry on two different days. Each point represents an independent primary culture derived from two 3- to 4-month-old mice. (**C**) Percentage of cells that incorporated 5-ethynyl-2′-deoxyuridine (EdU) (cells in S phase of the cell cycle) during a 3-hour pulse, as measured by intracellular flow cytometry. Mean ± SEM of values from nine different biological replicates analyzed on four different days. Each point represents one independent primary culture derived from two 3- to 4-month-old mice. (**D**) Differences in Proteostat^{+} area quantified from images. Mean ± SEM of values from five independent biological replicates from two independent experiments. (**E**) Percentage of cells that incorporated EdU measured as in (C). Mean ± SEM of values from nine different biological replicates analyzed on four different days. (**F**) Proteostat fluorescence measured as in (B). Mean ± SEM of values from nine biological replicates analyzed on three different days. rtTA, reverse tetracycline-controlled transactivator. (**G**) Percentage of cells that incorporated EdU measured as in (C). Mean ± SEM of values from six different biological replicates analyzed on two different days. For all panels, *P* values determined by one-sided Wilcoxon rank sum test. *$P \leq 0.05$; **$P \leq 0.01$.

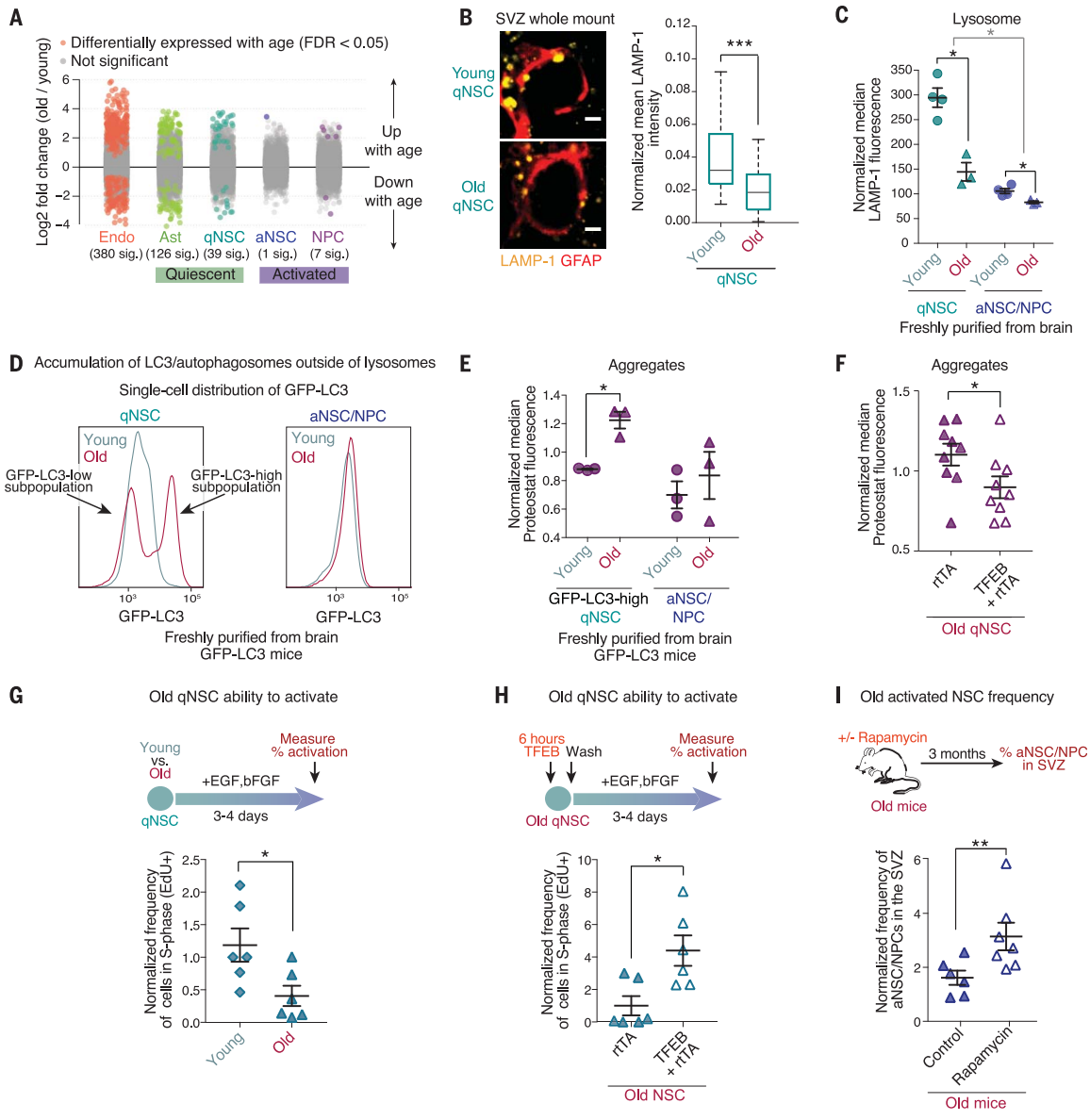


Fig. 4. Old qNSCs show lysosome defects, increased protein aggregates, and decreased activation, but these can be counteracted by lysosomal activation. (**A**) Log2 fold changes between young (3- to 4-month-old) and old (19- to 22-month-old) samples of each cell type for all detected genes using DESeq2 (*36*). Each point represents one gene. Genes in color are significantly ("sig.") up- or down-regulated with age with FDR = 0.05. Genes in gray are not significantly up- or down-regulated with age. (**B**) SVZ whole-mount staining of young (3-month-old) and old (21-month-old) qNSCs. Left: LAMP-1 (lysosomes, yellow) staining in qNSCs identified by the combination of a KI67-negative cell soma and a $GFAP^+$ (red) projection through ependymal cell pinwheels at the ventricular surface (KI67, proliferation marker). Scale bar, 2 μm. Right: Quantification of LAMP-1 fluorescence intensity in qNSCs normalized for cell size from three young and three old mice (n = 65 qNSCs). (**C**) Median fluorescence of LAMP-1 normalized for cell size. Mean ± SEM of values from three to four young and old mice, analyzed by flow cytometry on a single day. Each point represents FACS-sorted cells from a single mouse. (**D**) Representative histograms of the GFP-LC3 fluorescence per cell in qNSC and aNSC populations from a 5-month-old versus 25-month-old transgenic GFP-LC3 mouse, as measured by flow cytometry. See also fig. S9, C and D. (**E**) Median fluorescence of Proteostat in GFP-LC3-high qNSCs and aNSCs and NPCs from young (2- to 5-month-old) and old (25-month-old) GFP-LC3 mice normalized to the median value per replicate to combine samples prepared at different times. Mean ± SEM of values from three to four different mice analyzed by flow cytometry on a single day. Each point represents cells FACS-sorted from a single mouse. (**F**) Median fluorescence of Proteostat normalized to the median value per replicate. Mean ± SEM of values from nine biological replicates analyzed on three different days. Each point represents one independent primary culture derived from a 17- to 22-month-old mouse. (**G**) Percentage of cells in an individual culture that incorporated EdU (cells in S phase) during a 3-hour pulse, as measured by intracellular flow cytometry. Mean ± SEM of values from six different biological replicates analyzed on two different days. Each point represents one independent primary culture derived from two 3-month-old or 20-month-old mice. (**H**) Percentage of cells that incorporated EdU measured as in (G). Mean ± SEM of values from six different biological replicates analyzed on two different days. Each point represents one independent primary culture derived from a 17- to 22-month-old mouse. (**I**) Old mice (22 months old) were treated for 3 months with rapamycin in mouse chow. SVZ composition was then analyzed by FACS. Each point represents the abundance of aNSCs and aNPCs ($EGFR^+CD24^-CD31^-CD45^-$ cells) in a single mouse, expressed as the percentage of total live cells in the SVZ. Mean ± SEM of values from six to seven different mice. P values [except panel (B)] determined by one-sided Wilcoxon rank sum test; Panel (B) P values determined by two-sided Wilcoxon rank sum test. *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.01$.

subpopulation of old qNSCs that had increased amounts of GFP-LC3 fluorescence (Fig. 4D and fig. S9, C and D), indicative of reduced autophagic flux into acidified lysosomes (where GFP is normally denatured) (*25*). There was heterogeneity within the qNSC population with age (Fig. 4D and fig. S9, C and D), indicating that a subset of old qNSCs have defects in their lysosomes. By contrast, proteasome activity was not significantly altered with age in qNSCs or aNSCs (fig. S9E).

We tested how protein aggregates are affected by aging in qNSCs. The subpopulation of old qNSCs with high amounts of GFP-LC3 exhibited an increased amount of protein aggregates in vivo (Fig. 4E). Consistently, primary cultures of qNSCs from old mice also exhibited an increase in abundance of protein aggregates (fig. S9F). The age-dependent accumulation of GFP-LC3 and protein aggregates in vivo was reduced by subjecting mice to fasting for 48 hours (fig. S9G), indicating that these aggregates can be modulated by nutrient deprivation. The increased amount of protein aggregates in old qNSCs could also be reduced by expression of constitutively active TFEB (Fig. 4F). Thus, old qNSCs exhibit increased protein aggregation, which may result from age-dependent reduction in lysosome number or function, and this can be counteracted by TFEB expression.

The ability of qNSCs to activate declines with age (*9*–*11*). Consistently, old qNSCs in primary culture activated less efficiently than young qNSCs in response to growth factors (Fig. 4G). Interestingly, a pulse of expression of the active form of TFEB in old qNSCs restored activation in response to growth factors (Fig. 4H), suggesting that increased lysosomal function helps promote activation of old NSCs. Furthermore, systemic treatment of old mice with rapamycin, a mechanistic target of rapamycin (mTOR) inhibitor that enhances the lysosome-autophagy pathway (*26*, *27*), resulted in increased abundance of aNSCs in the SVZ (Fig. 4I). Although rapamycin could also exert its effects indirectly by acting in cells other than NSCs and/or by regulating other pathways (*26*), these data are consistent with the possibility that increasing lysosomal function can clear protein aggregates and counteract the decline in activation of old qNSCs (fig. S9H).

Our study identifies a previously uncharacterized organelle state in qNSCs: an enlarged lysosome that contains a large amount of insoluble protein aggregates. The clearance of qNSC aggregates by lysosomal activation may provide a burst of energy for activation or prevent the propagation of potentially toxic aggregates to NSC progeny. Although changes in autophagy during aging have been observed in hematopoietic and muscle stem cells (*27*–*29*), these studies did not examine the lysosomes and their protein aggregate content in these stem cells during aging. As lysosomes are central hubs for key cellular trafficking, signaling, and metabolic pathways, targeting of lysosome biogenesis and/or activation to clear aggregates in qNSCs may provide broad strategies to allow activation of old mammalian cells in vivo. This possibility is intriguing in light of the benefits of dietary restriction or starvation followed by refeeding to regenerative stem cell pools (*30*, *31*) and the effects of the lysosome-autophagy pathway on protein aggregates (*32*) and life span (*33*) in *Caenorhabditis elegans*. Identifying the conserved mechanisms by which protein aggregates are regulated in aging stem cells should help identify specific ways to restore function to old stem cells and tissues.

REFERENCES AND NOTES

1. W. E. Balch, R. I. Morimoto, A. Dillin, J. W. Kelly, *Science* **319**, 916–919 (2008).
2. M. S. Hipp, S. H. Park, F. U. Hartl, *Trends Cell Biol.* **24**, 506–514 (2014).
3. E. Wong, A. M. Cuervo, *Cold Spring Harb. Perspect. Biol.* **2**, a006734 (2010).
4. D. Vilchez, M. S. Simic, A. Dillin, *Trends Cell Biol.* **24**, 161–170 (2014).
5. D. A. Lim, A. Alvarez-Buylla, *Cold Spring Harb. Perspect. Biol.* **8**, a018820 (2016).
6. J. T. Gonçalves, S. T. Schafer, F. H. Gage, *Cell* **167**, 897–914 (2016).
7. P. Codega *et al.*, *Neuron* **82**, 545–559 (2014).
8. S. G. Kernie, J. M. Parent, *Neurobiol. Dis.* **37**, 267–274 (2010).
9. E. Enwere *et al.*, *J. Neurosci.* **24**, 8354–8365 (2004).
10. V. Capilla-Gonzalez, A. Cebrian-Silla, H. Guerrero-Cazares, J. M. Garcia-Verdugo, A. Quiñones-Hinojosa, *Glia* **62**, 790–803 (2014).
11. C. Giachino *et al.*, *Stem Cells* **32**, 70–84 (2014).
12. M. Knobloch, S. Jessberger, *Curr. Opin. Neurobiol.* **42**, 45–52 (2017).
13. F. J. Obermair *et al.*, *Stem Cell Res.* **5**, 131–143 (2010).
14. J. R. Valcourt *et al.*, *Cell Cycle* **11**, 1680–1696 (2012).
15. E. Llorens-Bobadilla *et al.*, *Cell Stem Cell* **17**, 329–340 (2015).
16. M. Sardiello *et al.*, *Science* **325**, 473–477 (2009).
17. M. Palmieri *et al.*, *Hum. Mol. Genet.* **20**, 3852–3866 (2011).
18. B. Martynoga *et al.*, *Genes Dev.* **27**, 1769–1786 (2013).
19. S. Pankiv *et al.*, *J. Biol. Chem.* **282**, 24131–24145 (2007).
20. C. R. Berkers *et al.*, *Mol. Pharm.* **4**, 739–748 (2007).
21. D. Shen *et al.*, *Cell Biochem. Biophys.* **60**, 173–185 (2011).
22. S. Alberti, R. Halfmann, O. King, A. Kapila, S. Lindquist, *Cell* **137**, 146–158 (2009).
23. J. Zhou *et al.*, *Cell Res.* **23**, 508–523 (2013).
24. C. Di Malta *et al.*, *Science* **356**, 1188–1192 (2017).
25. N. Mizushima, A. Yamamoto, M. Matsui, T. Yoshimori, Y. Ohsumi, *Mol. Biol. Cell* **15**, 1101–1111 (2004).
26. D. W. Lamming, L. Ye, D. M. Sabatini, J. A. Baur, *J. Clin. Invest.* **123**, 980–989 (2013).
27. L. García-Prat *et al.*, *Nature* **529**, 37–42 (2016).
28. T. T. Ho *et al.*, *Nature* **543**, 205–210 (2017).
29. A. H. Tang, T. A. Rando, *EMBO J.* **33**, 2782–2797 (2014).
30. S. Brandhorst *et al.*, *Cell Metab.* **22**, 86–99 (2015).
31. G. Valdez *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14863–14868 (2010).
32. K. A. Bohnert, C. Kenyon, *Nature* **551**, 629–633 (2017).
33. L. R. Lapierre *et al.*, *Nat. Commun.* **4**, 2267 (2013).

