A blood protein that predicts disease and death p. 254

Global lab biosafety needs a risk-based reboot p. 260

Multidimensional integrated quantum photonics p. 285

Science

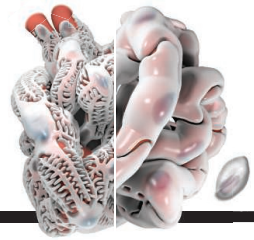
$15
20 APRIL 2018
sciencemag.org

AAAS

RESOLUTION REVOLUTION

Imaging cellular diversity in a developing zebrafish p. 284

CONTENTS



254
A molecular suspect in kidney failure

20 APRIL 2018 • VOLUME 360 • ISSUE 6386

NEWS

IN BRIEF

IN DEPTH

FEATURE

INSIGHTS

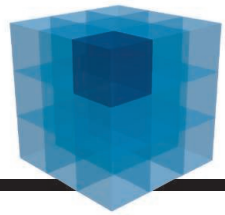
POLICY FORUM

PERSPECTIVES

BOOKS *ET AL.*

LETTERS

CONTENTS



269 & 283
A landscape of genetic interactions

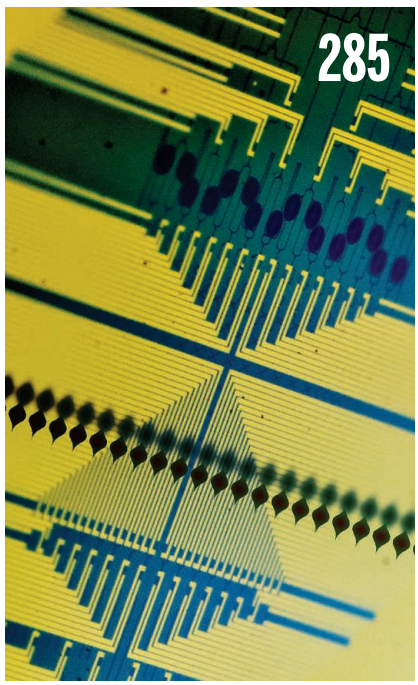
20 APRIL 2018 • VOLUME 360 • ISSUE 6386

RESEARCH

IN BRIEF

RESEARCH ARTICLES

REPORTS



DEPARTMENTS

ON THE COVER

A three-dimensional micrograph of computationally separated cells with their internal organelles, as captured by a movie of the developing zebrafish eye. Combining minimally invasive lattice light-sheet microscopy with adaptive optics to counter optical aberrations present in multicellular specimens enables the study of rapid subcellular processes at high resolution within living organisms, where all environmental factors that regulate cellular physiology are present. See page 284. *Image: Betzig Lab, Janelia Research Campus/HHMI; Kirchhausen and Megason Labs, Harvard Medical School*

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.

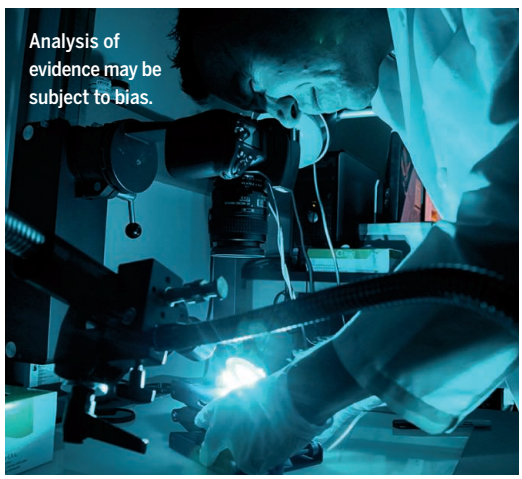
INFORMATION FOR AUTHORS See www.sciencemag.org/authors/science-information-authors

Published by AAAS

EDITORIAL

Biases in forensic experts

Forensic evidence plays a critical role in court proceedings and the administration of justice. It is a powerful tool that can help convict the guilty and avoid wrongful conviction of the innocent. Unfortunately, flaws in forensic evidence are increasingly becoming apparent. Assessments of forensic science have too often focused only on the data and the underlying science, as if they exist in isolation, without sufficiently addressing the process by which forensic experts evaluate and interpret the evidence. After all, it is the forensic expert who observes the data and makes interpretations, and therefore forensic evidence is mediated by human and cognitive factors. A U.S. National Research Council examination of forensic science in 2009, followed by a 2016 evaluation by a presidential panel, along with a U.K. inquiry into fingerprinting in 2011 and a 2015 guidance by the U.K. Forensic Science Regulator, have all expressed concerns about biases in forensic expert decision-making. Where does forensic bias come from, and how can we minimize it?



Analysis of evidence may be subject to bias.

"A consequence of cognitive biases is that science is misused, and sometimes even abused, in court."

Forensic experts are too often exposed to irrelevant contextual information, largely because they work with the police and prosecution. Extraneous information—from a suspect's ethnicity or criminal record to eyewitness identifications, confessions, and other lines of evidence—can potentially cause bias. This can give rise to conclusions that are incorrect or overstated, rather than what forensic decisions should be: impartial decisions, appropriately circumscribed by what the evidence actually supports. A consequence of cognitive biases is that science is misused, and sometimes even abused, in court. Not only can irrelevant information bias a particular aspect of an investigation, it often causes "bias cascade" from one component of an investigation to another and "bias snowball," whereby the bias increases in strength and momentum as different components of an investigation influence one another. Bias also arises when forensic experts work backward: Rather than having the evidence drive the forensic decision-making process, experts work from the target suspect to the evidence.

These problems in forensic decision-making have been largely ignored by the courts, even though there are simple procedural and context management solutions at hand. Biases that arise from exposure to irrelevant contextual information can be minimized by case managers who ensure that only relevant information gets to the appropriate expert. By blinding experts to extraneous information, they only get the particulars that are appropriate for them to have. Bias cascade and bias snowball can be minimized by compartmentalization. For example, the person collecting evidence from a crime scene should not be the expert who analyzes that data in the laboratory. In that way, any exposure to extraneous information at a crime scene does not influence the subsequent analysis. Such measures to minimize bias are standard scientific practices and are commonly used in applied sciences, but forensic science has yet to fully adopt them in practice. Target suspect–driven bias could be minimized by tools such as Linear Sequential Unmasking (LSU), whereby experts are only exposed to the target suspect after they have fully analyzed and documented the actual evidence (such as latent fingerprints, DNA, handwriting, or bullet cartridges found at the crime scene).

A major obstacle in adopting such countermeasures is that many forensic experts have a "bias blind spot" to these implicit biases and therefore tend to deny their existence. Forensic experts frequently present their decisions to the court with great confidence and then incorrectly take the court's acceptance of their findings as confirmation that they have not been biased or made a mistake. Acknowledging that bias can influence forensic science experts would be a substantial step toward implementing countermeasures that could greatly improve forensic evidence and the fair administration of justice.

If we want science to serve society, then it must be properly used in the halls of justice.

–Itiel E. Dror

Itiel E. Dror is a senior cognitive neuroscience researcher at University College London, London, UK. i.dror@ucl.ac.uk

10.1126/science.aat8443

PHOTO: ALEXANDRA GARCIA/THE WASHINGTON POST/GETTY IMAGES

NEWS

> "We can't wait for Washington, especially not now."
>
> **Fred Krupp**, president of the nonprofit Environmental Defense Fund, in *The New York Times*, on his group's plan to launch a satellite to pinpoint leaks of methane, a potent greenhouse gas, from industry worldwide.

IN BRIEF

Edited by **Jeffrey Brainard**

ACTIVISM

March for Science downsizes, evolves

A rally in Washington, D.C., which ended at the U.S. Capitol, was one of many on all seven continents.

Supporters of evidence-based government policies turned out in cities around the world on 14 April for the second March for Science, although in numbers significantly smaller than at the inaugural event a year ago, when by some estimates more than a million people took to the streets. "It's disappointing to see so few people," said John Cosgrove, a retired high school science teacher who traveled from Easton, Pennsylvania, to attend the flagship rally in Washington, D.C., as he did last year. "It's waned a little bit, but the energy is still there." Speakers at that event delivered nonpartisan messages of support for research, but many attendees made it clear, through signs and comments, that they were motivated by U.S. President Donald Trump's stance on climate change and other policies that they said ignored scientific findings. Not every rally involved a march: In Berlin, organizers instead invited scientists to talk about their work with neighbors and other interested people in bars and cafes, an initiative named *Kieznerds* ("neighborhood nerds"). The global grassroots movement has evolved to pursue science advocacy year-round in multiple ways, such as petition drives.