ACKNOWLEDGMENTS

We thank P. Codega, V. Silva-Vargas, and F. Doetsch for help with the NSC FACS protocol; C. Di Malta and A. Ballabio for the constitutively active TFEB vector; M. Hansen for the mCherry-GFP-LC3 vector and for helpful discussions; A. Roux and C. Kenyon for help with the protein aggregate protocol; R. Zoncu for suggestions regarding the lysosome; T. Palmer, S. Artandi, M. Winslow, C. Jan, A. Freund, C. Sidrauski, K. Rainbolt, and M. Pech for helpful discussions and input on the manuscript; D. Gaviño and J. Brett for sharing rapamycin-treated mice; the Stanford Shared FACS Facility and, in particular, C. Carswell-Crumpton for technical support; B. Benayoun, L. Booth, I. Harel, and other members of the Brunet lab for helpful discussions and feedback on the manuscript; C.-K. Hu and E. Noblin for advice on confocal microscopy and image analysis; and J. Butterfield for help with mouse husbandry and genotyping. This work is supported by NIH grant P01 AG036695 (A.B. and T.A.R.), Stanford Bio-X seed grant (A.B. and J.F.), the NSF (D.S.L. and E.A.P.), a National Defense Science and Engineering Graduate Fellowship (D.S.L.), NIH grant F31 AG043232 (E.A.P.), the Stanford Cancer Biology training grant T32 CA09302 (D.S.L. and E.A.P.), NIH training grant T32 GM7365 (B.W.D.), and the Glenn/American Federation for Aging Research (AFAR) Postdoctoral Fellowship Program for Translational Research on Aging (X.Z.). **Author contributions:** D.S.L. and A.B. planned the study. D.S.L. performed the experiments, analyzed the data, and wrote the manuscript with help from A.B. K.H. analyzed all RNA-seq data and performed all statistical analyses. T.R. designed, performed, and analyzed the TFEB experiments. A.E.W. and E.A.P. helped optimize RNA-seq, and A.E.W. generated the RNA-seq libraries. A.E.W. provided input into the analysis of proteostasis differences and subcloned mCherry-GFP-LC3 in a lentiviral vector. A.M. optimized and performed SVZ whole-mount staining and image analysis. E.A.P. helped with FACS protocol development and the generation of RNA-seq data. E.A.P. and B.W.D. helped with quantitative PCR validation. E.A.P. and B.W.D. optimized the culture system for NSC quiescence and activation. X.Z. and R.W.Y. helped optimize FACS. X.Z. helped with aggregate gels. R.W.Y. performed TFEB motif analysis. T.T.H. and E.P. provided young and old transgenic GFP-LC3 mice and input on GFP-LC3 sort. S.M. performed the upstream regulator analysis on RNA-seq data. K.D. helped with proteasome activity assay, LC3 antibody, and staining for FACS. T.A.R. provided intellectual contribution on stem cells and provided rapamycin-treated mice. J.F. provided intellectual contribution on proteostasis. All authors discussed the results and commented on the manuscript. **Competing interests:** None declared. **Data and materials availability:** Sequencing data have been deposited to ArrayExpress under accession number E-MTAB-5172 and to the Sequence Read Archive (SRA) under accession number SRP075993. The normalized RNA-seq read counts are also available in table S2. All code used for the RNA-seq analysis is available on Github (https://github.com/brunetlab/Leeman_et_al_2017.git).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1277/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S8
References (*34*–*47*)

7 June 2016; resubmitted 31 August 2017
Accepted 21 January 2018
10.1126/science.aag3048

SINGLE-CELL ANALYSIS

Mutation dynamics and fitness effects followed in single cells

Lydia Robert,[1,2]* Jean Ollion,[1] Jerome Robert,[1] Xiaohu Song,[3] Ivan Matic,[3] Marina Elez[1,4]*

Mutations have been investigated for more than a century but remain difficult to observe directly in single cells, which limits the characterization of their dynamics and fitness effects. By combining microfluidics, time-lapse imaging, and a fluorescent tag of the mismatch repair system in *Escherichia coli*, we visualized the emergence of mutations in single cells, revealing Poissonian dynamics. Concomitantly, we tracked the growth and life span of single cells, accumulating ~20,000 mutations genome-wide over hundreds of generations. This analysis revealed that 1% of mutations were lethal; nonlethal mutations displayed a heavy-tailed distribution of fitness effects and were dominated by quasi-neutral mutations with an average cost of 0.3%. Our approach has enabled the investigation of single-cell individuality in mutation rate, mutation fitness costs, and mutation interactions.

The pace of evolution and possible trajectories depend on the dynamics of mutation incidence and the effects of mutations on fitness. Mutation dynamics has never been analyzed directly. It is assumed to be Poissonian, but mutagenesis characterized by bursts, with transient mutation rate increases, has been demonstrated in several organisms (*1*–*3*). Mutation bursts may facilitate the acquisition of combinations of mutations, which is critical to the evolution of complex traits (*4*). Effects of mutations are known to range from beneficial to deleterious or lethal, but estimating the entire distribution of fitness effects (DFE), central to evolutionary modeling, has proven challenging (*5*). The DFE for spontaneous mutations has been estimated with the mutation accumulation (MA) approach (*6*–*8*), whereby independent lines of individuals are propagated in conditions that minimize the effects of natural selection and the evolution of their fitness is tracked. The precision of such DFE estimations is limited because of the small number of lines that can be monitored, typically fewer than 100. In addition, only an upper boundary for the average mutational effect is obtained, or strong assumptions are made regarding the DFE shape (*9*, *10*). Finally, in MA studies of microorganisms, the mutation sample is biased, as natural selection purges lethal and strongly deleterious mutations.

We quantitatively characterized the dynamics of spontaneous point mutations arising from replication errors, a major source of mutations (*11*, *12*), and their DFE in the bacterium *Escherichia coli*. For real-time detection of mutations in single living cells, we developed a mutation visualization (MV) experiment. We also designed a microfluidic MA (μMA) experiment, which allowed us to track fitness in the complete absence of natural selection in a high-throughput fashion during the accumulation of mutations.

In μMA experiments, we grew cells in a "mother machine" microfluidic chip (*13*) (Fig. 1A) and imaged them by phase-contrast microscopy to monitor their growth and survival at the single-cell level (*14*) (Fig. 1B and movie S1). The mother machine contained a series of separate microchannels where cells grew in a single row (Fig. 1A). These microchannels were closed on one side and retained the "mother" cell abutting the dead end through consecutive divisions. Therefore, at each generation, one individual was kept independently of its fitness (Fig. 1C), thus eliminating natural selection. Malthusian fitness, the exponential growth rate at the population level, can also be defined at the single-cell level as the cell exponential elongation rate. We imaged more than 1000 microchannels in parallel, in controlled and constant conditions, every 4 min

[1]Laboratoire Jean Perrin, UMR 8237 Sorbonne Universités, UPMC Université Paris 06, Paris, France. [2]Micalis Institute, Institut National de la Recherche Agronomique, AgroParisTech, Université Paris Saclay, Jouy-en-Josas, France. [3]INSERM U1001, Université Paris-Descartes, Sorbonne Paris Cité, Faculté de Médecine Paris Descartes, Paris, France. [4]Institute of Systems and Synthetic Biology, UMR 8030, CNRS, Commissariat à l'Energie Atomique et aux Energies Alternatives, Genopole, Université d'Évry Val-d'Essonne, Université Paris Saclay, Evry, France.
*Corresponding author. Email: lydia.robert@polytechnique.org (L.R.); marina.elez3@gmail.com (M.E.)

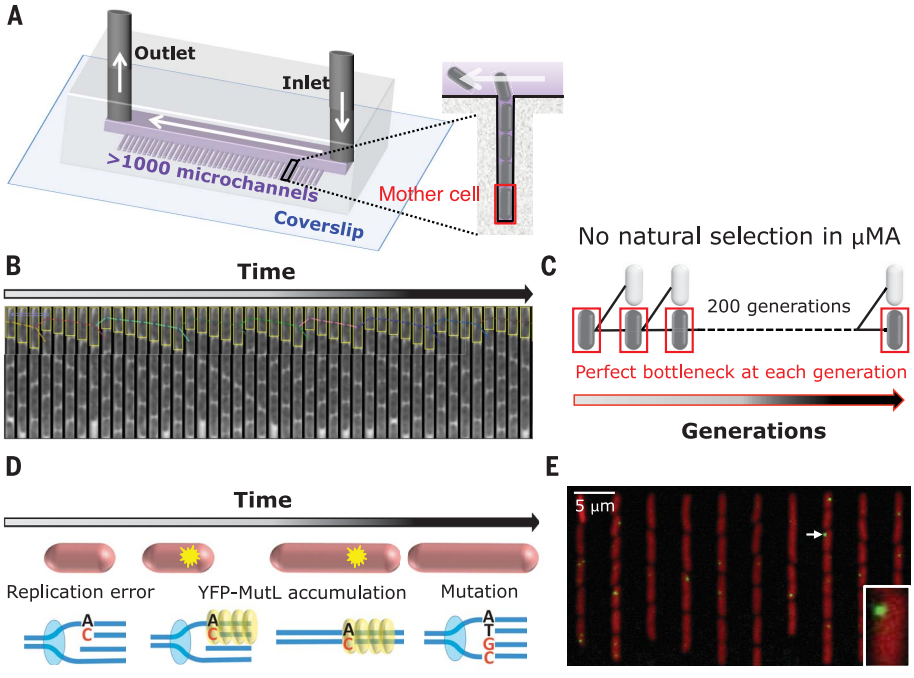


Fig. 1. Experimental setup. (**A**) The mother machine microfluidic chip, used for growing *E. coli* cells during μMA and MV experiments. μMA and MV experiments were performed independently, using phase-contrast microscopy and epifluorescence microcopy, respectively. (**B**) Phase-contrast images of one microchannel over time in a μMA experiment. We developed software that measures the mother cell length (represented by yellow boxes) at each frame and computes its fitness at each generation (*14*). (**C**) The mother machine allowed us to keep the mother cell at each division, regardless of its fitness, thus blocking natural selection. (**D**) Tracking point mutation emergence in single cells with YFP-MutL in the *mutH* strain, where all replication errors are converted into mutations. (**E**) Overlay of red (from the constitutive expression of tdCherry for automated cell segmentation) and yellow (YFP-MutL) fluorescence images from an MV experiment. The inset shows a magnified image of a cell with a YFP-MutL focus (arrow).

over the course of 3 days, which corresponded to ~200 generations. We obtained ~10^5 fitness measurements per experiment.

In MV experiments, cells grown in the mother machine were imaged by fluorescence microscopy (*14*). We visualized mutations using a fusion of yellow fluorescent protein (YFP) with MutL mismatch repair protein, which forms fluorescent foci at replication errors (*15*) (Fig. 1, D and E, and movie S2). To detect and track MutL foci, we developed an automated image analysis procedure (*14*). The dynamics of mutations originating from replication errors depends on both error production and repair. We examined the dynamics of error production by eliminating error repair through inactivation of the mismatch repair endonuclease MutH (*14*). Full inactivation was verified by control experiments (*14*). The YFP-MutL protein forms transient foci (*15*) (movie S2), and the average lifetime of the foci in the *mutH* strain corresponds to the time between the passage of two successive replication forks (*14*) (fig. S1). This pattern shows that in *mutH* cells, all MutL foci are converted into mutations by the next round of DNA replication.

Replication errors might occur at a constant rate with Poissonian dynamics, or alternatively, bursts of errors might be produced from fluctuations in intracellular composition, such as transient nucleotide pool imbalances (*1*, *3*). To investigate these possibilities, we performed MV experiments using *mutH* cells, tracking more than 3000 MutL foci in mother cells and deriving their interarrival times (the times between two successive focus occurrences) (Fig. 2A). These interarrival times were exponentially distributed and uncorrelated, which is characteristic of a Poissonian process (*14*) (Fig. 2, B and C, and fig. S2). In agreement with this finding, the number of MutL foci appearing between two frames followed a Poissonian distribution (fig. S3). As our method is limited by optical resolution, we cannot detect bursts of replication errors if they occur at a single replication fork on a short genomic region (typically less than 80 kilobases). In contrast, we can detect transient hypermutator states if they affect several replication forks, as would be expected for metabolic fluctuations, because different forks occupy different positions in the cell.

Our data indicate Poissonian dynamics rather than dynamics characterized by bursts, but they do not exclude some fluctuations in the replication error rate. In particular, we expect the error rate to depend on the number of replication forks, which changes during the cell cycle (*14*). We found a global proportionality between cell size, a proxy for the stage in the cell cycle, and the replication error rate across a range of cell sizes (fig. S4). Further analysis of the relationship between cell size and error rate revealed steplike variations reminiscent of variations in the number of replication forks (Fig. 2, D and E), with error rates varying by a factor of ~3 during a single cell cycle.

When exposed to certain external stresses, bacteria trigger specific responses that transiently increase the mutation rate (*16*–*18*). In our MV experiments, a small fraction of cells experienced endogenous stress, as demonstrated by reduced growth rate and/or abnormally large cell size. We investigated whether such subpopulations exhibit higher rates of replication error. To take into account the different replication fork numbers for different subpopulations, we estimated the number of forks in single cells and the mean error rate per fork in each subpopulation (*14*). Abnormally large cells and slowly growing cells did not present elevated error rates per fork (Fig. 2F and fig. S5), suggesting that stress-induced mutagenesis (SIM) acts mainly at the level of error repair or that endogenous stress in a favorable environment causes limited SIM.

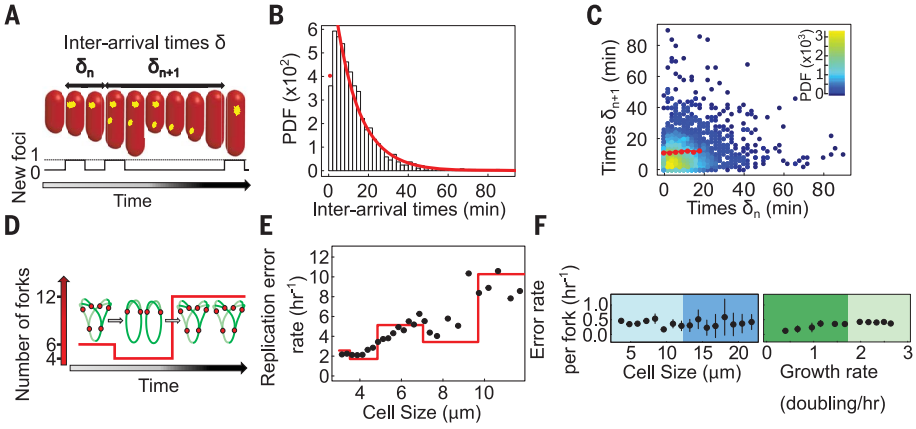


Fig. 2. Replication error dynamics. (**A**) The variables analyzed in MV experiments were the interarrival times for YFP-MutL foci (δ) and the number of new foci per mother cell. δ_n and δ_{n+1} indicate two successive interarrival times. (**B**) Histogram of focus interarrival times for a representative *mutH* experiment (n = 1873 foci). The red line and dot are predictions under the Poisson process hypothesis, which includes a correction accounting for discrete observation (*14*). PDF, probability density function. (**C**) Correlation between two successive focus interarrival times for the data in (B). Color indicates local density (inset). Red dots represent data binned according to δ_n. (**D**) Theoretical number of forks as a function of time for a doubling time of 26 min. Chromosomes are depicted in green, and the red dots denote replication forks. (**E**) Rate of replication error as a function of cell size in single cells. Black dots, experimental data; red line, prediction based on the number of forks (*14*). (**F**) Error rate per fork as a function of cell size and growth rate for single cells. Data were binned according to either cell size or growth rate. Results are means, and errors bars show ±2 SEM.

Visualization of mutations and fitness measurements could in principle be combined into a single time-lapse microscopy experiment. However, simply linking one mutation to one fitness effect is impossible for two reasons: First, there can be a substantial lag between the occurrence of a mutation and its effect on fitness. Second, a MutL focus indicates the insertion of an erroneous base in one of two newly replicated DNA strands. Consequently, the progeny from a cell with a MutL focus will contain a mixture of mutant and nonmutant cells (*14*) (fig. S6). Therefore, we performed μMA and MV experiments separately, which allowed us to increase the throughput of μMA experiments.

Using MV, we estimated mutation rates for *E. coli* wild-type (WT) (~0.0022 mutation/hour) and *mutH* mutant (~0.32 mutation/hour) strains (*14*), which agree with estimations from whole-genome sequencing of MA lines of WT and mismatch repair–deficient cells (*11*). We then performed a μMA experiment with WT cells, which accumulated an average of ~0.1 mutation per mother cell line over ~60 hours. As expected, the growth rate distribution for WT cells was stable over time (Fig. 3A, left). In contrast, for *mutH* cells, which accumulated ~20 mutations per mother cell line over ~60 hours, the mean growth rate slightly decreased (Fig. 3A, middle), and some cells underwent a steplike decrease in growth rate (Fig. 3B, blue trajectory) or a total cessation of growth, indicating cell death (fig. S7). Surprisingly, the average decrease in the growth rate of *mutH* cells during μMA (~0.1% per hour) was smaller than what would be expected from previous estimates of the average effect of mutations (*7*, *19*, *20*). We also performed μMA experiments using the strain MF1 (*14*), which has a mutation rate ~20 times that of the *mutH* strain (*14*) (table S1) because of impaired proofreading (*21*). As expected, the decrease in fitness in MF1 cells was ~20 times that in *mutH* cells (Fig. 3A, right, and table S2).

We then characterized the DFE by developing a nonparametric method, with no assumptions about DFE shape (*14*). Using the probabilistic framework of MA studies, we derived all the moments of the DFE (*14*). For nonlethal mutations, our estimate of mean fitness cost (Fig. 3C) is 0.31% ± 0.04%, which is lower than previous estimates (*7*, *19*, *20*). We found that the DFE for nonlethal mutations has a coefficient of variation (CV) of 9.5 ± 1.2, a skewness of 16.6 ± 0.7, and a kurtosis of 360 ± 90 (Fig. 3C). We also

obtained moments of higher order for the DFE (table S3). These results indicate a distribution that is wide, skewed, and heavy tailed, dominated by quasi-neutral mutations relative to infrequent, strongly deleterious mutations, such as the beta distribution in Fig. 3D.

The estimated DFE for MF1 was similar to the DFE for the *mutH* strain (table S4), although these strains have different mutation spectra (*11*, *21*). To further investigate the effect of the mutation spectrum on the DFE, we performed μMA experiments using a *mutT* mutant (fig. S8), which has a mutation rate similar to that of the *mutH* strain but a specific mutation spectrum of >99% AT-to-CG transversions (*11*, *12*, *14*) (table S1). Despite the differences in the way mutagenesis occurs, we found that the DFE for the *mutT* strain did not differ substantially from the DFE for the *mutH* or the MF1 strain (table S4).

Steplike growth rate decreases such as those shown in Fig. 3B (blue trajectory) occurred during μMA in all mutator strains and not in the WT strain. They might be provoked by a single deleterious mutation, independently of previously acquired mutations, or by the accumulation of interacting mutations. To investigate this issue, we studied the rate of occurrence of growth rate decreases in *mutH* and *mutT* cells. From the analysis of all single-cell trajectories (*14*) (fig. S9), we defined two categories of cells: slow-growing cells that underwent a >30% decrease in growth rate (Fig. 4A, top) and fast-growing cells corresponding to the rest of the population. The percentage of fast-growing cells decreased exponentially over time (*14*) (Fig. 4A, bottom), demonstrating that >30% growth rate decreases are caused by single deleterious mutations with >30% fitness cost. These mutations occur at a constant rate, independently of the accumulation of previous mutations, and represent 0.3% of all mutations. As expected from the quasi-optimal growth rate of our *mutH* strain, our analysis of single-cell trajectories detected no strongly beneficial mutations (>20% effect) during μMA (*14*).

Likewise, to assess whether lethal mutations occur at a constant rate, we studied survival during μMA (*14*). The WT strain, accumulating ~0.1 mutation over the course of μMA, showed age-related mortality (Fig. 4B, left), in agreement with previous observations (*13*). The high–mutation-rate *mutH*, *mutT*, and MF1 strains exhibited additional mutation-related mortality (Fig. 4B, left). During the first ~35 hours of μMA, when the mortality rate for the WT was low and approximately constant, the survival of the *mutH*, *mutT*, and MF1 mutator strains exponentially decreased (Fig. 4B, right, and fig. S10). To further disentangle age- and mutation-related mortality, we performed a μMA experiment using the strain pMQ, which has increased proofreading deficiency compared with MF1 (*21*). pMQ cells died rapidly, and their survival decreased exponentially (Fig. 4B, right, and fig. S11). These analyses demonstrate that lethal mutations occur at a constant rate in μMA, independently of previously acquired mutations, and that ~1% of mutations are lethal.

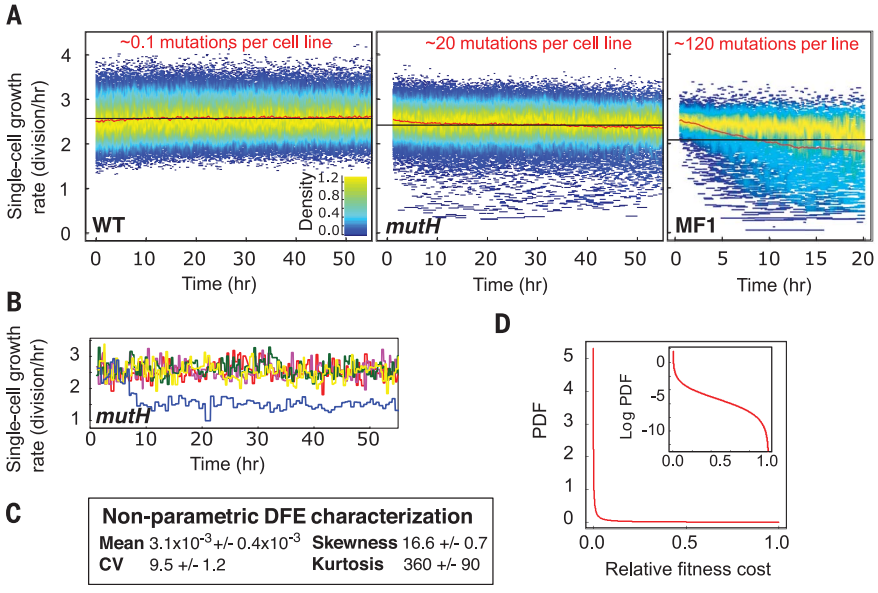


Fig. 3. μMA analysis allows nonparametric DFE characterization. (**A**) Representative μMA results for WT, *mutH*, and MF1 (*dnaQ926*) strains. For each time point, all single-cell growth rates for all mother cells are plotted (WT strain, $n = 1.8 \times 10^5$ cells; *mutH* strain, $n = 1.1 \times 10^5$ cells; MF1 strain, $n = 2.6 \times 10^4$ cells). Color depicts growth rate density. The red line represents the mean growth rate evolution. The black line indicates the mean of all growth rates. (**B**) Examples of single-cell growth rate trajectories for the *mutH* strain (one color per cell line). (**C**) Nonparametric estimation for the first four moments of the DFE allows calculation of the mean, the coefficient of variation (CV), skewness, and kurtosis (for each calculation, the mean ± 2 SEM for three independent *mutH* experiments is given). (**D**) Example of distribution with similar moments [beta distribution; parameters, $\alpha = 0.0074$ and $\beta = 2.4$, calculated to give the mean and CV as for the DFE in (C), giving skewness = 17 and kurtosis = 370]. The inset shows the same distribution on a log scale.

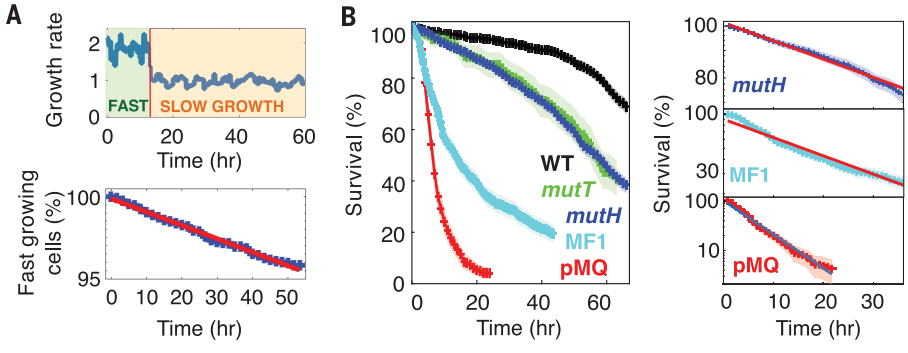


Fig. 4. Strongly deleterious (>30% cost) and lethal mutations occur at a constant rate during μMA. (**A**) The top panel shows an example of a *mutH* growth rate trajectory (blue line) corresponding to the occurrence of a strongly deleterious mutation in a mother cell. To analyze such events, we defined two categories of mother cells, slow-growing cells that underwent an abrupt, >30% growth rate decrease and fast-growing cells corresponding to the rest of the population. The bottom panel shows the temporal evolution of the percentage of fast-growing cells (represented by blue dots, on a log scale; we analyzed 5272 single-cell trajectories and found 156 strongly deleterious mutations). The 95% confidence interval is narrower than the line thickness. The red curve is an exponential fit (coefficient of determination $R^2 = 0.99$). (**B**) The left panel shows survival curves of the population for the WT ($n = 1283$ single-cell trajectories), *mutH* ($n = 1480$), *mutT* ($n = 216$), MF1 ($n = 838$), and pMQ ($n = 472$) strains. The 95% confidence intervals are represented by the shaded areas (they are narrower than the line thickness for $n > 400$). The right panel shows the same survival curves on a log scale for *mutH* (data in blue, exponential fit in red; $R^2 = 0.99$), MF1 (data in cyan, exponential fit in red; $R^2 = 0.99$), and pMQ (data in red, exponential fit in blue; $R^2 = 0.99$) strains. The plot is limited to ~35 hours, during which age-related mortality was low and approximately constant. The 95% confidence intervals are represented by the shaded areas.

Our results show that the accumulation of mutations with <30% fitness cost does not change the rate of lethal or strongly deleterious mutations (>30% cost), as would be expected from the saturation or induction of a buffering mechanism (*22*, *23*). We anticipate that examination of various suboptimal genotypes in μMA experiments would allow further investigation of mutation interactions and provide a complete picture of epistasis, central to the evolution of sexual reproduction and recombination (*24*). Our approach may also allow for investigation of how the DFE and mutation dynamics change in response to the environment, shedding light on the adaptation potential of populations.

REFERENCES AND NOTES

1. J. W. Drake, *Crit. Rev. Biochem. Mol. Biol.* **42**, 247–258 (2007).
2. S. Uphoff *et al.*, *Science* **351**, 1094–1097 (2016).
3. J. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8403–8408 (2007).
4. M. R. Stratton, P. J. Campbell, P. A. Futreal, *Nature* **458**, 719–724 (2009).
5. T. Bataillon, S. F. Bailey, *Ann. N.Y. Acad. Sci.* **1320**, 76–92 (2014).
6. T. Mukai, *Genetics* **50**, 1–19 (1964).
7. T. T. Kibota, M. Lynch, *Nature* **381**, 694–696 (1996).
8. P. D. Keightley, A. Caballero, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3823–3827 (1997).
9. A. Eyre-Walker, P. D. Keightley, *Nat. Rev. Genet.* **8**, 610–618 (2007).
10. D. L. Halligan, P. D. Keightley, *Annu. Rev. Ecol. Evol. Syst.* **40**, 151–172 (2009).
11. H. Lee, E. Popodi, H. Tang, P. L. Foster, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2774–E2783 (2012).
12. P. L. Foster, H. Lee, E. Popodi, J. P. Townes, H. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5990–E5999 (2015).
13. P. Wang *et al.*, *Curr. Biol.* **20**, 1099–1103 (2010).
14. Materials and methods and supplementary text are available as supplementary materials.
15. M. Elez *et al.*, *Curr. Biol.* **20**, 1432–1437 (2010).
16. R. S. Galhardo, P. J. Hastings, S. M. Rosenberg, *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
17. P. L. Foster, *Crit. Rev. Biochem. Mol. Biol.* **42**, 373–397 (2007).
18. R. Maharjan, T. Ferenci, *Mol. Biol. Evol.* **32**, 380–391 (2015).
19. S. Trindade, L. Perfeito, I. Gordo, *Philos. Trans. R. Soc. London Ser. B* **365**, 1177–1186 (2010).
20. L. Perfeito, A. Sousa, T. Bataillon, I. Gordo, *Evolution* **68**, 150–162 (2014).
21. K. M. Esvelt, J. C. Carlson, D. R. Liu, *Nature* **472**, 499–503 (2011).
22. S. Maisnier-Patin *et al.*, *Nat. Genet.* **37**, 1376–1379 (2005).
23. T. A. Sangster, S. Lindquist, C. Queitsch, *Bioessays* **26**, 348–362 (2004).
24. R. D. Kouyos, O. K. Silander, S. Bonhoeffer, *Trends Ecol. Evol.* **22**, 308–315 (2007).

ACKNOWLEDGMENTS

We thank M. F. Bredeche, S. Fleurier, L. Becerra, and M. Rosticher for technical assistance; O. Tenaillon, A. Amir, N. Henry, P. Thomen, G. Debregeas, A. F. Bitbol, V. Vargas, R. Robert, D. Chatenay, and M. De Paepe for discussions and comments on the manuscript; and N. Kuperwasser for manuscript editing. **Funding:** This study was funded by the Agence Nationale de Recherche (grant ANR-14-CE09-0015-01). **Author contributions:** M.E. and L.R. designed research and performed experiments. L.R., J.O., and M.E. analyzed the data. M.E. and L.R. wrote the manuscript with input from all authors. L.R. performed modeling. J.O. developed image analysis software. J.R., I.M., and X.S. contributed tools. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data described in the paper have been archived at Dryad (doi:10.5061/dryad.75625), and software codes and programs are available at Github (https://github.com/LabJeanPerrin/BOA).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6381/1283/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S12
References (*25*–*46*)
Movies S1 and S2

1 March 2017; resubmitted 19 October 2017
Accepted 30 January 2018
10.1126/science.aan0797

TECHNICAL RESPONSE

INORGANIC CHEMISTRY

Response to Comment on "Synthesis and characterization of the pentazolate anion *cyclo*-N_5^- in $(N_5)_6(H_3O)_3(NH_4)_4Cl$"

Chao Jiang,[1]* Lei Zhang,[2,3]* Chengguo Sun,[1,4] Chong Zhang,[1] Chen Yang,[1] Jun Chen,[2,3]† Bingcheng Hu[1]†

Huang and Xu argue that the *cyclo*-N_5^- ion in $(N_5)_6(H_3O)_3(NH_4)_4Cl$ we described in our report is theoretically unfavorable and is instead protonated. Their conclusion is invalid, as they use an improper method to assess the proton transfer in a solid crystal structure. We present an in-depth experimental and theoretical analysis of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ that supports the results in the original paper.