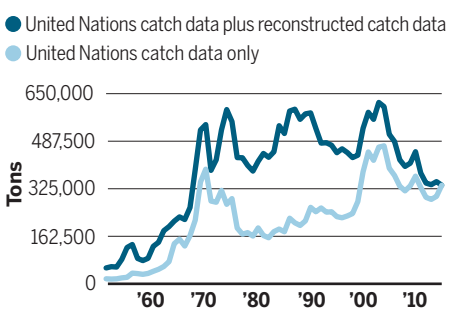
AI-based eye screen approved

MEDICAL DEVICES | The U.S. Food and Drug Administration (FDA) last week announced the first approval of artificial intelligence–based disease screening software that doesn't need a clinician to interpret the results. The computer program, developed by Iowa-based company IDx LLC, analyzes digital images of a patient's retina to detect diabetic retinopathy, a common complication of diabetes that is the leading cause of blindness in U.S. adults. In tests of the software's accuracy, it correctly identified 87.4% of patients with the disease and 89.5% of those without it. Patients who test positive would be referred to an eye care professional. A patient could take the test in a general practitioner's office without visiting a specialist, although a special retinal camera is required. If widely adopted, the software could make it easier for more patients to get regularly screened for the condition.

Deep-sea catches undercounted

MARINE ECOLOGY | Landings from deep-sea commercial fisheries might have been underreported by as much as 42% over the past 6 decades, because fleets excluded bycatch or didn't report all hauls, researchers reported last week in *Frontiers in Marine Science*. Catches of deep-sea species amounted to just 0.5% of global landings from 1950 to 2015, but populations of Greenland halibut, orange roughy, and other species have collapsed under the fishing pressure because they tend to

Catches of deep-sea (>400 meter) species by bottom trawling



CREDITS: (PHOTO) C. AYCOCK/*SCIENCE*; (GRAPHIC) J. BRAINARD/*SCIENCE*; (DATA) L. VICTORERO *ET AL.*, *FRONTIERS IN MARINE SCIENCE*, 5:98 (2018)

Published by AAAS

mature slowly and reproduce at low rates, say Lissette Victorero of the University of Southampton in the United Kingdom and colleagues. Her team analyzed data on deep-sea catches reported to the Food and Agriculture Organization of the United Nations and drawn from a database assembled by the Sea Around Us project at the University of British Columbia in Vancouver, Canada. The undercount means that conservation and management plans for these species may be inadequate, the authors contend.

NIH alcohol probe expands

BIOMEDICINE | The U.S. National Institutes of Health (NIH) in Bethesda, Maryland, last week broadened the scope of its investigation into ties between the alcohol industry and NIH's alcoholism institute to include funding decisions by the institute's director. The investigation initially focused on a National Institute on Alcohol Abuse and Alcoholism (NIAAA) study that began this year, funded largely with $68 million from beverage companies, that will track 7800 volunteers to determine whether one alcoholic drink a day has health benefits. Last month, *The New York Times* reported that NIAAA staff met with beverage industry officials in late 2013 and early 2014 about financing the study, despite a ban on soliciting such contributions. More recently, STAT reported that since joining NIAAA in 2014, Director George Koob has cut back on studies of alcohol advertising that the industry disliked. "We're looking into this in a very aggressive way," said NIH Director Francis Collins at a congressional hearing. In a separate move that underscored concerns over industry funding of biomedical research, Collins said last week that NIH will decline cash contributions from the private sector for a newly launched research project on the U.S. epidemic of opioid overdoses. He left open that NIH and pharmaceutical companies might cooperate in other ways, such as sharing data.

A U.S.-Indian quest for neutrinos

PARTICLE PHYSICS | India and the United States have agreed to help each other build large experiments to study particles called neutrinos. India will contribute $10 million in hardware to the Long-Baseline Neutrino Facility (LBNF) and the Deep Underground Neutrino Experiment (DUNE), which together will fire neutrinos 1300 kilometers from Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, to a detector housed in a former mine in South Dakota. The United States will contribute equally to the proposed India-based Neutrino Observatory (INO) near Theni, which would study neutrinos created as cosmic rays strike the atmosphere. The $220 million INO may be in jeopardy, however, as a tribunal recently ordered developers to obtain new environmental permits. But India's participation in LBNF/DUNE isn't contingent on the INO's success, says Fermilab Director Nigel Lockyer, who hopes India will eventually pitch in more for LBNF/DUNE, which will cost more than $2 billion. Scientists from 31 countries are working on the project, which broke ground in July 2017.

PHOTO: TIMOTHY ALLEN/GETTY IMAGES

Sea nomads are adapted to their underwater lifestyle.

PHYSIOLOGY

The sea nomad's evolutionary advantage

The Bajau people are known as sea nomads because they have for centuries lived on houseboats on the ocean between Malaysia, the Philippines, and Indonesia, spending 60% of their work hours underwater spearing fish and gathering sea cucumbers and other foods. Now, a research team has found that the Bajau may have evolved adaptations for those oxygen-deprived conditions. The Bajau carry a gene variant that seems to lead to unusually large spleens that can supply an extra boost of oxygenated red blood cells as needed, Melissa Ilardo, an evolutionary genomicist at the University of Utah in Salt Lake City and colleagues report this week in *Cell*. Spleens of Bajau people are 50% larger than those of farmers who live in the same region, the group showed. And by comparing the DNA of the Bajau, those farmers, and the Han Chinese, the researchers found 25 genetic changes in the sea nomads that show signs of natural selection, the evolutionary process by which adaptations become fixed. One notable change is in a gene called *PDE10A,* which can affect thyroid activity, which in turn influences spleen size. The study complements others showing that humans living at high altitudes have also evolved changes to thrive in the thin air.

Deal cuts ship emissions

CLIMATE CHANGE | Members of the United Nations's International Maritime Organization in London last week agreed to cut greenhouse gas emissions from international shipping to 50% of 2008 levels by 2050. International shipping, which contributes about 3% of global carbon dioxide emissions, had been the

Women wait in a Ugandan hospital as their malaria-infected children receive blood transfusions.

PUBLIC HEALTH

Malaria parasites are common in African blood donations

In some regions of Africa, as many as three quarters of blood donors carry malaria parasites in their blood despite having no symptoms of the disease, putting people who receive blood transfusions at risk of infection. At the Multilateral Initiative on Malaria Pan African Malaria Conference in Dakar this week, researchers from the Worldwide Antimalarial Resistance Network reported an analysis of two dozen studies of blood donors from across sub-Saharan Africa indicating that between 6.5% and 74% carried malaria parasites, depending on the region. A second study analyzed 200 blood samples from the Malabo blood bank in Equatorial Guinea using sensitive polymerase chain reaction techniques. It found that 29.5% were contaminated with malaria parasites. The clinical risk to people who receive transfusions from infected donors isn't clear; an earlier study found that only 2% of patients who received blood that tested positive for parasites were infected by the transfusion. But for those with weak natural immunity to malaria—especially young children and pregnant women who often receive transfusions to treat anemia—the risk is likely higher. The scientists emphasize that transfusion recipients in malaria risk areas should receive prophylactic malaria treatment. And they say new techniques for screening the blood supply are sorely needed.

only sector of the modern economy not covered by plans to accomplish the aims of the Paris climate agreement. (Aviation struck a similar accord 2 years ago.) The deal, an "initial strategy," does not lay out measures to make such cuts, which are expected to be difficult for large container and tanker ships to achieve because most are powered by cheap but energy-dense heavy fuel oil. The agreement was finalized over the objections of several countries, including Russia, Saudi Arabia, and the United States.

U.K. university strike suspended

WORKPLACE | A strike over pensions by scientists and other scholars at 65 universities across the United Kingdom has been called off. Citing a deficit of £12.7 billion, the universities had proposed changing their pension scheme from a guaranteed benefit to one based on investment performance. University and College Union members disagreed and in late February began 14 days of strikes. Last week, the union voted 64% to 36% to accept a proposal to establish a joint committee of experts that will review the valuation of the fund and propose ways to maintain a guaranteed pension.

New head for Geological Survey

SCIENCE POLICY | The U.S. Senate confirmed former astronaut James Reilly to lead the U.S. Geological Survey on 9 April by voice vote. With a background in petroleum geology, Reilly is one of a handful of scientists who have been nominated by President Donald Trump for administrative posts and won confirmation. Reilly has pledged to protect scientific integrity at the agency, where the administration has sought to scale back work focused on climate change.

NSF awardee breaks barriers

SCIENTIFIC PRIZES | Kristina Olson, a social and developmental psychologist at the University of Washington in Seattle, last week became the first psychologist to win the National Science Foundation's (NSF's) most prestigious prize for young scientists. Best known for her TransYouth Project, the first-ever longitudinal study of a large group of transgender children, Olson is also the first woman since 2004 to receive the $1 million Alan T. Waterman Award, named for NSF's first director. At 36—and 10 years after her Ph.D.—she qualified under NSF's new eligibility requirements, which raised the age ceiling from 35 to 40 and added 3 years to the 7-year limit for time after degree. The change is intended to level the playing field for researchers whose careers may have been slower to take off because of family obligations, financial pressures, or physical challenges, and this was the first year it was in effect. For good measure, Olson may also be the first Waterman winner to use a good chunk of the award to create a summer research program for undergraduates to foster greater diversity in the scientific workforce. "My goal is to use the money to move us in new directions, because things aren't going to just change on their own," Olson says. "And unless we make room in science for all of the best people, it will be hard to make progress on any of society's problems."