In their comment, Huang and Xu (*1*) bring up the point that the ellipsoid of the O_2 atom in the asymmetry unit plot of the pentazolate salt was much larger than those of the remaining atoms in $(N_5)_6(H_3O)_3(NH_4)_4Cl$, and conclude that there should actually be less electron density than assigned in the model. They claim the O_2 atom could also be a nitrogen atom. However, they ignore the anisotropic displacement parameters of the disordered atoms; thus, such a conclusion is meaningless. Huang and Xu also calculated separate ion pair $N_5^-\cdots H_3O^+$ energetics for the solid crystal structure and drew the conclusion that HN_5 should exist in place of N_5^-.

In our paper (*2*), the $(N_5)_6(H_3O)_3(NH_4)_4Cl$ crystal containing the stabilized pentazolate anion was prepared by the rupture of the C–N bond in 3,5-dimethyl-4-hydroxyphenylpentazole through treatment with *m*-chloroperbenzoic acid (*m*-CPBA) and ferrous bisglycinate [$Fe(Gly)_2$]. Although the high-quality x-ray crystallographic data provided the visualized proof of the structure of the pentazolate salt, confirmation of its structure and understanding of the cation-anion interactions have puzzled and intrigued researchers. Initial attempts via nuclear magnetic resonance (NMR), mass spectrometry (MS), and infrared (IR) spectroscopy to confirm the structure of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ were complicated. Fortunately, the expected single crystals were obtained by low-temperature evaporation of a solution in ethanol and ethyl acetate. At the very early stage, HN_5 was considered the most likely structural unit in pentazolate salt on the basis of NMR and MS spectra. However, the x-ray structure determination suffered from charge balance and the *cyclo*-N_5^- showed high symmetry with five equal N-N bond lengths. We concluded that the *cyclo*-N_5^- should exist in its anionic form rather than its acid form, HN_5. The non-hydrogen atoms were identified by their peak weights and sources from starting materials. Cl atoms gave the highest peak weights; N atoms gave the lowest. After all the non-hydrogen atoms were confirmed and refined, all the H atoms were smoothly located from Q maps from the high-quality crystal data. Each N atom was bonded to four H atoms, which gave an NH_4^+ cation. Each O atom was coordinated by six atoms. After searching similar ions and considering electric neutrality, the O atom was assigned as H_3O^+ with disordered H atoms with half occupancy. There were two oxygen atoms in the ellipsoid plot of $(N_5)_6(H_3O)_3(NH_4)_4Cl$, one a bit larger, the other cigar-shaped. The difference can be attributed to the disorder of the water molecules in the crystal structure, rather than to incorrect atom assignment (*3*).

Shown in Fig. 1, A and B, are two ^{15}N NMR spectra of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ in which the essential difference is the ^{15}N labeling in their starting material, 3,5-dimethyl-4-hydroxyphenylpentazole. The labeling effect with ^{15}N isotope can be discerned by analyzing the spectra (*4*). The unlabeled pentazolate salt shows a lone visible ^{15}N signal resonant at −356.18 ppm, whereas the labeled pentazolate salt has the same signal at

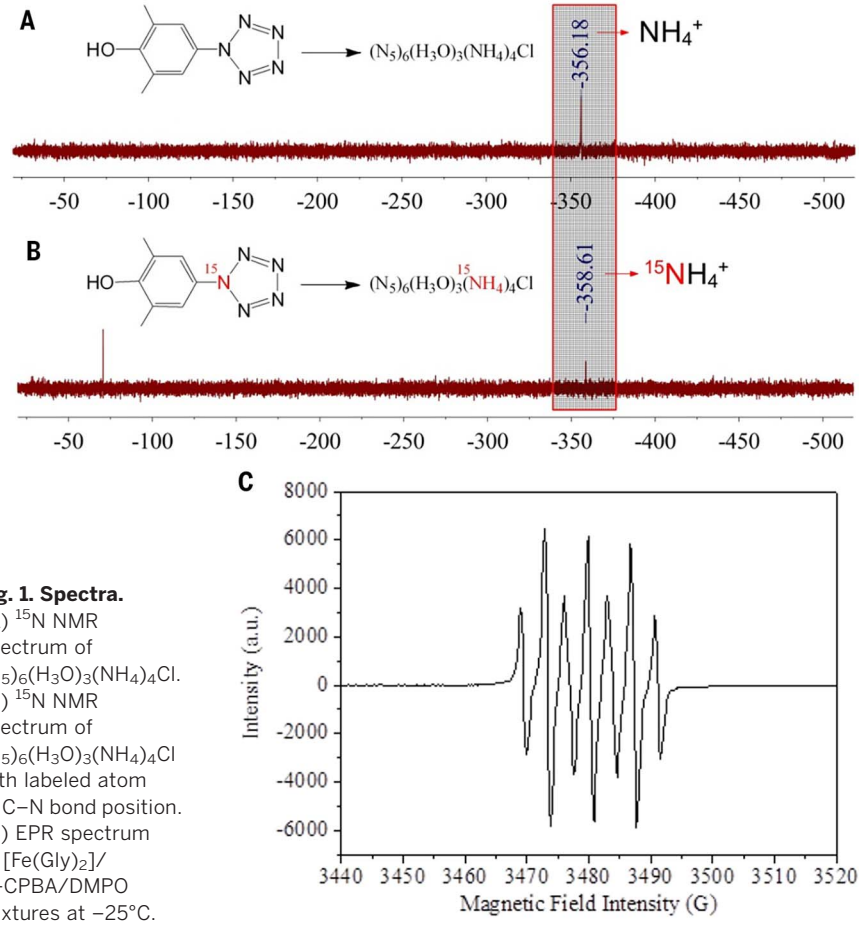


Fig. 1. Spectra. (**A**) ^{15}N NMR spectrum of $(N_5)_6(H_3O)_3(NH_4)_4Cl$. (**B**) ^{15}N NMR spectrum of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ with labeled atom at C–N bond position. (**C**) EPR spectrum of [$Fe(Gly)_2$]/*m*-CPBA/DMPO mixtures at −25°C.

[1]School of Chemical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China. [2]Software Center for High Performance Numerical Simulation, Institute of Applied Physics and Computational Mathematics, Beijing 100088, China. [3]Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Beijing 100088, China. [4]School of Chemical Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning, China.
*These authors contributed equally to this work.
†Corresponding author. Email: hubc@njust.edu.cn (B.H.); jun_chen@iapcm.ac.cn (J.C.)

–358.61 ppm. Both signals are reasonably assigned to NH_4^+ (*5*). The chemical shift difference of ~2.4 ppm comes from the ^{15}N labeling of the NH_4^+ cation, which indicates that the NH_4^+ cation is derived from the N atom bonded to the aryl ring in aryl pentazole through oxidative decomposition.

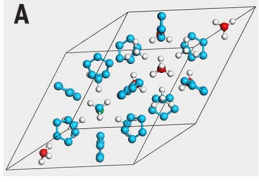
As far as the source of Cl^- is concerned, the only chlorine source is *m*-chloroperbenzoic acid. We concluded that the Cl^- ion must come from the dechlorination reaction of *m*-CPBA (*6*). Previous studies of iron-catalyzed oxidation reactions (*7*) indicate that the *m*-CPBA might react with ferrous glycinate and initiate a free radical process responsible for the cleavage of the C–N bond and the generation of Cl^- ion. The isolation of 2,6-dimethyl-1,4-benzoquinone from the reaction mixture supports the C–N bond cleavage pathway. In order to explore this free radical–based hypothesis, we recorded low-temperature electron paramagnetic resonance (EPR) spectra (Fig. 1C), which showed a strong EPR signal when mixing [$Fe(Gly)_2$], *m*-CPBA, and DMPO (dimethyl pyridine *N*-oxide). In summary, excess *m*-CPBA plays a dual role as the oxidant to initiate a free radical pathway combining with [$Fe(Gly)_2$] and the source of chloride.

To evaluate the thermodynamic stability of $(N_5)_6(H_3O)_3(NH_4)_4Cl$, we used density functional theory calculations to quantify the electrostatic interactions among the constituent ions in pentazolate salt (*8*). Taking the lattice parameters and atomic coordinates from single-crystal x-ray diffraction analysis as input, the geometry optimization was performed on the basis of the conjugate gradient method. One approximation for the hydronium ions is that the hydrogen atoms are fixed at three occupancy sites, whereas they practically occupy six resonance states at 50% probability level. The simulated structure was considered as finally optimized when the residual forces were less than 0.03 eV/Å and the stress components were less than 0.01 GPa. The simulated structure (Fig. 2A) shows satisfactory agreement with the characterized structure by x-ray diffraction, with the discrepancy of the volume at 0.4% and the discrepancy of lattice vectors at ~1%. At the equilibrium state of the crystal, each chloride ion is stabilized by four ammonium ions. In Fig. 2B, the four ammonium ions around the chloride ion form a tetrahedron, resulting in four N-H···Cl hydrogen bonds (2.44 Å and 8.02 kcal/mol). All the residual hydrogen atoms affiliated with ammonium ions directly contact the pentazolate anion, and the N···H-N hydrogen bond is 2.10 Å and 12.06 kcal/mol. All the hydrogen atoms affiliated with hydronium ions interact with their neighboring pentazolate anion, resulting in two types of N···H-O hydrogen bonds, the first being 2.17 Å and 13.26 kcal/mol and the second being 2.26 Å and 11.81 kcal/mol. With the aid of the electrostatic interactions among the ions, the entire pentazolate crystal is thermodynamically stable. Binding energy is defined as the difference between total energy of constituent ions in the free state and total energy of the crystal. The binding energy of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ crystal is calculated to be 1000 kcal/mol, much higher than the binding energy of NaCl (187.9 kcal/mol) (*9*).

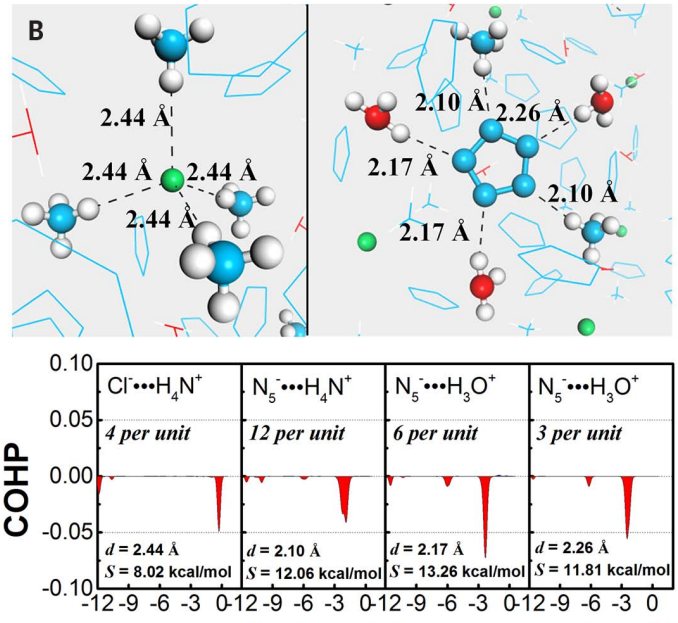
To better understand the basicity of *cyclo*-N_5^- ion, a known amount of pentazolate salt can be dissolved in deionized water at ambient temperature, enabling the pH of the solution to be accurately and easily determined by pH meter. When 50 mg of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ was dissolved in 5 ml of deionized water (1.71×10^{-2} mol/liter), the pH of the solution was measured to be ~5.55 at ambient temperature (22.3°C). In contrast, under the identical concentration of 1.71×10^{-2} mol/liter, the pH of aqueous ammonium chloride is 5.51 (*10*), hence the solution of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ is mildly acidic. The measured pH values for different concentrations of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ solutions are listed in Table 1. The results lead us

Table 1. The pH value of pentazolate salt solutions. Specific amounts of $(N_5)_6(H_3O)_3(NH_4)_4Cl$ (from 10 to 50 mg) were dissolved in 5 ml of deionized water at 22.3°C. The obtained solutions were analyzed using a pH meter with a pH probe.

Concentration (mol/liter)	pH
3.4×10^{-3}	5.64
6.8×10^{-3}	5.60
1.02×10^{-2}	5.57
1.36×10^{-2}	5.56
1.70×10^{-2}	5.55
2.04×10^{-2}	5.53



	a (Å)	*b* (Å)	*c* (Å)	*α* (°)	*β* (°)	*γ* (°)	Volume (Å^3)
Calc.	12.59	12.90	12.72	60.20	59.82	59.54	1456.21
Expt.	12.71	12.71	12.71	60.00	60.00	60.00	1450.26
Error (%)	-0.94	1.49	0.08	0.33	-0.30	-0.77	0.41



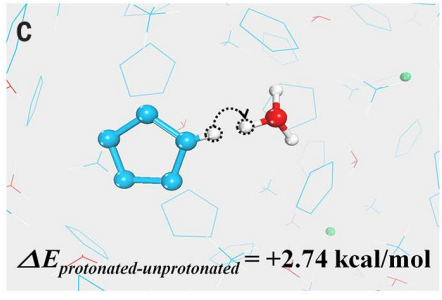


Fig. 2. Calculations. (**A**) Calculated lattice vector lengths (Å), angles (°), and volumes (Å^3) of the primitive cell of pentazolate salt, along with results from experiments. (**B**) Hydrogen bonds stabilizing $(N_5)_6(H_3O)_3(NH_4)_4Cl$ at equilibrium state. Crystal orbital Hamilton population (COHP), the result of multiplying the electron density of states by the overlap element of the Hamiltonian, shows the contribution of a specific contact to band energy. The integrated value of COHP at band energy directly quantifies the interactive strength (*S*, in kcal/mol). (**C**) Proton transfer between *cyclo*-N_5^- ion and hydronium ion/ammonium ion in crystal and in vacuum. $\Delta E_{protonated\text{-}unprotonated}$ is the total energy difference of the systems between the protonated state and the unprotonated state. The state with lower total energy is the thermodynamically more stable one. Color code: sapphire, N; red, O; green, Cl; white, H.

to conclude that the correctly calculated equilibrium constant of *cyclo*-N_5^- ion is $pK_b^{N_5^-} = 8.425 \pm 0.002$, indicating that it is more basic than acetate ion ($pK_b^{Ac^-} = 9.24$) (*10*).

In general, NH_3 has a stronger affinity for the proton than for H_2O (*11*). If the HN_5 instead of *cyclo*-N_5^- were present in the crystal of $(N_5)_6(H_3O)_3(NH_4)_4Cl$, the *cyclo*-$N_5^-$ ion should accept protons from H_3O^+ rather than NH_4^+. On the basis of this hypothesis, we constructed a crystal structure model (Fig. 2C) in accordance with $(N_5)_6(H_3O)_3(NH_4)_4Cl$ except that one proton had shifted from the hydronium ion to the *cyclo*-N_5^- ion (with the H-N bond length set at 0.95 Å). Subsequently, the geometry optimization of the created structure was performed on the basis of the conjugate gradient method. One unanticipated finding was that the proton on the *cyclo*-N_5^- ion quickly transferred to the nearest H_2O molecule to form H_3O^+. The total energy was lowered by 2.74 kcal/mol, implying that the *cyclo*-N_5^- ion is thermodynamically more stable than HN_5 in the crystal structure. The highly symmetrical forces in the crystal also support the *cyclo*-N_5^- ion in place of HN_5.

REFERENCES AND NOTES

1. R.-Y. Huang, H. Xu, *Science* **359**, eaao3672 (2018).
2. C. Zhang, C. Sun, B. Hu, C. Yu, M. Lu, *Science* **355**, 374–376 (2017).
3. P. Müller, R. Herbst-Irmer, A. L. Spek, T. R. Schneider, M. R. Sawaya, *Crystal Structure Refinement: A Crystallographer's Guide to SHELXL* (Oxford Univ. Press, 2006).
4. P. Kelemen, J. Lugtenburg, B. Klumperman, *J. Org. Chem.* **68**, 7322–7328 (2003).
5. H. Knicker, H.-D. Lüdemann, K. Haider, *Eur. J. Soil Sci.* **48**, 431–441 (1997).
6. H. W. Johnston, G. G. Briggs, M. Alexander, *Soil Biol. Biochem.* **4**, 187–190 (1972).
7. J. Kiwi, A. Lopez, V. Nadtochenko, *Environ. Sci. Technol.* **34**, 2162–2168 (2000).
8. L. Zhang *et al.*, *J. Phys. Chem. B* **120**, 11510–11522 (2016).
9. D. A. Johnson, *Metals and Chemical Change* (Royal Society of Chemistry, 2002), vol. 1.
10. J. W. Ball, D. K. Nordstrom, *User's Manual for WATEQ4F, with Revised Thermodynamic Data Base and Text Cases for Calculating Speciation of Major, Trace, and Redox Elements in Natural Waters* (U.S. Geological Survey, Menlo Park, CA, 1991).
11. L. W. Chung *et al.*, *Chem. Rev.* **115**, 5678–5796 (2015).

ACKNOWLEDGMENTS

Supported by National Natural Science Foundation of China grants 11604017 and 21772092, the JASMIN infrastructure for parallel computing, Fundamental Research Funds for the Central Universities grant 30917011101, and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

2 January 2018; accepted 23 January 2018
10.1126/science.aas8953

TECHNICAL COMMENT

INORGANIC CHEMISTRY

Comment on "Synthesis and characterization of the pentazolate anion *cyclo*-N_5^- in $(N_5)_6(H_3O)_3(NH_4)_4Cl$"

Rong-Yi Huang* and Heng Xu

Zhang *et al*. (Reports, 27 January 2017, p. 374) reported synthesis of a *cyclo*-N_5^- ion putatively stabilized in a solid-state salt by hydrogen bonding from surrounding counterions. We performed theoretical calculations suggesting that HN_5 would be favored over the anion in the reported pentazolate salt via proton transfer.

Zhang *et al*. (*1*) reported a *cyclo*-N_5^- ion stabilized in the $(N_5)_6(H_3O)_3(NH_4)_4Cl$ salt via hydrogen bonding from the adjacent NH_4^+ and H_3O^+ ions. The structure was determined by single-crystal x-ray diffraction (XRD) analysis. However, the ellipsoid of the O_2 atom in the asymmetry unit plot of the pentazolate salt was much larger than those of the remaining atoms in the reported pentazolate salt, indicating that there should actually be less electron density than assigned in the model. The O_2 atom could also be a nitrogen atom. Therefore, we suggest that the authors should provide solid evidence to confirm the O_2 atom in the pentazolate salt. Moreover, the phase purity of the pentazolate salt and the stability of its crystal structure at room temperature could not be determined on the basis of the physical evidence afforded at present. In particular, the *cyclo*-N_5^- ion in $(N_5)_6(H_3O)_3(NH_4)_4Cl$ possesses C_{2v} symmetry, and thus the analysis of the infrared spectrum of the *cyclo*-N_5^- ion based on D_{5h} symmetry is questionable. In addition, the out-of-plane mode peak of the H_3O^+ ion cannot be observed from the infrared spectrum of the pentazolate salt. On the basis of the nature of the *cyclo*-N_5^- ion and H_2O, the two expected competitive reaction pathways are

$$HN_5\cdots H_2O \leftarrow [N_5\cdots H\cdots H_2O] \rightarrow N_5^-\cdots H_3O^+$$

Thus, in the reported pentazolate salt, the question becomes whether the purported *cyclo*-N_5^- is actually protonated. Recently, Steele *et al*. (*2*) reported that the crystal of pentazole HN_5 discovered using first-principles evolutionary search at high pressures is thermodynamically stable, so HN_5 may also exist in the pentazolate salt. Because the position of the H atom is difficult to confirm by single-crystal XRD methods, we performed theoretical calculations for all the relevant species as an aid to ascertaining the H atom position in the reported pentazolate salt.

All computations were performed with the Gaussian 09 D.01 program (*3*). Density functional theory calculations were carried out using the B3LYP (*4*, *5*), *w*B97X-D (*6*), and M06-2X (*7*) methods in combination with the 6-311++G(d,p) (*8*) basis set and the very large aug-cc-pVQZ (*9*, *10*) basis set for H, O, and N atoms, respectively. For all the species explored here, the initial geometries were further optimized and were followed via frequency analysis to confirm minimum energy structures. The vibrationally stable *cyclo*-N_5^- ion possesses D_{5h} symmetry at different levels, in distinct contrast to that of the *cyclo*-N_5^- ion (C_{2v} symmetry) in the reported pentazolate salt. However, the experimental C_{2v} symmetry of the *cyclo*-N_5^- ion coincides with that of HN_5 (C_{2v} symmetry) obtained at different levels. This indicates the possible formation of HN_5 during the reaction process.

To illustrate that HN_5 can exist in the reported pentazolate salt, we obtained the molecular electrostatic potentials $V(r)$ and average local ionization energies $\bar{I}(r)$ on the molecular surfaces encompassing *cyclo*-N_5^- and H_2O by means of the Multiwfn program (*11*). Figure 1A presents the distribution of the local minima values, $V_{S,min}(r)$ and $\bar{I}_{S,min}(r)$, on the van der Waals surface of *cyclo*-N_5^- and H_2O, which are the electrophilic reactive positions. $V_{S,min}(r)$ and $\bar{I}_{S,min}(r)$

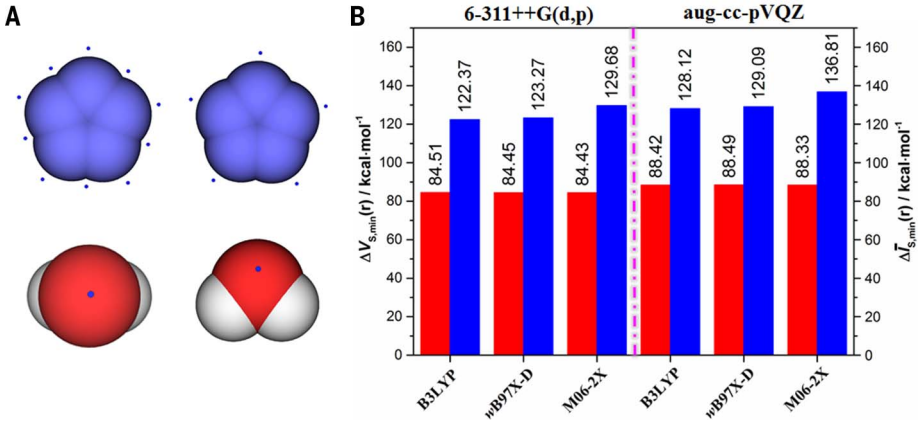


Fig. 1. Computational results. (**A**) Distribution of electrostatic potential minima (left) and average local ionization energy minima (right) on the van der Waals surfaces of *cyclo*-N_5^- (top) and H_2O (bottom) at the B3LYP/6-311++G(d,p) level. (**B**) The calculated $\Delta V_{S,min}(r)$ [$V_{S,min}(r)$ of H_2O – $V_{S,min}(r)$ of *cyclo*-N_5^-] (red) and $\Delta \bar{I}_{S,min}(r)$ [$\bar{I}_{S,min}(r)$ of H_2O – $\bar{I}_{S,min}(r)$ of *cyclo*-N_5^-] (blue) at different levels.

Table 1. Topological parameters at bond critical points of N-H and O-H bonds for HN_5 and H_3O^+ by atoms-in-molecules analysis based on the aug-cc-pVQZ basis set. Values are atomic units.

Parameter	Method	N-H/HN_5	O-H/H_3O^+
ρ	B3LYP	0.355724	0.346522
	*w*B97X-D	0.357297	0.352855
	M06-2X	0.355734	0.341603
$\nabla^2\rho$	B3LYP	−2.089781	−3.071623
	*w*B97X-D	−2.111553	−3.336325
	M06-2X	−2.141610	−2.944185
H	B3LYP	−0.567125	−0.819008
	*w*B97X-D	−0.572378	−0.884981
	M06-2X	−0.576408	−0.786941

Anhui Key Laboratory of Functional Coordination Compounds and College of Chemistry and Chemical Engineering, Anqing Normal University, Anqing 246011, P. R. China.
*Corresponding author. Email: huangry@aqnu.edu.cn

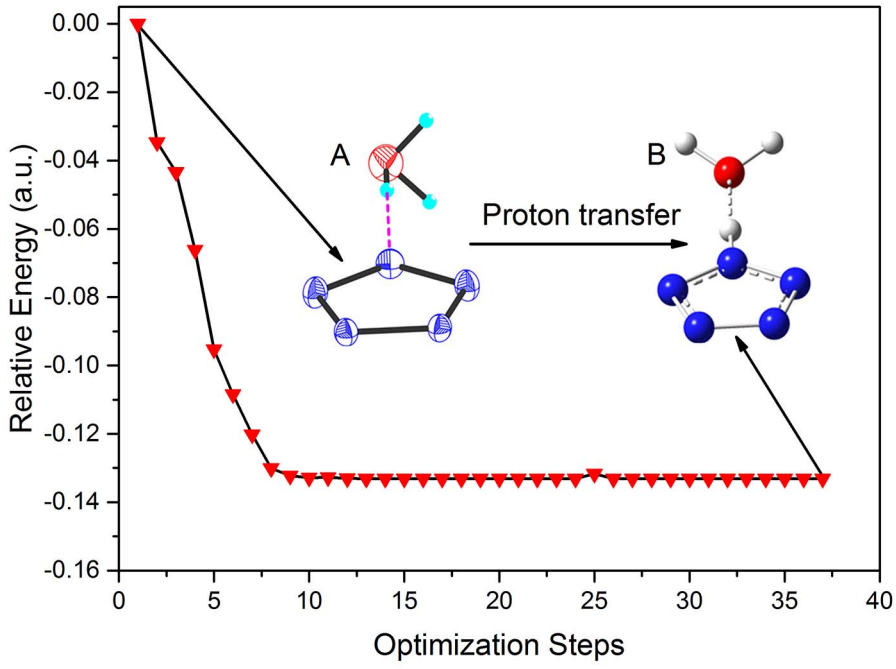


Fig. 2. Geometry optimization. (**A**) Ion pair $N_5^-\cdots H_3O^+$ obtained from the crystal structure of the pentazolate salt. (**B**) A barrierless proton transfer during optimization leads to the final optimized structure at the *w*B97X-D/aug-cc-pVQZ level.

are all associated with the N/O atoms in *cyclo*-N_5^- and H_2O, except for the partial $V_{S,min}(r)$ of *cyclo*-N_5^- located between the two adjacent nitrogen atoms. In comprehensively considering the distribution of $V_{S,min}(r)$ and $\bar{I}_{S,min}(r)$, the strong reactive sites are those associated with the N/O atoms in *cyclo*-N_5^- and H_2O, respectively, which can be attacked by the electrophilic reagents (e.g., M^{n+} and H^+).

As is well known, the addition product, H_3O^+, is stable in corresponding systems. Zhang *et al.* (*12*) reported a cobalt-based metal complex containing the *cyclo*-N_5^- ligand, in which the Co^{2+} and N_5^- ions are linked together though a Co–N bond. Burke and colleagues (*13*), using ^{15}N nuclear magnetic resonance spectroscopy, concluded that HN_5 was also produced and held in solution as a N_5^- ion with Zn^{2+}. Steele *et al.* (*2*) reported a thermodynamically stable crystal of pentazole HN_5 discovered at pressures above 50 GPa using first-principles methods. All of these results, obtained experimentally and theoretically, indicate that our predictive electrophilic reactivities of *cyclo*-N_5^- and H_2O are reasonable. Further, in looking at the $V_{S,min}(r)$ and $\bar{I}_{S,min}(r)$ values associated with the N/O atoms for *cyclo*-N_5^- and H_2O, it is notable that the $V_{S,min}(r)$ and $\bar{I}_{S,min}(r)$ values of *cyclo*-N_5^- were much lower than those found for H_2O at different levels. The computed $\Delta V_{S,min}(r)$ [$V_{S,min}(r)$ of H_2O – $V_{S,min}(r)$ of *cyclo*-N_5^-] and $\Delta\bar{I}_{S,min}(r)$ [$\bar{I}_{S,min}(r)$ of H_2O – $\bar{I}_{S,min}(r)$ of *cyclo*-N_5^-] were also obtained. As shown in Fig. 1B, the $\Delta V_{S,min}(r)$ and $\Delta\bar{I}_{S,min}(r)$ between H_2O and *cyclo*-N_5^- were more than 84 and 122 kcal mol^{-1}, respectively. The large difference values of $V_{S,min}(r)$ and $\bar{I}_{S,min}(r)$ indicate that the electrophilic reactivity of *cyclo*-N_5^- was much higher than that found for H_2O, indicating that HN_5 was more favored in the reaction mixture of H^+, *cyclo*-N_5^-, and H_2O. The topological parameters at the bond critical points (BCPs) of the N-H and O-H bonds for HN_5 and H_3O^+ at different levels are listed in Table 1. The electronic density at the BCPs demonstrates that the N-H bond of HN_5 was also stronger than the O-H bond of H_3O^+. Thus, HN_5 can exist in the reported pentazolate salt.

Deeper insight into possible proton transfer (PT) from H_3O^+ to *cyclo*-N_5^- emerged from theoretical calculations performed for the ion pair $N_5^-\cdots H_3O^+$ (Fig. 2A), which were taken from the crystal structure of the reported pentazolate salt. The optimized stable structure of the ion pair $N_5^-\cdots H_3O^+$ was also obtained, as shown in Fig. 2. We obtained new species in which the proton was bonded to the nitrogen atom of *cyclo*-N_5^-, namely the hydrogen-bonding dimer $HN_5\cdots H_2O$ (Fig. 2B). In addition, the figure presents a continuous decrease in the energy for the species studied, which occurs as a result of the PT from H_3O^+ to *cyclo*-N_5^- and the formation of O···H-N hydrogen bonds and N-H bonds with *cyclo*-N_5^-. The above analysis shows a barrierless PT mechanism during optimization. Furthermore, the calculations at the *w*B97X-D/aug-cc-pVQZ level show that the deprotonation of H_3O^+ requires an energy of 166.62 kcal mol^{-1}, whereas the protonation of *cyclo*-N_5^- can provide an energy of 319.77 kcal mol^{-1}. As a result, besides enough energy for the PT process, there can be extra energy for release, so it is easy for PT to occur in the ion pair studied. The above analysis further indicates that the HN_5 can exist in the reported pentazolate salt via PT.