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

PHOTO: YANNICK TYLLE/GETTY IAMGES

IN DEPTH

PLANETARY SCIENCE

NASA lander to probe interior of Mars

By listening for marsquakes, InSight will measure Red Planet's core, mantle, and crust

By **Paul Voosen**

NASA has sent orbiters to study the atmosphere of Mars and rovers to study its surface. Now, the agency plans to look inside the planet. The $814 million InSight lander, due to launch next month, carries three instruments designed to peer through Mars's rusty shell, including a seismometer that will detect "marsquakes." "We've got a black hole that starts 5 meters below the surface and goes all the way down to the center," says Bruce Banerdt, InSight's principal investigator and a geophysicist at the Jet Propulsion Laboratory (JPL) in Pasadena, California. He and his colleagues hope that by measuring the thickness and composition of the planet's crust, mantle, and core, InSight will provide clues to how Mars lost its magnetic field and whether it once hosted plate tectonics.

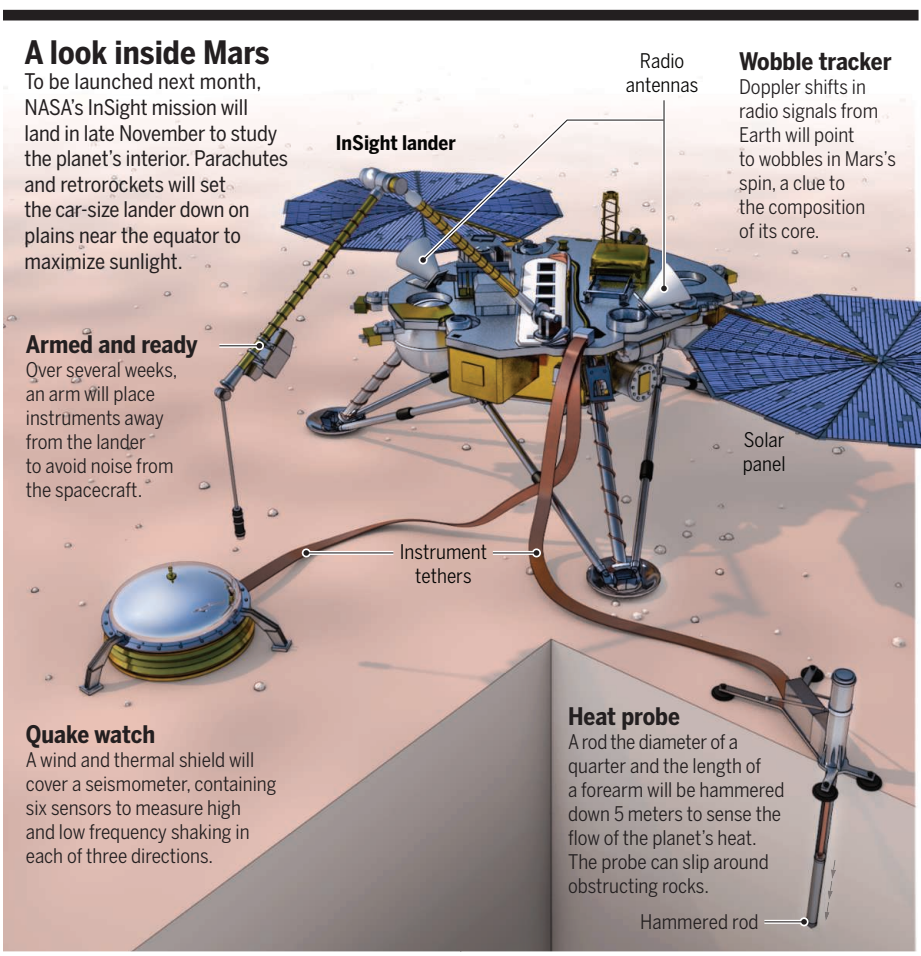
The mission came close to cancellation after a leak was found in the seismometer's vacuum. Ultimately, the launch was delayed for 2 years to find a fix. "We are a much better mission compared to the one we had 2 years ago," says Philippe Lognonné, a planetary seismologist at Paris Diderot University who leads the seismology instrument.

InSight marks NASA's return to planetary seismology after 4 decades. Apollo astronauts deployed five seismometers that detected moonquakes—tremors that helped identify the moon's core. The two Viking landers on Mars both carried seismometers, though one failed and the other sent no reliable signals. Ambitious efforts since then to put multiple seismic stations on Mars have sputtered. But the geophysicists kept pushing, Banerdt says. "I've been getting up at meetings and berating people for not getting behind this for decades," he says. "Sometimes I think they selected my mission just to shut me up."

Developed by JPL with Lockheed Martin and European partners, InSight is built on the same platform as 2008's Phoenix lander. Like its predecessor, it will use parachutes and retrorockets to reach the surface. The target landing site is a smooth plain of lava near the equator—perhaps "the most geologically boring site on the planet," Banerdt says. That's for a reason: InSight could probably do its job from anywhere on Mars, so the team picked a site with few landing hazards and, thanks to its tropical location, plenty of sunlight for the probe's solar panels. After the craft touches down in late November, its robotic arm will deploy the volleyball-size seismometer and a heat probe, driving a rod 5 meters into the surface with thousands of strokes of a tungsten hammer.

The heat probe will measure how much heat is escaping from the planet, and how quickly—a clue to its history. From chemical analysis of the chunks of Mars that arrive on Earth as meteorites, researchers have a sense of the martian mantle's composition. Combining this with the heat gradient and interior dimensions divined by InSight, they can estimate how much of Mars's internal heat comes from radioactive elements in the planet's interior. The remainder is primordial energy left-over from Mars's formation. Based on the rates at which those two heat sources ebb, researchers can estimate when volcanoes were most vigorous on Mars. "The evolution of a planet is driven completely by how heat moves out to space," says Steven Hauck, a planetary scientist at Case Western Reserve University in Cleveland, Ohio.

Tiny Doppler shifts in radio broadcasts sent from Earth to receivers on InSight will reveal other details of the martian interior. The signals will track how the planet wob-



GRAPHIC: C. BICKEL AND A. CUADRA/*SCIENCE*

bles in its rotation, which reflects the internal tug of its core and mantle. Just as raw eggs, with their liquid interiors, spin differently from cooked ones, Mars's wobbles should hint at the core's size, density, and whether it is partially molten, says Véronique Dehant, a geophysicist at the Royal Observatory of Belgium in Brussels. That information could, in turn, shed light on its composition and whether it is crystallizing from the inside out as it cools, like Earth's core, or from the outside in. Ultimately, the results will improve models of how the planet lost the magnetic field that its core once generated, says George Helffrich, a geophysicist at the Tokyo Institute of Technology's Earth-Life Science Institute. "Mars's core could represent what Earth's will look like in the future."

A third set of clues will come from marsquakes. Because Mars lacks the tectonic plates that grind together on Earth, its tremors are likely to be weaker and less frequent than earthquakes. "We might see five to 10" over InSight's 2-year mission, Banerdt says. "Or we might see 1000." In his dreams, the lander will see dozens of marsquakes with a magnitude more than five—a bounty that could help InSight identify their sources, even though it will be a single seismic station.

Typically, three stations are needed to triangulate an earthquake's source from so-called body waves, which dive through the planet. But the InSight team has devised a workaround for Mars by relying on waves that vibrate along the surface. On Earth, features such as the oceans quickly dampen such waves. But on Mars, surface waves from big quakes should race around the planet multiple times. By detecting surface waves from three different global paths, the researchers hope to pinpoint each tremor's source, which will enable them to make sense of how the body waves it produces change speed or reflect off structures in the interior.

If all goes as hoped, the resulting seismic x-rays of Mars will reveal the dimensions of its crust, mantle, and core and any layering within them. A thick crust would mean that Mars melted thoroughly at its start, allowing larger amounts of less dense minerals to rise and collect at the surface. Banerdt says a thick crust, resistant to fracturing into plates, would also suggest that Mars never had plate tectonics.

If InSight survives until 2021, a second seismic station could join it and corroborate its readings: an instrument mounted on the landing platform for Europe's ExoMars rover. But even by itself, InSight is a boon, says Yosio Nakamura, a planetary seismologist at the University of Texas in Austin, who started his career working on Apollo. "A single station is much, much better than no station at all." ■

SCIENCE DIPLOMACY

Department of State's air pollution sensors go global

Fine-particle sentinels prod governments to curb emissions

By **Eli Kintisch**

In October 2010, as heavy smog hung over Beijing, the U.S. embassy's Twitter feed said its rooftop pollution sensor had detected "crazy bad" levels of hazardous microparticles. So-called PM2.5 had shot up to about 550 micrograms per cubic meter—a level to which programmers had given the sardonic label because they thought it would never be reached. The undiplomatic language ruffled feathers in Beijing; embassy staff apologized, deleted the tweet, and replaced the label with "beyond index."