REFERENCES AND NOTES

1. C. Zhang, C. Sun, B. Hu, C. Yu, M. Lu, *Science* **355**, 374–376 (2017).
2. B. A. Steele, I. I. Oleynik, *J. Phys. Chem. A* **121**, 1808–1813 (2017).
3. M. J. Frisch *et al.*, Gaussian 09, Revision D.01 (Gaussian Inc., Wallingford, CT, 2013).
4. A. D. Becke, *Phys. Rev. A* **38**, 3098–3100 (1988).
5. C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **37**, 785–789 (1988).
6. J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).
7. Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **110**, 5121–5129 (2006).
8. M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **80**, 3265–3269 (1984).
9. R. A. Kendall, T. H. Dunning Jr., R. J. Harrison, *J. Chem. Phys.* **96**, 6796–6806 (1992).
10. D. E. Woon, T. H. Dunning Jr., *J. Chem. Phys.* **98**, 1358–1371 (1993).
11. T. Lu, F. Chen, *J. Comput. Chem.* **33**, 580–592 (2012).
12. C. Zhang *et al.*, *Angew. Chem. Int. Ed.* **56**, 4512–4514 (2017).
13. R. N. Butler, J. C. Stephens, L. A. Burke, *Chem. Commun.* **2003**, 1016–1017 (2003).

ACKNOWLEDGMENTS

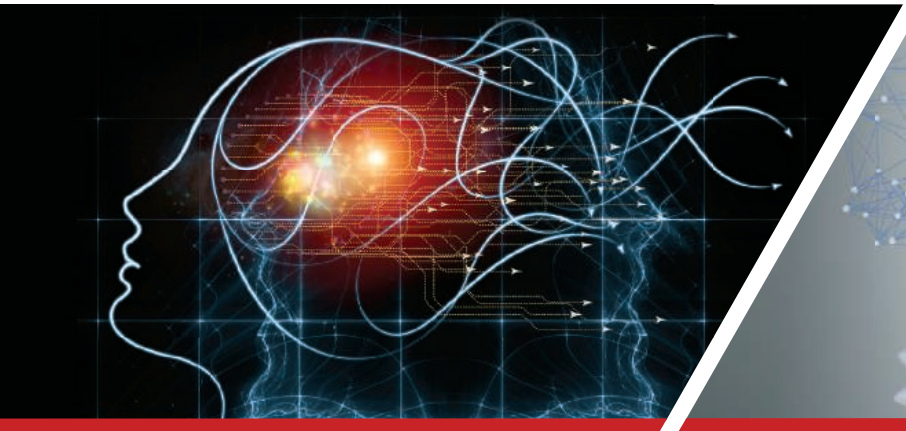
Supported by Natural Science Foundation of the Education Commission of Anhui Province of China grant KJ2017A347.

12 July 2017; accepted 23 January 2018
10.1126/science.aao3672

Science Robotics | AAAS

The First Journal of Interdisciplinary Robotics Research

UPCOMING SPECIAL ISSUES IN 2018



Social Robotics: Summer 2018

Learning—Beyond Imitation: Fall 2018

Submit Your Article for Publication in *Science Robotics*:

- Rapid review, scoop protection, and no article page limits.
- Exposure to an international audience, including 11,000 science journalists.
- Supportive and knowledgeable editorial staff.
- Promotion on the *Science Robotics* website and the potential for accompanying commentary or podcasts.
- AAAS social media and media relations support.

Learn more at **robotics.sciencemag.org**.
Send pre-submission inquiries to **sciroboteditors@aaas.org**.

Waseda University pushes forward with global academic network

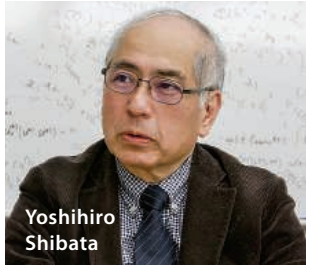
Innovative programs and prioritized funding have propelled Waseda University to record highs in world university rankings, underscoring the university's reputation for openness, dynamism, and diversity.

The role of mathematics in science and technology
—Multiscale Analysis, Modelling and Simulation

"Some people may wonder about the importance of mathematics in society," says **Yoshihiro Shibata, head of the Multiscale Analysis, Modelling and Simulation Unit.** "However, it is difficult to initially test complex new ideas and concepts using experiments alone, such as investigating the flow of blood in arteries. The development of science and technology often requires a cycle of mathematical analysis, computer simulations, and experimental verification."

In this unit, members collaborate with international partners to apply nonlinear partial differential equations, geometry, and quantum mechanics to practical applications such as fluidics, or the flow of gas and liquids. As part of their activities, Shibata and his colleagues also started a doctoral course at the Faculty of Science and Engineering in April 2017, with lectures in quantum mechanics, quantum mathematics, fluid mathematics, numerical fluid dynamics, experimental hydrodynamics, and geometry. Students participate in international symposia as well.

The course is run in collaboration with researchers from Imperial College


Yoshihiro Shibata

London, Technical University Darmstadt, the University of Pisa, Rice University, the University of Pittsburgh, and the University of Hamburg. "Students must spend at least three months at one of our partner institutes," explains Shibata. "Beginning in 2018, we will offer co-teaching courses with the University of Pisa and intend to expand our network to include institutes from China."

Specific research topics include mathematical modelling based on imaging the interaction of molten metals with concrete, related to the decommissioning of nuclear reactors destroyed in the Fukushima accident; analyzing cloud formations to better understand climate change; and observing blood flow for early diagnosis of cardiovascular issues such as aneurysms.

"The emphasis of our education and research is to combine expertise in mathematics and physics to produce useful models to solve major problems facing mankind."

Multiscale Analysis, Modelling and Simulation Unit
www.waseda.jp/inst/sgu/en/unit/multiscale-analysis-modeling-and-simulation

Integrating methodology of political science and economics for innovative models solving global issues
—Positive/Empirical Analysis of Political Economy


Aiji Tanaka

"It is almost impossible to separate economics and politics on the global stage," says **Aiji Tanaka, head of the Positive/Empirical Analysis of Political Economy.** "Raising taxes is both a political and economic decision. So is importing rice or fruit products. But universities in Japan teach politics and economics as separate subjects. Waseda University has combined these subjects in the Faculty of Political Science and Economics."

Tanaka is initiating further reforms to create an international hub for education and research. Notably, students at the Faculty of Political Science and Economics, Faculty of Commerce, and Waseda Institute for Advanced Study could complete their studies entirely in English. "We are hiring international faculty members to teach and conduct research at Waseda," explains Tanaka. "It's a stringent selection process, assessing both publications and teaching ability. We hired six international faculty in 2016, and are interviewing more for the coming year."

Tanaka and colleagues are also collaborating with Stanford University and the University of Essex to organize summer schools at the Waseda campus in Tokyo. Plans include expanding courses to contain wider global perspectives, hiring more international staff, and collaborating on positive theory/empirical analyses of the behavior of political/economic actors and the governance of political systems and multinational corporations with the Said Business School at Oxford University, the School for Advanced Studies in the Social Sciences, the University of California, San Diego, and other universities abroad.

"Waseda University has a 130-year history of pioneering political economics–based education and research," says Tanaka. "Our faculty members are training our students to be fluent in the languages of both economics and politics, and we will continue this tradition on a global scale to resolve the challenges of the 21st century."

Positive/Empirical Analysis of Political Economy Unit
www.waseda.jp/inst/sgu/en/unit/empirical-analyses-of-political-economy

eppendorf

The Next Benchmark

The new refrigerated Centrifuge 5910 R

The Centrifuge 5910 R sets the next benchmark in versatility, capacity and user convenience.

Its main swing-bucket rotor holds both conical tubes and plates – no need to change rotor buckets or adapters. This improves handling and saves time.

> Max. capacity: 4 x 750 mL or 36 x 50 mL conical
> Excellent rotor versatility
> Modern operating system with outstanding functionality
> Advanced temperature management that keeps your samples safe

www.eppendorf.com/next-benchmark

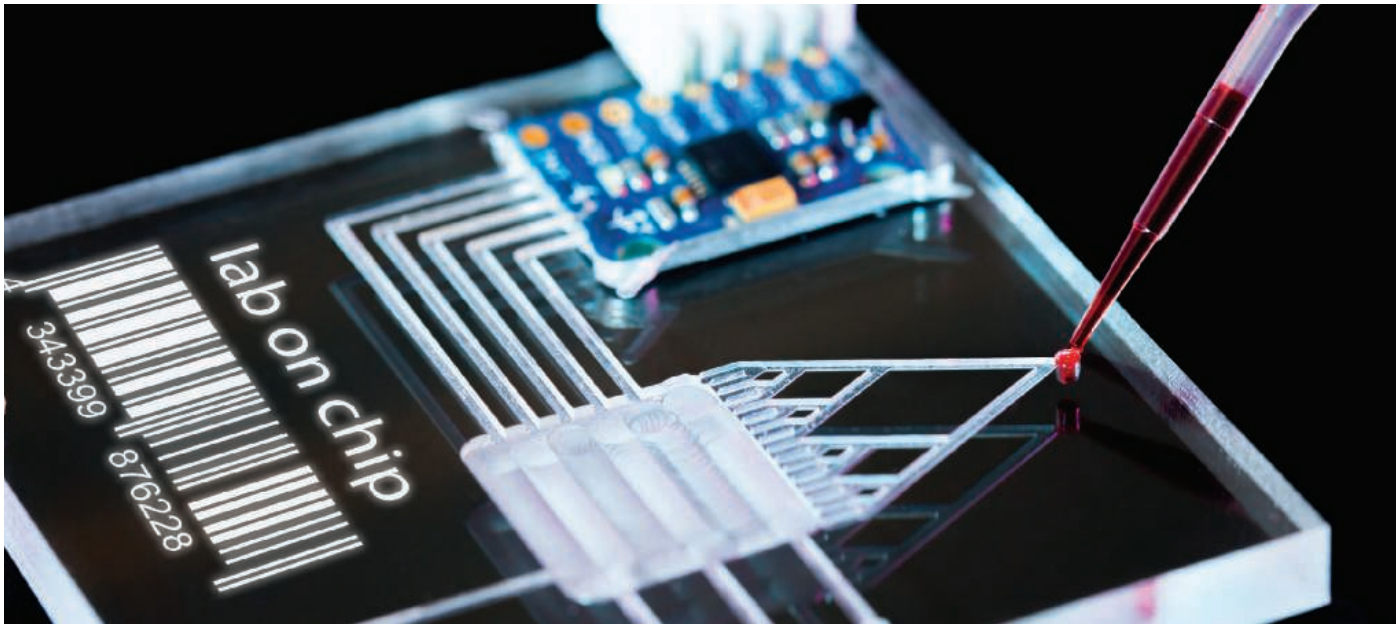


IMAGE: © SCIENCE PHOTO/SHUTTERSTOCK.COM

In vitro veritas:
Biosensors and microarrays come to life

Advances in biosensors and microfluidic devices are driving a quiet revolution in biomedical research, which could lead to the reduction or elimination of animal use in many experiments. **By Alan Dove**

For decades, laboratory biologists have regarded animal models as a necessary evil. While some activists decry their use on moral grounds, even the most practical-minded researchers acknowledge fundamental problems with them. Animals are expensive, provide only imperfect replicas of human biology, and introduce numerous variables into experiments that can be difficult or impossible to control.

These flaws aren't purely academic. Pharmaceutical researchers have struggled for years with late-stage development failures, in which drugs that look promising in multiple animal systems turn out to be useless or even toxic in humans. Nonhuman models have simply been the least bad tool for detailed studies on human biology.

Thanks to advances in completely unrelated fields, though, that grim situation is starting to change. Using improved biosensors that can monitor microscopic compartments, and microfluidic devices that combine the miniaturized features of computing chips with living cells and tissues, researchers are now building systems that can reduce or even eliminate the need for laboratory animals while simultaneously yielding better data.

Diamonds are for sensors

Pharmacokinetics researchers have been among the most enthusiastic proponents of improving on animal models. Ideally, researchers would like to know exactly where a drug goes in the body and how and when it's processed, generating a detailed history from dosing to metabolism to excretion. In practice, that has entailed cumbersome techniques such as medicating numerous animals and sacrificing them for analysis at different times. Besides being laborious and expensive, these experiments give only coarse measures of drug metabolism across time and body compartments.

In recent years, investigators have implanted tiny, electrochemical carbon-based sensors into animals, which provide measurements of metabolic changes in real time in a single animal. That's worked for natural signals such as dopamine and serotonin levels in the brain, but the high background noise and low dynamic range of carbon probes make them poor choices for studying drug metabolism.

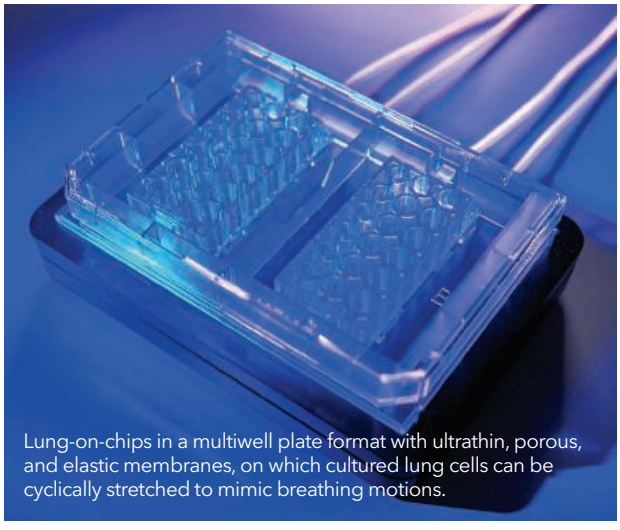
Chemists and biologists are collaborating to overcome those limitations. Yasuaki Einaga, professor of chemistry at **Keio University** in Yokohama, Japan, has worked on electrochemical sensors for decades. Einaga's group has found that boron-doped diamond probes are particularly good at detecting electrochemical changes in solutions. His researchers have tested these probes in systems ranging from wastewater treatment to chemical synthesis.

"To further explore their biological applications, in 2007 [we] succeeded in downsizing the ... probes to micro scale," says Einaga. Hiroshi Hibino, professor of molecular physiology at **Niigata University School of Medicine** in Niigata, Japan, saw the system's potential for pharmacokinetics applications, and began collaborating with Einaga in 2011.

The two labs have since found that boron-doped diamond probes can detect electrochemical changes caused by several classes of drugs in live animals and explanted organs, cont.>

providing real-time data on drug concentrations. "The number of drugs measurable by these sensors is much more than that of any conventional electrode," says Einaga. In one recent paper, for example, the team accurately measured concentrations of a diuretic, an anticonvulsant, and a chemotherapeutic agent in guinea pigs' inner ears.

"[We] have a plan to test the kidney, [and] both groups are discussing a further collaboration to develop an implantable microsensing system to track [a drug] and its effects longitudinally in organs such as the brain," says Hibino. He adds that the two labs are now in discussions with several other researchers who want to use the technology in fields ranging from oncology to psychopharmacology.