Yet that incident and others like it spurred public complaints that eventually elicited more aggressive efforts to tame the pollution. By now, rooftop sensors like those that drew attention to Beijing's pollution sprout from 26 diplomatic posts in 16 countries. Their immediate goal is to protect the health of U.S. diplomats. But they are raising concerns about air pollution from Sarajevo to New Delhi and supplying data to research efforts. The "little-air-monitor-that-could," as physicist and former U.S. diplomat David Roberts calls it, has become a worldwide watchdog.

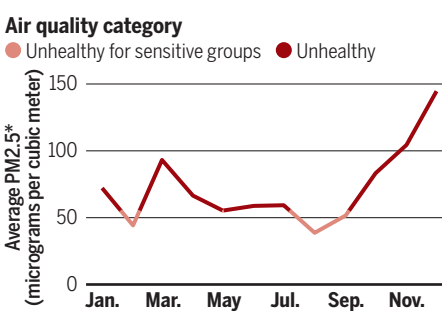
The first Beijing sensor, installed a decade ago, was meant to provide warnings of bad-air days to U.S. citizens. The fine particles it measures, mostly residues of coal burning and vehicle emissions, are linked to respiratory and heart disease. Third party apps made the readings widely available to Chinese who wanted to monitor PM2.5—and got little information from their own government.

China had been struggling to rein in air pollution ever since the late 1990s, after Beijing won the bid for the 2008 Olympics. But officials were embarrassed by the slow progress that the sensor readings highlighted, especially when the U.S. data undercut the Beijing government's rosy pronouncements of "blue sky" days, when pollution supposedly fell. Officials demanded that the embassy stop releasing the data. "We said we couldn't, since the data regarded the health of U.S. citizens," recalls Gary Locke, who served as ambassador to China from 2011 to 2014.

Beijing officials also challenged the usefulness of the U.S. data, pointing out, for instance, that the embassy sensor was in only one location and therefore could be giving an incomplete picture. In response, the embassy teamed up with U.S. Environmental Protection Agency (EPA) scientists to validate the

Winters of discontent

In 2016, as in previous years, the U.S. embassy's sensor in Beijing tracked a surge in PM2.5 late in the year.



*Data are not fully verified or validated.

CREDITS: (PHOTO) KEVIN FRAYER/STRINGER/GETTY IMAGES; (GRAPHIC) J. YOU/*SCIENCE*; (DATA) U.S. DEPARTMENT OF STATE

The U.S. embassy's hourly smog updates have irked Beijing officials—and spurred mitigation measures.

PHOTO: ROBERT S. MERRILLEES

results. "Since we were being criticized, we wanted to make sure we were doing it correctly," says Erica Thomas, a Department of State official who led air monitoring at the embassy in Beijing from 2010 to 2014.

Tensions came to a head in late 2011, when Beijing's smog got so bad that its airport canceled hundreds of flights. Based on monitoring of larger particles, municipal authorities insisted that the air was only "slightly polluted"—provoking ridicule from bloggers citing U.S. data. "It was the straw that broke the camel's back," says Angel Hsu of Yale University, an expert on pollution in China. Days later, China proposed new air quality standards that included PM2.5 for the first time. It now runs the biggest PM2.5 monitoring system in the world.

The Department of State, too, expanded its PM2.5 monitoring efforts, first within China and then around the world, and is using the sensors for research as well. Drawing on PM2.5 data from five diplomatic posts in China, a 2015 study in *Atmospheric Environment* revealed previously unknown variations in PM2.5 levels; for instance, in Beijing the particles tended to peak around midnight and bottom out in spring, because of weather patterns. And the U.S. embassy in Sarajevo is now testing cheaper sensors with an aim of deploying them "in different pollution environments," says Caroline D'Angelo, who runs the Department of State's monitoring network in Washington, D.C.

Findings are radiating into other disciplines. During a stint at the U.S. consulate in São Paulo, Brazil, Tommy Flynn, a program manager with the South Carolina Department of Health and Environmental Control, is providing technical assistance on the monitors. But he has also learned from the consulate's practice of distributing filter masks to high-risk groups during ultra–high-pollution events. That's a lesson, he says, that could be applied in South Carolina during forest fires. And scientists from NASA and the World Bank are using embassy PM2.5 data, now posted on EPA's AirNOW monitoring website, to groundtruth satellite measurements of pollution, filling in gaps in the global map of reliable air quality measurements.

The U.S. sensors continue to serve as a diplomatic cudgel, as well. In India, where air pollution in New Delhi and other cities has skyrocketed, "the data has flowed from the embassy to the media and then created outcry," says Christa Hasenkopf, a former Department of State official who now runs the air pollution nonprofit OpenAQ in Washington, D.C. As Locke says, the monitoring program "has taken on a life its own." ■

ARCHAEOLOGY

Cannabis, opium use part of ancient Near Eastern cultures

Growing body of data suggests ritual drug use by ancient Mesopotamians, Cypriots, and others

By **Andrew Lawler**, *in Munich, Germany*

For as long as there has been civilization, there have been mind-altering drugs. Alcohol was distilled at least 10,000 years ago in the Fertile Crescent, about the same time that agriculture took hold there. Elsewhere, for example in Mesoamerica, other psychoactive drugs were an important part of culture. But the ancient Near East had seemed curiously drug-free—until recently.

Now, new techniques for analyzing residues in excavated jars and identifying tiny amounts of plant material suggest that ancient Near Easterners indulged in a range of psychoactive substances. Recent advances in identifying traces of organic fats, waxes, and resins invisible to the eye have allowed scientists to pinpoint the presence of various substances with a degree of accuracy unthinkable a decade or two ago.

Cypriot jugs were crafted in the shape of the poppy seed pod 3000 years ago.

For example, "hard scientific evidence" shows that ancient people extracted opium from poppies, says David Collard, senior archaeologist at Jacobs, an engineering firm in Melbourne, Australia, who found signs of ritual opium use on Cyprus dating back more than 3000 years. By then, drugs like cannabis had arrived in Mesopotamia, while people from Turkey to Egypt experimented with local substances such as blue water lily.

Some senior researchers are still dubious, pointing out that ancient texts are mostly silent on such substances. Others consider the topic "unworthy of scholarly attention," Collard says. "The archaeology of the ancient Near East is traditionally conservative."

But the work is prompting fresh thinking on the relationship between substances and societies. At the International Congress on the Archaeology of the Ancient Near East here last week, for example, one scholar even reinterpreted well-studied ancient images as representing drug-taking rituals and drug-induced distortions.

Drug use almost certainly began in prehistory and spread with migrations. For example, the Yamnaya people, who swept out of Central Asia about 5000 years ago and left their genes in most living Europeans and South Asians (see p. 252), appear to have carried cannabis to Europe and the Middle East. In 2016, a team from the German Archaeological Institute and the Free University, both in Berlin, found residues and botanical remains of the plant, which originates in East and Central Asia, at Yamnaya sites across Eurasia. It's difficult to know whether the Yamnaya used cannabis simply to make hemp for rope or also smoked or ingested it. But some ancient people did inhale: Digs in the Caucasus have uncovered braziers containing seeds and charred remains of cannabis dating to about 3000 B.C.E.

Once people organized into city states, they may also have started large-scale production of pharmaceuticals, says archaeologist Luca Peyronel of the International University of Languages and Media in Milan, Italy. A decade ago, before the onset of Syria's brutal civil war, he was part of a team that gathered samples from an unusual kitchen in a palace in the northwestern Syrian city of Ebla, which flourished 4 millennia ago on the outskirts of the Sumerian and Akkadian empires.

The room lacked the plant and animal remains typically associated with food preparation. But residue analyses on pots found there may explain the mystery, as Peyronel and his colleagues described in a paper last year: The researchers found traces of wild plants often used for medicine, such as poppy for opium to dull pain, heliotrope to fight viral infections, and chamomile to reduce inflammation. Given

that the space contained eight hearths and pots that could hold 40 to 70 liters, the drugs could have been made in large quantities, Peyronel says.

Some of these extracts, such as opium, can induce hallucinations, although it's unclear whether the potions were used in ritual or medicine. The kitchen's location near the heart of the palace suggests its products were used for ceremonial occasions, and cuneiform tablets from the building mention special priests associated with ritual beverages, Peyronel says. The distinction between medicine and mind-altering drug may have been lost on ancient peoples. "The two hypotheses are not necessarily at odds," he adds.

Three hundred kilometers due west and several centuries later, the ancient people of Cyprus used opium in religious ceremonies, Collard says. Residue analyses show that between 1600 and 1000 B.C.E., people poured opium alkaloids into pots crafted in the shape of the seed capsule of the opium poppy, in what Collard calls "prehistoric commodity branding." All the jugs were found in temples and tombs, suggesting a role in ritual. Opium jugs made on Cyprus have been found in Egypt and the Levant—the first clear example of the international drug trade.

Other substances less well known today may have played a role in healing or ecstatic rituals in the ancient Near East. When King Tutankhamun's tomb, dating to the 14th century B.C.E., was opened in 1922, archaeologists found the boy-king's body covered with the flowers of blue water lily, a common motif in many Egyptian tomb paintings. Steeped in wine for several weeks, the plant yields a sedative that produces a calm euphoria.

Diana Stein, an archaeologist at Birkbeck University of London, claims archaeologists have long studied scenes of rituals involving drugs and their effects without realizing it. She argues that the banquet scenes that often adorn small seals found Anatolia, Syria, Mesopotamia, and Iran actually show people imbibing psychoactive potions. Another common motif, interpreted as a scene of contest, may instead represent the internal conflict that results when the imbiber faces an alternative reality, Stein proposes. In these images, "everything is distorted and pulsing—but they certainly knew how to carve things realistically when they wanted to," she said at the meeting here.

"I find Diana's arguments convincing and even energizing, as they open up a new avenue for research," says Megan Cifarelli, an art historian at Manhattanville College in Purchase, New York.

But others are more cautious. "Scholars have tended to shy away from the possibility that the ancient Near Easterners partook of 'recreational' drugs, apart from alcohol, so it's good that someone is brave enough to look into it," says archaeologist Glenn Schwartz at Johns Hopkins University in Baltimore, Maryland. But he says Stein's suggestions "seem to go too far on too little evidence," a view echoed by many at the meeting.

Collard, however, is confident that additional residue and botanical analyses, along with study of iconography and texts, will gradually persuade skeptics. Cifarelli notes that the ancients likely used drugs not just to heal, but to forge sets of beliefs, and contact a spiritual realm where healing and religion were entwined. "Most of us," she says, "are so far removed from that kind of transformative magic." ■

Poppies, shown here with seed pods, have been used to produce opium in the Near East for some 5000 years.

PHOTO: ISTOCK.COM/OZTURK

DEMOGRAPHY

Plan for 2020 U.S. census is fatally flawed, critics say

Using existing records to ensure accuracy is unlikely to succeed, experts predict

By **Jeffrey Mervis**

A plan by Commerce Secretary Wilbur Ross to ensure a "complete and accurate" 2020 U.S. census is seriously flawed, former Census Bureau officials and other experts in survey research charge.

For critics, the problem is not simply Ross's controversial, eleventh-hour decision last month to add a citizenship question to the decennial census, although civil rights groups, local and state governments, and business leaders believe the question will depress participation and jeopardize an accurate head count in 2020. It is also Ross's plan to "maximize" the Census Bureau's ability to carry out the constitutionally mandated count by using so-called administrative records, the massive amount of information that other government agencies already possess on U.S. residents. But social scientists predict harnessing administrative records will be a mission impossible, at least in the 2 years left before Census Day.

"It would be incredibly challenging" to carry out Ross's strategy, says Amy O'Hara, a senior research scholar at Stanford University's Institute for Economic Policy Research in Palo Alto, California, who until last fall managed the agency's efforts to use administrative records. And the barriers aren't likely to be overcome before the census begins on 1 April 2020, adds Hermann Habermann, a former deputy census director under President George W. Bush who's now an occasional consultant to the Committee on National Statistics at the National Academies of Sciences, Engineering, and Medicine in Washington, D.C. "The process sounds easy to a layperson," he says. "But it's not."

On 26 March, Ross released an eight-page memo outlining the plan. It responded to a December 2017 request from the Department of Justice (DOJ) to add a question to the cen-

New U.S. citizens celebrate at a naturalization ceremony last month. Asking about citizenship on the 2020 census has become a controversial topic.

sus that would provide block-by-block information on citizenship. DOJ said the data are essential to prevent voter discrimination under the 1965 Voting Rights Act.

Critics disagree, noting that previous administrations never needed such details to enforce the law. Instead, they believe the real purpose of the new question is to scare immigrant communities that traditionally have voted for Democrats away from participating in the census. If that happens, the resulting undercount could tilt the process of apportioning the 435 congressional seats in favor of Republicans. Some 17 states and seven cities have sued Ross and DOJ, arguing the move would "fatally undermine the accuracy" of an exercise that is also used to distribute some three-quarters of a trillion dollars annually in federal aid.

In his memo, Ross states that he held "extensive" conversations with outside experts before making his decision, who told him there is no evidence that a citizenship question would "materially" reduce response rates. That's correct as far as it goes, says Habermann, one of the three anonymous experts whom Ross cites. (*Science* was able to identify the experts because Ross used previous or current job titles in his memo.) But Habermann says Ross is ignoring something even more important.

"When someone wants to put a question on the census, there's a high burden of proof that must be met about its value," he says. "I told him that I don't think the case has been made that [the question] is so important that it's worth endangering this fragile instrument."

A second expert, Robert Groves, who led the Census Bureau during the 2010 census and who is now provost of Georgetown University in Washington, D.C., has voiced a similar concern. In January, he was one of five former census directors who urged Ross to reject DOJ's request. "It is highly risky to ask untested questions," Groves and the others wrote in a 26 January letter. "There is a great deal of evidence that even small changes in survey question order, wording, and instructions can have significant, and often unexpected, consequences for the rate, quality, and truthfulness of response."

The bureau typically conducts exhaustive testing before adding any question to the decennial census, to ensure that the data are of the highest quality; such testing has never been applied to a question about citizenship. (The agency's American Community Survey, which goes to some 3.5 million households annually, asks three citizenship questions that have been vetted.)

> ***"When someone wants to put a question on the census, there's a high burden of proof that must be met ..."***
> **Hermann Habermann**, former census official

Ross's memo doesn't address the problems associated with fielding an untested question. But he says administrative records will somehow allow the bureau to maintain quality despite a lower response rate, which he concedes may occur.

The reality, experts say, is that the bureau has yet to test the potential impact of using more administrative records. In general, experts say they can help improve the accuracy of a count. For example, the Census Bureau can use postal or tax records to identify households not already in its master address file. Those addresses would then be targeted for follow-up mailings and, as a last resort, repeated knocks on the door. Administrative records can also be used to check whether the answers to a census survey are accurate, or to fill in a blank if respondents have skipped a question.

But administrative records are not a gold standard. They may contain errors, or data at odds with what the Census Bureau has collected through its surveys.

Ross's memo is intended to jump-start the bureau's use of administrative records. With respect to citizenship, O'Hara says, the first step might be to track down citizenship records on individual U.S. residents at another federal agency or a commercial firm. Then the agency would have to negotiate a data-sharing agreement. Such negotiations can take many years, she notes.

The next hurdle is to convert the data into a format that the Census Bureau can handle. O'Hara says, "I can't estimate how difficult [harmonization] would be because these data have never been drawn together."

Moreover, citizenship data from another agency may not be up-to-date. The Social Security Administration, for example, asks people about citizenship when they apply for a number. (A noncitizen may be on a visa that allows them to work legally.) But although residents typically notify the agency about a change in marital status, for example, they are not required to inform the agency of a change in citizenship status.

Another problem is what to do when the records contain conflicting information about a person. Race is a good example, O'Hara says. "People respond differently to a question about race and ethnicity depending on how it's asked," she notes. "And I think that interpreting answers to a question about citizenship is likely to be even more complicated."

Multiple addresses in a file can also confound the search for an individual, even if that person's name and date of birth are known. For the census, O'Hara explains, "you need to be able to identify a person, at a particular place, with certain attributes."

Congress will have several chances to explore the agency's plans for the 2020 census. Last week, census officials discussed the new question at a private meeting with a House of Representatives panel that oversees the census, and on 8 May they will reprise the conversation in public. Democrats on the corresponding Senate panel have also asked for a chance to grill Ross. But unless Congress or the courts intervene, the impact of using administrative records may not be known until after the 2020 census is over. ■

PHOTO: MARIO TAMA/STAFF/GETTY IMAGES

HUMAN EVOLUTION

Ancient DNA untangles South Asian roots

Study shores up steppe as long-sought Proto-Indo-European homeland

By **Lizzie Wade**, *in Austin*

Today, the population of South Asia is divided into dozens of ethnic, linguistic, and religious groups that live side by side—but not always in harmony. A contentious border separates India and Pakistan; political movements draw stark lines between India's Muslim and Hindu populations. Groups don't mix much, as people tend to marry those who share their ethnicity and tongue.

Now, a study of the first ancient DNA recovered from South Asia shows that populations there mingled repeatedly thousands of years ago. Nearly all of the Indian subcontinent's ethnic and linguistic groups are the product of three ancient Eurasian populations who met and mixed: local hunter-gatherers, Middle Eastern farmers, and Central Asian herders. Three similar groups also mingled in ancient Europe, giving the two subcontinents surprisingly parallel histories.