Lung-on-chips in a multiwell plate format with ultrathin, porous, and elastic membranes, on which cultured lung cells can be cyclically stretched to mimic breathing motions.

"We want to be able to take cells from a patient and test them ... and see which therapy is going to be best for that patient."
– Olivier Guenat, CEO of AlveoliX.

Plumbing codes

While improved biosensors can extract more and better data from each animal, they can also be combined with microfluidic devices to replace an animal outright, at least for some types of experiments. Microfluidic device makers borrow techniques from the electronics industry to manufacture miniature laboratories on chip-like wafers. The small size of the chips' channels and chambers means that fluids can flow through them rapidly. Microfluidic devices can also include complex structures that mimic biological compartments, making cultured cells behave more naturally. Finally, the chips can be mass-produced on semiconductor manufacturing equipment, making them relatively cheap.

Over the past few years, biologists have built a series of progressively more complex microfluidic devices, which have essentially evolved into artificial, miniaturized human organs. Biosensors built into these systems allow researchers to watch in real time as, for example, a tiny human-like liver or kidney reacts to an incoming dose of a drug.

The power of these new systems can be intimidating, though. "The biggest challenge for us instrument providers is to educate people," says Fabien Crespo, head of marketing and sales at **Elveflow** in Paris, France. Crespo adds that "people are kind of afraid of microfluidics, which is a lot about plumbing."

Elveflow and a few other companies have made this microscopic "plumbing" their main focus. "Researchers can now find microchips suited to their applications, and we are providing everything to control liquid flow inside those microchips," says Crespo. Because the microfluidics field has standardized the different types of fittings that go into and out of chips, a single liquid-handling system can adapt to changing laboratory needs. In Elveflow's setup, for example, a point-and-click computer interface allows researchers to build multistep protocols controlling liquid flows. More advanced users can use a scripting interface to drive the system programmatically.

Such flexible, modular systems are undoubtedly one reason microfluidics are becoming so popular. "We are seeing a lot of expansion in the field, especially over the last two years," says Crespo. He adds that while the first users of microfluidics were mostly in academic research laboratories, he's seen increasing demand from industrial scientists. That's likely driven by interest in developing new preclinical drug screening assays, but Crespo also expects microfluidics to start showing up in point-of-care diagnostic tests within the next few years.

The government chips in

The groundwork for the new boom in microfluidics began nearly a decade ago, through a combination of basic research and farsighted government assistance. When researchers in a few academic laboratories began building "organ-on-a-chip" systems, administrators at the **U.S. National Institutes of Health** (NIH), the U.S. Food and Drug Administration, and the Defense Advanced Research Projects Agency took notice. "It became apparent ... that this was going to be a promising tool and technology," says Danilo Tagle, associate director for special initiatives in the Office of the Director at NIH in Bethesda, Maryland.

Tagle and his counterparts at other agencies convened a meeting with researchers in 2011 to discuss turning the new tissue and organ chips into practical models for drug testing and regulatory approval. "Numerous studies in the past few years have indicated that as much as 90% of the attrition in drug development is caused by failure to predict safety as well as efficacy when using [cell] culture systems and in vivo animal models," says Tagle. "What we're hoping [is that] these tissues on chips or organs on chips can fill in the missing information we need in order to have better success in drug development."

To help make that happen, NIH created the **Tissue Chip for Drug Screening** program in 2012. Collaborating with several academic labs, the initiative's five-year goal was to build organs on chips that could yield accurate predictions about drug responses in humans. Researchers funded by the program had to build devices that could keep cells alive in a setting that would mimic specific organs or tissues, and incorporate biosensors to measure the cells' physiology. The program has now established several independent testing centers for organ chips, using over 100 drugs that cleared traditional preclinical testing only to fail in clinical trials. "We're asking the question, 'Would [a given chip] have predicted the adverse event that the 2D culture systems and the animal models were unable to predict?'" says Tagle. Organ chips

PHOTO: COURTESY ALVEOLIX AND UNIVERSITY OF BERN, SWITZERLAND

Featured participants

AlveoliX
www.alveolix.com

ARTORG Center for Biomedical Engineering Research
www.artorg.unibe.ch

Elveflow
www.elveflow.com

Emulate
emulatebio.com

Hesperos
www.hesperosinc.com

Keio University
www.keio.ac.jp/en

Niigata University School of Medicine
www.med.niigata-u.ac.jp/eng/top.html

Tissue Chip for Drug Screening Program, U.S. National Institutes of Health
ncats.nih.gov/tissuechip

Wyss Institute
wyss.harvard.edu

that do well in these tests could be used to supplement or even replace animal data in future regulatory filings by drug makers.

Having concluded its first phase in July, with several promising systems being tested for assaying drug toxicity, the Tissue Chip program is now focusing on models for drug efficacy. These single- and multiorgan systems will be built to mimic particular diseases, including Parkinson's disease, amyotrophic lateral sclerosis, and osteoarthritis.

Deep breaths

The main advantage organ chips have over traditional cell culture systems is their ability to simulate the complex structures and dynamics of intact organs. Instead of being restricted to flat, solid surfaces, chip makers can incorporate microscopic channels, curves, pores, and layers for the cells to populate. The carefully controlled, programmable fluid flows of a microfluidic device add to the realism of the technology, recreating the forces those same cell types would encounter in living humans.

Lungs provide a good example of both the challenges and the potential of this approach. In the human lung, cells maintain a semipermeable barrier with two distinct sides, allowing gas exchange between air and blood, while keeping the two fluids separate and withstanding regular cycles of flexing with each breath. Researchers at Harvard University's **Wyss Institute** first mimicked this system with a lung chip in 2010. **Emulate** in Boston, Massachusetts, is continuing development on that and other organ chips.

The Boston team isn't alone, though. "We saw this very interesting paper from the Wyss Institute, [and] we wanted to go further [with the concept]," says Olivier Guenat, CEO of **AlveoliX** in Bern, Switzerland. Guenat is also a group head at the University of Bern's **ARTORG Center for Biomedical Engineering Research,** where his lab collaborates with clinical teams treating patients with lung diseases.

To simulate the structure and three-dimensional deformation of a lung, Guenat's team developed a platform that seeds cells onto thin layers of silicone containing regularly spaced holes. Only 3 micrometers thick, the layers are nonetheless strong enough to withstand repeated, breath-like flexing. Having accomplished that, "the second biggest challenge was that we wanted ... to develop something that is very easy to use," says Guenat.

After several design iterations, AlveoliX now has a prototype system that maintains 12 lung chips on a standard-size multiwell plate. This arrangement enables users to handle and test the chips with existing microscopes, plate readers, and other common laboratory equipment. Ultimately, "we want to be able to take cells from a patient and test them ... and see which therapy is going to be best for that patient," says Guenat.

The chips should also prove useful for preclinical studies, allowing scientists to control fluid flows, mechanical stresses, and other parameters with far greater precision than they can in living models, while simultaneously eliminating the challenges of animal handling. "I don't know any biologist who likes to sacrifice animals, and with organs on chips [now available], we really want to reduce animal testing," says Guenat.

Concerto for organs

As organ chips become more established, scientists in the field are already pushing microfluidics and biosensor technology to the next logical step: multiorgan systems. In theory, one could simply pump culture medium through different chips in series, circulating drugs and metabolites within a high-tech homunculus. The reality turns out to be considerably more challenging.

"There's just a lot of things that can go wrong when somebody tries to do this," says Mike Shuler, president and CEO of **Hesperos** in Orlando, Florida. The company specializes in multiorgan chip systems. Shuler, who is also a professor of engineering at Cornell University in Ithaca, New York, says even people intimately familiar with multiorgan chip systems can have trouble getting them running. "When I'm transferring technology from my [academic] lab down to the company, sometimes it takes a few iterations for us to get it to work right," he says.

As a result, Hesperos has built its business on offering multiorgan models as a service, rather than trying to sell and support them as stand-alone products. To date, the company has collaborated with several pharmaceutical companies interested in testing drug leads on chips. Shuler says Hesperos typically builds systems with four or five interconnected organs. "The liver is almost always the critical one, then cardiac and neuromuscular junctions ... are probably the [next] most popular," he says, adding that the company has also built systems incorporating artificial skin, gastrointestinal tracts, and blood–brain barriers.

One of the biggest challenges has been keeping the systems running long enough for extended metabolic testing. "We try to operate out to 28 days," says Shuler. At that timescale, the difference in solubility between oxygen and carbon dioxide in the cells' medium can cause gas bubbles to accumulate, disrupting the system's tightly controlled fluid flows. Hesperos has addressed that problem by eliminating the pumps normally used to control microfluidic devices, and using a carefully designed gravity flow system instead.

As the technology continues to develop, proponents of microfluidic systems expect their popularity to skyrocket in the next few years. "We're already being put into workflows for large and small pharma, [and] at this point we've been able to meet just about every single milestone that people have set out for us," says James Hickman, Hesperos' chief scientific officer. Shuler projects that human organ chips "eventually will, I think, replace animals [or] so greatly reduce animal use that it's a much less significant part of the drug development process, because you are getting data on human systems."

Alan Dove is a science writer and editor based in Massachusetts.

Cell Culture Flow Systems

Quasi Vivo Systems are a series of advanced, interconnected cell culture flow systems, specifically designed to improve cellular phenotypes and human in vivo relevance of in vitro cell culture models. By introducing flow to the in vitro environment, Quasi Vivo enables you to generate models with improved predictive value, greatly improving confidence in the validity of your results. The system consists of interconnected cell culture chambers and a peristaltic pump to create a continuous flow of media over cells. It is available with three different culture chambers (QV500, QV600, and QV900) to support a wide range of applications, including submerged cell culture, coculture, and modeling of air–liquid and liquid–liquid interfaces. Not only is the system easy to set up, it also enables close monitoring of variables during an experiment. Furthermore, its large scale and user configurability allow researchers to perform assays that are not possible with microfluidic systems.

Lonza
For info: 800-638-8174
www.lonza.com/quasi-vivo

Micropatterned Hepatocyte Co-Cultures

HepatoPac Kits are in vitro tools used for predicting likely in vivo outcomes and mechanisms of toxicity with respect to transport and metabolism of drugs and chemicals. HepatoPac plates contain micropatterned co-cultures (MPCCs) of primary hepatocytes and stromal cells. This technology replicates the physiological microenvironment of the liver and allows hepatocytes to exhibit normal metabolic activity for over 4 weeks for both short- and long-term toxicology and efficacy studies during preclinical drug discovery. Hepatocyte health, functionality, and liver enzyme activity can be extensively characterized on the HepatoPac platform. HepatoPac products are available with human, rat, dog, or monkey hepatocytes, or in multispecies formats.

Ascendance Biotechnology
For info: 781-391-0205
ascendancebio.com

Disease States Cell Culture

HemaCare supplies researchers with a wide variety of well-characterized human disease–state primary cells and blood components through our U.S.-based collection network for use in basic, clinical, and translational research studies. The company can provide whole blood, plasma, serum, and peripheral blood mononuclear cells for key research areas from diseases such as acute myeloid leukemia, Crohn's disease, and syphilis. In addition to the products listed in the catalog, HemaCare can offer fresh and cryopreserved primary cells and blood components from other disease states—including various cancers such as renal cancer and multiple myeloma, allergies, and many other conditions.

HemaCare
For info: 877-310-0717
www.hemacare.com/store/products/disease-state.html

Single-Cell Analysis System

The Cyto-Mine Single-Cell Analysis System automates single-cell analysis, sorting, imaging, and dispensing. It is a single, compact, GLP-compliant system that fits in a standard biological safety cabinet. Traditionally, up to three different instruments would be required for each step, resulting in a costly, time-consuming process, using up valuable lab space, and increasing the risk of sample contamination. Cyto-Mine can deliver up to 10 million tests per day, far more than other systems on the market that deliver 10,000 tests in three weeks. The high-throughput system uses Sphere Fluidics' patented picodroplet technology to encapsulate a single cell in growth media and trap secreted molecules, such as antibodies, from the cell as it grows. Since each cell is compartmentalized in the disposable Cyto-Cartridge, monoclonality is assured, and the unique workflow enables selective screening of single cells to find rare lead candidates.

Sphere Fluidics
For info: 888-258-0226
www.spherefluidics.com/products/integrated-systems/cyto-mine

Automated Pipetting Robot

The BRAND Liquid Handling Station (LHS) pipetting robot handles routine tasks at high speed and with the highest precision. It closes the gap between electronic pipettes and complex, expensive pipetting robots. The LHS is a quiet, compact instrument with a unique front door, allowing it to be set up in tight quarters. The instrument has seven working positions and five available liquid ends with single-channel volumes up to 1,000 µL and multichannel volumes up to 300 µL. The system has the flexibility to move liquids between reservoirs, tubes, and plates. Applications include PCR, quantitative PCR, and ELISA setup; serial dilutions; microplate replication and reformatting; DNA normalization; cherry picking; cell cultures; and general liquid transfers. Adapters and racks bring consumables to an even height, reducing unnecessary vertical movements to save time during pipetting. The LHS is supplied with extremely user-friendly software—liquid handling protocols are created with ease.

BrandTech Scientific
For info: 888-522-2726
www.brandtech.com/product/liquid-handling-station

Syringe Pump

Hamilton Company expands its PSD/4 precision syringe drive product family with the launch of PSD/4 Smooth Flow. The Smooth Flow delivers exact flow rates in nanoliters per minute with total dispense times up to 8 hours. It is perfect for engineers seeking an economical way to automate processes that were developed on lab-scale pumps, and is also ideal for flow cytometry and microfluidics applications where smaller sample volumes, faster analysis time, and more sensitive detection mechanisms require accurate and extremely small precision flow rates. The compact PSD/4 design now has 192,000 steps of resolution across the 30-mm stroke. The pump is compatible with a variety of catalog and custom valves as well as standard and long-life syringes from 12.5 µL to 12.5 mL.

Hamilton Company
For info: 800-648-5950
www.hamiltoncompany.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

The Wenzhou FGF Team Seeks Top Talent To Discover Opportunities:

Zoom in on the therapeutic applications of FGFs in China

Based at both Wenzhou Medical University and Wenzhou University the research team led by Prof. Xiaokun Li is devoted to FGF basic and translational research with their clinical application as an ultimate goal. The therapeutic modalities using FGF formulations have evolved from the initial external and topical administration to implantable medical device as well as injection. The clinical indications have expanded from trauma and diabetic ulcer treatment, to neurological repair/regeneration. In the immediate future, FGFs will also be instrumental in the treatment of major metabolic diseases, such as diabetes and atherosclerosis.

The fibroblast growth factor (FGF) family is comprised of a group of structurally related protein ligands that signal through FGF receptor tyrosine kinases (FGFRs) to carry out a plethora of vital functions in development, metabolism, tissue homeostasis and repair after injury. Historically, translational research of FGF/FGFR has focused on the oncology aspect in the West, whereas the research team led by Prof. Xiaokun Li based at both Wenzhou Medical University and Wenzhou University is devoted to FGF biotechnology and the delivery of engineered clinical grade FGFs for tissue repair and regeneration.