The study, presented here last week at the meeting of the American Association of Physical Anthropologists and in a preprint on the bioRxiv server, sheds light on where these populations came from and when they arrived in South Asia. It also strengthens the claim that Proto-Indo-European (PIE)—the ancestral language that gave rise to modern languages from English to Russian to Hindi—originated on the steppes of Asia.

"It's first-rate work," says Partha Majumder, a geneticist at the National Institute of Biomedical Genomics in Kalyani, India. He found hints of similar genetic patterns in his previous studies, but the addition of ancient DNA makes the new conclusions stronger, he says. "It's absolutely stunning."

Priya Moorjani, a geneticist at the University of California, Berkeley, studies how South Asian populations relate to each other and to others around the world. In previous work, she analyzed the genomes of nearly 600 modern Indians and Pakistanis from 73 ethnolinguistic groups in South Asia. Her team found that almost all people living in India today carry ancestry from two ancient populations: Ancestral North Indians, who were more related to people from Central Asia, the Middle East, the Caucasus, and Europe; and Ancestral South Indians, who were more related to indigenous groups living in the subcontinent today. But without DNA from ancient people, Moorjani couldn't be sure who gave rise to those ancestral populations, or when.

Moorjani, David Reich of Harvard University, and Kumarasamy Thangaraj of the Centre for Cellular and Molecular Biology in Hyderabad, India, spent years searching for ancient DNA in South Asia, where hot climates might degrade it. Finally, their team recovered and analyzed ancient genomes from 65 individuals who lived in northern Pakistan between 1200 B.C.E. and 1 C.E. They also analyzed 132 ancient genomes from Iran and southern Central Asia, and 165 from the steppes of Kazakhstan and Russia, and compared them with published ancient and modern genomes. These data allowed them to reconstruct when different populations arrived in South Asia and how they interacted.

Three ancestral groups gave rise to the diverse people of South Asia today.

PHOTO: MARY DOGGETT/ALAMY STOCK PHOTO

Between 4700 and 3000 B.C.E., farmers from Iran mixed with hunter-gatherers indigenous to South Asia, Moorjani said. This combination of ancestries was found in the DNA of skeletal remains from sites in Turkmenistan and Iran known to have been in contact with the Indus Valley civilization, which thrived in Pakistan and northwest India starting around 3300 B.C.E. (*Science*, 6 June 2008, p. 1276). The researchers dub this population "Indus periphery." The 65 ancient people from Pakistan also show this combination, although they all lived after the Indus civilization declined. The researchers suspect that "Indus periphery" people actually may have been the founders of Indus society, although without ancient DNA from Indus Valley burials, they can't be sure.

Still, Moorjani's team sees this ancient mixture of Iranian farmers and South Asian hunter-gatherers all over South Asia today. As the Indus Valley civilization declined after 1300 B.C.E., some Indus periphery individuals moved south to mix with indigenous populations there, forming the Ancestral South Indian population, which today is more prominent in people who speak Dravidian languages such as Tamil and Kannada, and in those belonging to lower castes.

Meanwhile, herders from the Eurasian steppe moved into the northern part of the subcontinent and mixed with Indus periphery people still there, forming the Ancestral North Indian population. Today, people who belong to higher castes and those who speak Indo-European languages such as Hindi and Urdu tend to have more of this ancestry. Shortly after, these two already mixed groups mixed with each other, giving rise to the populations living in India today.

"Strikingly, this is very similar to the pattern we see in Europe," Moorjani said. Around 7000 B.C.E., agriculture spread into both Europe and South Asia with farmers from Anatolia and Iran, respectively, who each mixed with local hunter-gatherer populations. After about 3000 B.C.E., Yamnaya pastoralists from the Central Asian steppe swept both east and west, into Europe and South Asia, bringing the wheel and perhaps cannabis (see p. 249).

Earlier genetic work had linked the arrival of these herders to the spread of Indo-European languages in Europe. But other researchers, including archaeologist Colin Renfrew of the University of Cambridge in the United Kingdom, had argued that the earlier Anatolian farmers were the original PIE speakers. The new data "make a strong case" for the Yamnaya as carriers of Indo-European languages, Renfrew says. But he still thinks Anatolian farmers could have spoken the earliest language in that family. ■

Winning an academic position and stable funding can be an uphill climb for young scientists.

U.S. SCIENTIFIC WORKFORCE

Proposal to rescue postdocs from limbo draws darts

Expert panel offers mix of old and new ideas to help young biomedical researchers build stable careers

By **Michael Price**

An expert panel last week offered some new solutions to an old problem in U.S. biomedical research: too many young scientists vying for scarce academic research jobs and stable funding. Although many of its recommendations echo past proposals that have largely failed to gain traction, a few—including one to limit postdoctoral researchers to receiving just 3 years of funding from their principal investigators (PIs), and another to create a new federal oversight council—are catalyzing debate.

Two statistics highlight the "fissures and areas of stress" facing young biomedical researchers trained in the United States, suggests the panel assembled by the National Academies of Sciences, Engineering, and Medicine (NASEM) in Washington, D.C. In 1973, more than half of those who earned a biomedical science Ph.D. were employed in a tenured or tenure-track position within 6 to 10 years of completing their degree. Today, that number is less than one-fifth. And the average age at which a researcher wins their first major grant from the National Institutes of Health (NIH) has risen from 36 in 1980 to 43 in 2016 (see graphic, right). Such trends have helped create "a research career path that is increasingly unattractive," the report notes, and causes many hopeful scientists to languish in relatively poorly paid postdoc positions with little chance of reaching the tenure track.

To fix the problems, the panel offers more than a dozen major recommendations aimed at Congress, research funders, and academic institutions. Most radically, it asks NIH to develop policies that would limit a postdoc to receiving just 3 years of funding from their PI's research grants. The problem with current practices, the panel says, is that they encourage some PIs to view postdocs primarily as labor and provide little career mentoring.

To prevent such limbo, the panel recommends a three-pronged approach. The 3-year funding cap would spur postdocs to demonstrate whether they are able to win funding for independent research. The second prong would quintuple within 5 years the number of certain kinds of grants that allow young scientists to develop greater autonomy, increasing their chance of success. A third prong would create a greater number of better-paid, nontenured staff scientist positions. That could allow PIs to offer stable jobs to talented researchers who don't find tenure-track jobs or independent funding.

The committee also recommends raising NIH's fellowship salaries—on which many universities base their own postdoc pay—which would discourage PIs from seeing postdocs as a source of cheap labor, and suggests universities regularly share data on career outcomes. And it calls on federal lawmakers to charter a new Biomedical Research Enterprise Council (BREC) that would include representatives from an array of stakeholders. The panel says the BREC could serve as a forum for figuring out how to implement the report's recommendations, which it estimates could cost $1 billion over 5 years.

Although it is not clear whether Congress would give such a body enough authority to be effective, creating the BREC could be "a good step forward as a policy process," says Hal Salzman, a science and engineering workforce researcher at Rutgers University in New Brunswick, New Jersey. But Gregory Petsko, a neuroscientist at Weill Cornell Medicine in New York City and chair of a similar 2014 NASEM report, is skeptical that it would carry much weight. "I'd love to be wrong," he says. "I'm not opposed to it as an experiment."

The proposed 3-year cap has drawn fiercer opposition. Michael Baym, a bioinformatics researcher at Harvard Medical School in Boston, says such a cap would have prematurely ended his career. "This would have effectively kicked me out of a postdoc" before he established himself, he tweeted.

Mark Peifer, a cell biologist at the University of North Carolina in Chapel Hill, says that if the rule were in place now, he would have to let go of two of his promising postdocs, who he believes will ultimately win academic jobs. And he resents the implication that PIs aren't doing enough to help postdocs prepare for a variety of careers. "That may be the case at large labs with lots of postdocs," he says. "But at most small and medium-size labs, we're already mentoring [and] tracking career outcomes." Still, Peifer says a funding cap is needed, but believes 5 or 6 years is more realistic.

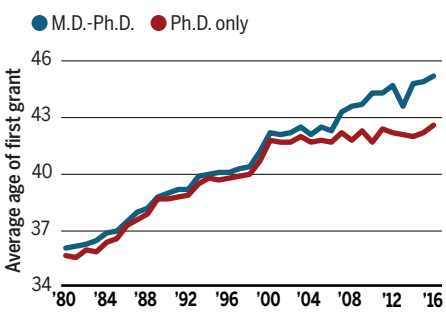
Others worry an arbitrary time limit on funding will hurt postdocs involved in long-term studies, or those who need to take time off for medical or family leave.

Postdocs shouldn't worry about a rapid change, says biologist Jessica Polka, a member of the NASEM panel and director of ASAPBio, based in California. She notes that the report recommends that the cap be implemented "only after one [or] more pilot studies, so that deleterious consequences ... can be understood and mitigated." ■

Michael Price is a freelance journalist in San Diego, California.

Long wait

The average age at which a biomedical researcher wins their first Research Project Grant, or an equivalent, from the National Institutes of Health is rising.



CREDITS: (GRAPHIC) D. MALAKOFF/*SCIENCE*; (DATA) NATIONAL INSTITUTES OF HEALTH

OMEN IN THE BLOOD

A protein marker predicts health crises—but can it cause them?

By **Stephen S. Hall**

Nick Henry first experienced the symptoms of kidney disease in 2004, shortly after the 19-year-old had a severe reaction to a spider bite. "I woke up one morning, and I was just swollen from head to toe," he recalls. But doctors managed Henry's disease, allowing him to return to his unusually active lifestyle—including baseball, softball, basketball, flag football, golf, and fishing—in his northeast Louisiana hometown of West Monroe. Shortly after he witnessed the death of his mother in a motor scooter accident in 2012, however, Henry's renal health took a dramatic turn for the worse. "It's almost as if my body went into shock," he says. "Within a couple months, boom, I started swelling up again."

That swelling was a sign that his kidneys were no longer working normally. A biopsy confirmed that he had focal segmental glomerulosclerosis (FSGS), a severe form of kidney disease. In FSGS, the kidney's glomeruli—the microscopic filtration units that sieve excess fluid and waste products from the blood—become overly leaky; essential proteins such as albumin seep out, disrupting blood chemistry and causing fluid to leak from the blood vessels into tissues throughout the body. Henry's condition deteriorated so rapidly that by July 2014, his doctors in Shreveport, Louisiana, decided to remove both diseased kidneys. The next month, Henry received a transplanted kidney from his identical twin, Nate, who was healthy, even though FSGS can be genetic in origin.

Within a day of the transplant, however, Henry felt like the swelling was coming back. At first, his doctors reassured him that

Nick Henry had a kidney transplant, but his new organ quickly deteriorated, and he spends his nights on dialysis.

PHOTO: KEVIN BEASLEY

he was doing fine. "Once they checked my urine, saw me spilling a bunch of protein again," he says, "they realized [FSGS] was attacking the new kidney." Three days after the surgery, Henry's doctors conceded that the newly transplanted kidney had already become diseased. His transplant doctor, Neeraj Singh of Louisiana State University in Shreveport, says the recurrence was "one of the most dramatic cases I've seen."

The sudden failure of Henry's new kidney is a recent chapter in a long-running medical mystery, dating to when kidney transplants became routine in the 1970s. Up to 30% of transplanted kidneys fail in FSGS patients—not because of immune rejection by the body, as doctors first suspected, but because the new organ immediately begins succumbing to the same disease process that ravaged the original ones. As he struggled to cope with that devastating turn of events (and relied on dialysis to stay alive), Henry traveled to Chicago, Illinois, to consult with Jochen Reiser, a kidney disease specialist who is chairperson of internal medicine at Rush University Medical Center there.

Ever since he learned about such transplant failures 2 decades earlier, Reiser has been convinced that "there is something in the blood circulating that attacks the kidney. And we were out to catch that." What he and colleagues claim to have "caught," in an elegant but still unfolding story of molecular detective work over the past 10 years, is a protein known as soluble urokinase plasminogen activator receptor (suPAR). When Reiser analyzed blood samples from the Henry twins, the results aligned with the message he has been preaching with evangelical fervor for years. Nate, the healthy brother, had relatively low levels of suPAR; Nick's were high—a driving force, Reiser believes, of his kidney failure.

Chronic kidney disease affects 14% of the U.S. population, with estimates suggesting nearly 600 million people affected worldwide. The disease steadily erodes the kidneys' ability to filter the blood, often leading to cardiovascular disease and premature death. Kidney disease—which can directly attack the filtration process, as in FSGS, or damage the kidney's support structure—is particularly insidious because by the time the first diagnostic signs appear, patients have irreversibly "burned off" much of their kidney function. Historically, the leading risk factors have been high blood pressure, diabetes, and African-American ancestry. (Several mutations associated with increased risk are more common in African-Americans.)

But research by Reiser and others has dramatically challenged that traditional picture of risk. If suPAR levels are low, people with the high-risk genes are no more likely to develop kidney disease than people without those gene variants, Reiser says. If suPAR levels are high, people are at greater risk of developing the disease regardless of whether they have the mutations.

Molecular studies in animals as well as a growing number of analyses of large human populations associating suPAR with kidney disease have bolstered his confidence—and convinced him the disease could be treated by suppressing suPAR. Some other researchers aren't convinced, noting that several clinical studies found no clear association between suPAR levels and FSGS. But on both sides of the debate there is widespread fascination with suPAR, a ubiquitous, Zelig-like bystander molecule that, at elevated levels in the blood, seems to presage many health calamities, such as heart attacks, diabetes, and premature death. Whatever suPAR's precise role in kidney disease, the molecule appears to be a potent signal broadcast by an immune system under siege. It is exquisitely sensitive to inflammation, an accelerant for many diseases.

"What is inflammation?" asks biochemist Jesper Eugen-Olsen of the University of Copenhagen, a pioneer in suPAR research. "It's the language of cells. It's how cells communicate with each other. When something is going wrong, the immune system is activated. It produces suPAR … and suPAR is a voice that just shouts, 'Get on with it! Something is going on!'"

IN NEITHER BACKGROUND nor appearance does Reiser conform with the public image of the director of a major metropolitan medical center. His 10th floor office at Rush sits just behind a corridor lined with photographs of hospital administrators going back to the 19th century—stern-faced, all-knowing medical patriarchs. Inside, Reiser, 46, sports a stylish striped blue suit, fashionably stubby beard, red socks, and slick dark hair. Known among colleagues as ambitious and scientifically gregarious, he has been eager to collaborate with anyone interested in exploring suPAR biology, and his brash, full-on style extends to the conspicuous display of large-format books celebrating the history of Aston Martins (he owns one) and Porsches on the coffee table in his office. Describing the speed of data collection for a paper that several years ago ended up in *The New England Journal of Medicine* (*NEJM*), he says, "It was like going from zero to 200 in no time," adding sheepishly, "Car analogy."

Nate Henry (right), Nick's identical twin, is healthy. Nick's high levels of a molecule called suPAR may explain his illness.

PHOTO: KEVIN BEASLEY

Born and raised in the small German village of Remchingen, on the eastern edge of the Black Forest, Reiser got his medical degree and Ph.D. from Heidelberg University and did an overseas residency at Albert Einstein College of Medicine in New York City. Specializing in kidney disease, he went on to conduct research at Harvard Medical School in Boston and became chief of nephrology at the University of Miami Leonard M. Miller School of Medicine in Florida before being hired by Rush in 2012.

Reiser's arrival in the United States in 1999 coincided with renewed interest in solving the mystery of why up to 30% of FSGS patients who receive transplants see the disease recur in the new kidney. Just 3 years earlier, a group headed by Flavio Vincenti, a transplant specialist at the University of California, San Francisco (UCSF), and Virginia Savin, at the Medical College of Wisconsin in Milwaukee, announced a major clue. They reported in *NEJM* that they had amassed evidence for an FSGS-promoting factor in the blood of transplant recipients who'd experienced recurrences; they couldn't isolate the exact protein, but when colleagues later injected an extract of such patients' blood into rodents, the animals' kidneys became permeable and spilled protein in the urine. That mysterious "permeability factor" became "the holy grail" of the field, according to Sanja Sever, a molecular biologist who studies kidney disease at Massachusetts General Hospital in Boston.

While still in Germany, Reiser had trained his research efforts on a unique renal cell called the podocyte (so named because of its amoebic, faintly footlike extensions). That choice turned out to be fortunate. The kidney has about 1 million glomeruli, and in each one, hundreds of podocytes bridge the gap between the bloodstream and the urinary system. Their footlike extensions wrap around capillaries snaking through the kidneys and, along with two other layers of tissue, form a physical mesh of cells, like a threeply screen door, that allows only small molecules—sodium ions, potassium ions, and metabolic wastes—to pass into the urinary tract. When the podocytes become damaged, however, they essentially lose their architectural integrity. The kidney filters then become leaky, allowing larger essential proteins such as albumin to escape from the blood and pass into the urine.

It's like a coffee filter, Sever says. "If there are holes in your filter, then you get some coffee grounds in your urine." Podocyte damage can be reversed early in kidney disease. But, she says, "If you keep losing them, there's a point of no return. … You are basically walking toward end-stage renal disease."