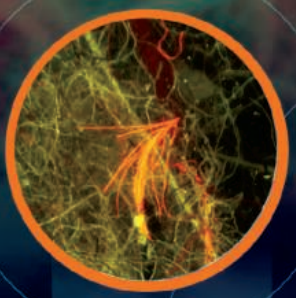
The FGF family consists of 18 mammalian FGFs divided into 6 subgroups on the bases of their sequence homology, phylogenetics and structural characteristics. Five of the subgroups are considered canonical FGFs capable of high affinity binding to heparin sulfate (HS) and acting locally as paracrine molecules. The potential of FGFs in promoting cell proliferation, survival, angiogenesis, migration, and differentiation, have been explored for therapeutic applications in the setting of tissue repair/regeneration and also cancer therapy. Li' s team, which started from Jinan University, Guangzhou, overcame several major bottlenecks for FGF protein engineering and recombinant production such as poor protein solubility and stability, enabling them to develop and license the first FGF protein drug (FGF2) in the world for clinical use. The clinical trial of the topical FGF2 biologics, led by Prof. Xiaobing Fu and first reported in the Lancet, proved its beneficial effect in accelerating healing of burn wound, skin flap transplantation, and diabetic ulcers. As of June 2017, in China alone, FGF biologics have been used in 80 million cumulative patients/cases with great clinical and socioeconomic benefit. Importantly, the safety record of this treatment has been excelling; during a 20-year period post-FGF treatment of clinical follow-up, no excessive hyperplasia or other major adverse effects have been observed. The FGF biologics have transformed the clinical practice of trauma management from the traditional anti-infection and anti-inflammation therapies to include FGF-induced coordinated pro-active repair and functional regeneration. In the past 10 years, Li and his colleagues have also successfully developed novel formulations for FGF1 and FGF2, as well as FGF7 and FGF10 biologics, which are currently in different phases of clinical evaluation. A non-mitogenic mutant FGF1 formula is also under preclinical evaluation for the treatment of Type 2 Diabetes.

The endocrine FGFs are relatively new members of the FGF family with much reduced affinity toward HS and activate FGFRs with Klotho as a cofactor and exhibit distinct regulatory activity in various metabolic processes including glucose, lipid, bile acid, vitamin D and phosphate metabolism and energy homeostasis. These atypical FGFs, as exemplified by FGF21, present therapeutic potentials for a myriad of major metabolic diseases such as diabetes, obesity, cardiovascular and renal diseases, amongst others. Besides the aforementioned paracrine FGFs, Li' s team has also undertaken major effort on basic and translational research on endocrine FGFs, particularly FGF21 and more recently FGF23. Li and his colleagues first discovered the role of adiponectin in mediating the metabolic effect of FGF21 on energy metabolism and insulin sensitivity, as well as protection against atherosclerosis. Currently the team has completed preclinical studies and is in the process of applying for CFDA approval for clinical trial on FGF21 for the treatment of diabetes. In collaboration with Prof. Mohammadi from New York University, Li' s team provides structural insight into the activation of Phospholipase C by the concerted action of two FGF receptor molecules. A recent Nature publication further solved the atomic structure of the α-Klotho/FGFR1c/FGF23 ternary complex providing a basis for understanding how klothos and endocrine FGFs interact, therefore opening up new avenues for structure-based drug design. Studies from Li' s team have already unveiled unexpected therapeutic activity of FGF1 toward diabetic nephropathy and demonstrated that mitogenic and metabolic activities can be uncoupled by tuning FGF1/FGFR dimer stability. These mechanistic findings will likely lead to future drug discoveries targeting FGFs for the treatment of a variety of human diseases.

Armed with the rich and validated experience of "Chinese Style" protein drug development, the Wenzhou FGF team and its network of international collaborators, which together constitute the "Wenzhou FGF family" is pushing ahead to make significant contributions to translate FGF research discoveries into clinical application in the years to come.

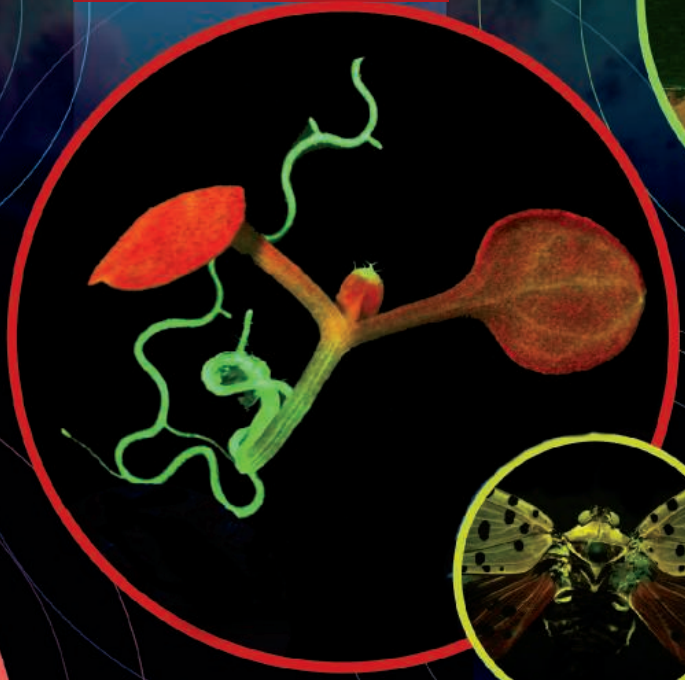
The Wenzhou FGF team welcomes scholars from around the world to join them. To learn more about available opportunities, interested candidates should email (profxiaokunli@163.com or Zhang.jinsan@mayo.edu) the Wenzhou FGF team.

there's more to see...
WITH FLUORESCENCE

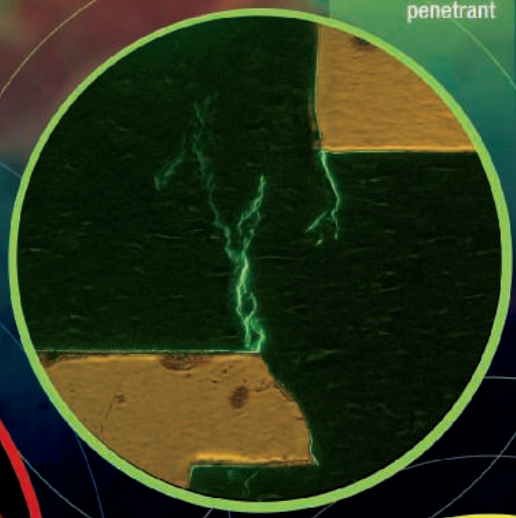


FORENSIC SCIENCE
'Dust bunny' under blue excitation

LIFE SCIENCE
Arabidopsis sp. expressing green fluorescent protein (GFP) in roots

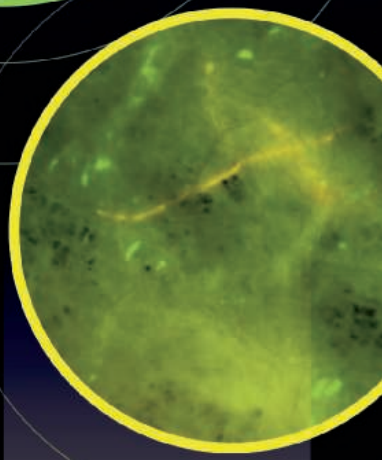


INDUSTRY
Crack feature, fluorescent penetrant



PHOTOGRAPHY
Inchworm on leaf, fluorescence, Maine

ENTOMOLOGY
Natural fluorescence of a Leafhopper

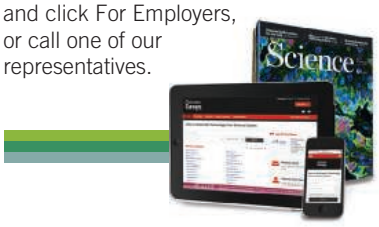


ENVIRONMENTAL SCIENCE
Microplastics in the oceans and water supplies, the food chain, and consumer products have become pervasive.

Science Careers

SCIENCE CAREERS ADVERTISING

For full advertising details, go to ScienceCareers.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

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

KING'S College LONDON

Professor of Analytical & Forensic Sciences

School of Population Health & Environmental Sciences,
Department of Analytical, Environmental & Forensic Sciences

We are seeking a resilient and inspirational Director able to work strategically within King's Forensics. This full time permanent post is based within the Faculty of Life Sciences & Medicine, School of Population Health & Environmental Sciences and the Department of Analytical, Environmental and Forensic Sciences located at the King's Waterloo Campus. It offers an exciting opportunity to establish and lead research in the field of analytical science within an internationally recognised and multi disciplinary university environment.

It is anticipated that you will have research expertise in an area of analytical science that both complements and broadens our current research interests of toxicology, drug misuse in sport, forensic genetics, and environmental forensics including evidence detection. Good knowledge, preferably with practical experience, of operational forensic science is desirable. Applicants should have a track record of internationally excellent high impact research, demonstrated by a record of quality publications and proven ability to attract external funding. Experience of supervising postgraduate research students is essential. Applicants should have experience of developing and delivering innovative graduate and/or professional education. The salary for this post will be in accordance with the level of appointment.

The appointment will be made, dependent on relevant experience and settled by negotiation. The minimum professorial salary at King's College, London is £66,084 per annum, plus £2,923 London Weighting Allowance per annum.

Further details and application packs are available on the College's website at www.kcl.ac.uk/jobs. All correspondence should clearly state the job title and reference number A9/JVA/0474/18-SS. If you have any queries please contact the recruitment team at recruitmentteam2@kcl.ac.uk.

The deadline for applications is midnight on 3rd April 2018.

香 港 大 學
THE UNIVERSITY OF HONG KONG

Founded in 1911, the University of Hong Kong is committed to the highest international standards of excellence in teaching and research, and has been at the international forefront of academic scholarship for many years. The University has a comprehensive range of study programmes and research disciplines spread across 10 faculties and over 140 academic departments and institutes/centres. There are 28,000 undergraduate and postgraduate students who are recruited globally, and more than 2,000 members of academic and academic-related staff coming from multi-cultural backgrounds, many of whom are internationally renowned.

Post-doctoral Fellowships

Applications are invited for a number of positions as Post-doctoral Fellow (PDF), at the University of Hong Kong. Appointments will be made for a period of 2 to 3 years and the appointees must be in post on or before February 28, 2019.

PDF posts are created specifically to bring new impetus and vigour to the University's research enterprise. Positions are available from time to time to meet the strategic research needs identified by the University. Positions are available in the following Faculties/Departments/Schools/Centres/Units:

- Faculty of Dentistry
- Faculty of Education
- Department of Electrical and Electronic Engineering
- Department of Mechanical Engineering
- School of Biomedical Sciences
- Department of Clinical Oncology
- Department of Microbiology
- Department of Paediatrics and Adolescent Medicine
- Department of Pathology
- Department of Pharmacology and Pharmacy
- School of Public Health
- Department of Surgery
- School of Biological Sciences
- Department of Chemistry
- Department of Physics
- Department of Statistics and Actuarial Science
- Department of Psychology
- Partner State Key Laboratory for Liver Research

Post-doctoral Fellows

PDFs are expected to devote full-time to research. Applicants should be doctoral degree holders having undertaken original research that has contributed to the body of knowledge. A highly competitive salary commensurate with qualifications and experience will be offered. Annual leave and medical benefits will also be available.

Procedures

Prospective applicants are invited to visit our webpage at http://jobs.hku.hk/ to view the list of the Faculties/Departments/Schools/Centres/Units and their research areas for which PDF positions are currently available. Before preparing an application, they should contact the Head of the appropriate academic unit, or the contact person as specified, to ascertain that their research expertise matches the research area for which a vacant PDF post is available.

Applicants must submit a completed University application form, which should clearly state **which position** they are applying for; and in which academic discipline. They should also provide further information such as details of their research experience, publications, research proposals, etc.

Application forms (341/1111) can be downloaded at http://www.hr.hku.hk/apptunit/form-ext.doc and further particulars can be obtained at http://jobs.hku.hk/. **Closes April 22, 2018**. The University thanks applicants for their interest, but advises that only candidates shortlisted for interviews will be notified of the application result.

The University is an equal opportunities employer and is committed to equality, ethics, inclusivity, diversity and transparency; and is committed to a Non-smoking Policy

By **Meghan Wright**

Instagram won't solve inequality

Science Sam is a big name on campus. She's a Ph.D. candidate in the sciences who wants to pursue a career outside of academia, like me. But unlike me, she is our school's science communication, or #scicomm, superstar. Her Instagram page, which aims to show the "fun and trendy" side of science, was recently celebrated in the school's newsletter for increasing the public's trust in scientists. At a career workshop, graduate students were urged to follow Science Sam's example and use #scicomm to build our personal brands as we enter the job market. I already have an Instagram account, but it reflects my interests in photography and baking more than my love of science. The workshop got me thinking: Should my posts focus less on pastries and more on pipettes?

I followed Sam on Instagram, and I soon found many other female graduate students and postdocs whose Instagram pages are filled with pretty selfies, fun videos, and microscope images captioned with accessible language and cute emojis. These researchers assert themselves as scientists who don't fit the stereotypes that are typically applied to women in the field. They are not boring or unfashionable. Instead, their posts demonstrate that they're interested in clothes and makeup, that they're physically active, and that they are attractive romantic partners. By visibly contradicting stereotypes about female scientists, it is clear that they hope to inspire girls to pursue science and to encourage female scientists to showcase their femininity in our male-dominated work spaces.

"Publicly documenting ... the sweet smile I brandish in the lab isn't going to help me."

If I had seen these posts when I was 17, maybe I would have been more confident that I would excel in engineering even though I spent my time in the darkroom developing photos, not in the computer lab learning to code like the male engineers I saw in popular culture. But instead of cheering on Instagram's dynamic and vibrant #scicomm women, I felt an increasing bitterness with each post I came across, though I struggled to understand why.

After long conversations with other women in my lab, I realized that I am not bitter toward the authors of these posts. Although I am annoyed that the majority of the posts seem to celebrate a very narrow representation of femininity, my real bitterness comes from the systemic challenges that these posts are working to address, and from seeing so many young female scientists compelled to turn to their personal social media pages to try to correct the system's failures.

I have come to understand #scicomm on Instagram as a digital demonstration of the efforts that many female scientists exert daily to correct for gender disparities. Female scientists spend demonstrably more time teaching, mentoring, and participating in community outreach than their male colleagues, just as there are far more female #scicommers on Instagram than male. I liken the many hours that Science Sam spends on her Instagram content to the volunteer work a female professor in my department put into organizing a summer program to introduce teens from underrepresented communities to biomedical engineering.

A lot of good comes from the work of female scientists on Instagram. But it disturbs me that these efforts are celebrated as ways to correct for the long held and deeply structured forms of discrimination and exclusion that female scientists face. I wonder whether our efforts should instead be directed toward advocating for policy changes at institutional and governmental levels.

When I next interview for a job, I won't have an Instagram page to show that my love of science doesn't make me boring and unfriendly. Publicly documenting the cute outfit I wear and the sweet smile I brandish in the lab isn't going to help me build a fulfilling career in a field where women hold less senior positions, are paid less, and are continuously underrated. Time spent on Instagram is time away from research, and this affects women in science more than men. That's unfair. Let's not celebrate that. ■

Meghan Wright is a doctoral candidate in the Institute of Biomaterials and Biomedical Engineering at the University of Toronto in Canada.

ILLUSTRATION: ROBERT NEUBECKER