What causes such damage? Reiser suspected that the mysterious blood-borne factor disrupts podocytes through receptor molecules on their cell surface. He focused on one: β_3-integrin, a molecule whose activation perturbs the shape and motility of cells. When he looked for the molecular key that turned the lock of the integrin receptor, he discovered that oncologists had already been working on one such protein, urokinase PAR (uPAR), a cell surface receptor that plays a role in cancer metastasis. Reiser became even more intrigued when he learned that uPAR can be cleaved from cell surfaces and circulate in the blood—at which point it becomes a soluble cousin known as suPAR. Maybe suPAR was the mysterious kidney-destroying factor.

> ***"The data gets stronger and stronger that suPAR is the worst toxin you can have for the kidneys."***
>
> **Jochen Reiser**, Rush University Medical Center

In 2011, Reiser and colleagues reported in *Nature Medicine* that in cell culture, suPAR damaged human podocytes through the integrin pathway. The researchers supplemented that evidence with three mouse models showing that rodents with elevated levels of suPAR suffered kidney damage, although sometimes more slowly than in FSGS. With human clinical data suggesting that elevated suPAR levels correlated with the recurrence of FSGS in patients, a picture emerged in which the protein triggers a pathogenic process that ultimately produces holes in the coffee filter, leading to kidney disease.

The findings both electrified and polarized the nephrology community. In a commentary for *Nature Medicine*, Martin Pollak of Harvard Medical School, who studies the genetics of kidney disease, and nephrologist Stuart Shankland of the University of Washington in Seattle described the findings as "paradigm shifting for our understanding of the pathogenesis of FSGS."

But some groups could not find the same clinical association between suPAR levels and recurrent disease in FSGS patients, and other groups questioned the protocol and interpretation of the animal models. And regardless of whether suPAR actually destroys the kidney, many nephrologists thought its levels were not very informative—by the time those specialists saw patients with kidney disease, suPAR levels were already high and offered no prognostic value. With Reiser claiming to have found the "holy grail" even as several groups were reporting discordant results, says one source, "People felt very emotional."

BY THAT POINT, another key strand of the suPAR story had emerged in Europe. There, the focus was on the molecule as a potential biomarker for a range of diseases.

The first clues came from AIDS patients. In Copenhagen, Eugen-Olsen and others examined blood collected from more than 300 HIV patients in the early 1990s, before life-saving antiretroviral therapies became available. All those patients had died, but a retrospective analysis showed their suPAR levels eerily correlated with disease progression: Higher levels were associated with an earlier death. Eugen-Olsen then spent several years collaborating with a hospital in the West African nation of Guinea-Bissau, testing suPAR levels in patients suspected of being HIV-infected. Again, higher suPAR levels predicted a quicker death among the infected. Surprisingly, however, suPAR also predicted mortality in patients who didn't have AIDS; many turned out to have tuberculosis. That finding led him to hypothesize that suPAR might be a more general biomarker for chronic inflammation.

In 2001, Eugen-Olsen founded the company ViroGates, which began to manufacture a relatively inexpensive test to measure suPAR levels in the blood. With the test in hand, he and colleagues in Copenhagen began to look at collections of blood samples banked in large-cohort prospective studies. In one called MONICA, which monitored healthy members of the Danish population for about 13 years, elevated levels of suPAR were associated with a higher risk of cardiovascular disease, type 2 diabetes, cancer, and premature death. Two other large European populations, enrolled in the Malmo Diet and Cancer Study and the Danish Inter99 Study, showed similar associations.

The findings caught the attention of researchers at Emory University School of Medicine in Atlanta who had been looking for new and better biomarkers to predict risk of adverse cardiac events in people with heart disease. The researchers had built the Emory Cardiovascular Biobank with serum from several thousand patients. "We draw blood, and we follow them for years," says Salim Hayek, a physician and research fellow at Emory. When two of Hayek's colleagues, Danny Eapen and Arshed Quyyumi, delved into the biobank, they found that higher suPAR levels predicted heart attacks and death, as they reported in the *Journal*

of the American Heart Association in 2014. (At the annual meeting of the American College of Cardiology last month, Hayek presented further evidence from the Emory group, suggesting that suPAR is a better predictor of cardiac events including heart attacks and death than any other biomarker in widespread clinical use.)

In addition to serving as an omen of ill health, suPAR seems to be a remarkably sensitive indicator of lifestyle insults. Studies have shown that the protein's blood levels typically rise with obesity and with smoking. (Eugen-Olsen, an inveterate smoker, confesses that he quits when his suPAR levels rise and resumes when they subside again.) "Just looking at the data," Hayek says, "clearly the environment is a much larger contributor to suPAR than genetics."

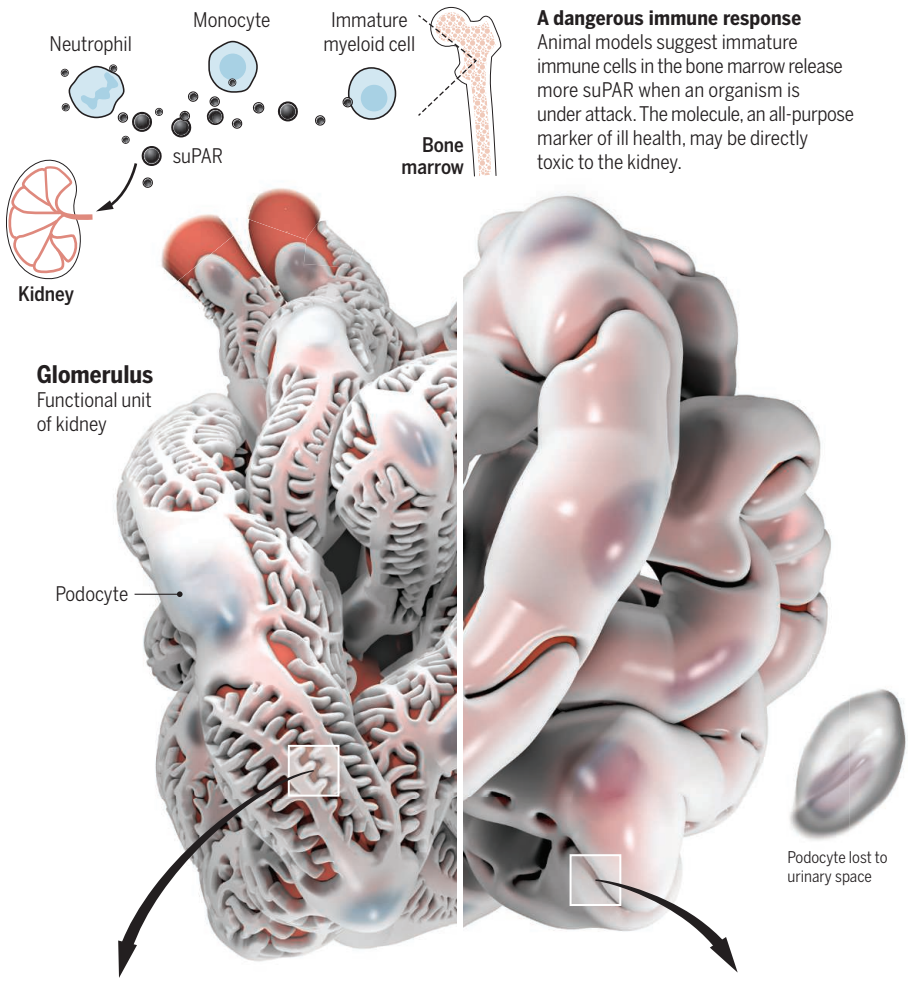
With its links to multiple diseases and environmental stresses, suPAR appears to sit at the nexus of immune signaling, chronic inflammation, and tissue damage. Among the protein's normal sources are fat cells, immune cells, and endothelial cells, which produce low baseline levels. But a team led by Reiser and David Scadden of the Harvard Stem Cell Institute showed in 2017 that in mice, immature "stemlike" cells in the bone marrow can release a pulse of suPAR when the immune system detects an attack.

Reiser believes suPAR is an ancient and unspecific way for the immune system to send urgent signals to the major organ systems when an organism faces a severe challenge from disease or the environment. Kidney damage, he says, is the long-term cost of that vital signaling mechanism. "As one example," he says, "you get infected, you release more suPAR, you open your kidneys up, and you can dump the big molecules out into the urine. Almost like a primitive coupling of the immune system to vital organs." In an acute infection, he says, the body urgently needs to flush out bacterial toxins, relatively big molecules. But if that inflammatory signaling becomes chronic, it takes its toll on kidney function—a trade-off that may have been acceptable earlier in human history, Reiser suggests, but is less so now. "If you live 40 years long, you can burn off the kidney this way, no problem," he says. "If you live to be 80, 90, 100, you might burn off your kidneys too soon."

For Reiser, the Emory cardiac biobank offered a chance to put to rest the notion that high levels of suPAR are simply a nonspecific sign of failing kidneys, not a cause. When he saw the 2014 heart risk paper from the Emory group, he had the obvious question: Could the large databank show whether suPAR levels predicted the onset of kidney disease years later? He immediately fired off an email to Hayek.

An organ under attack

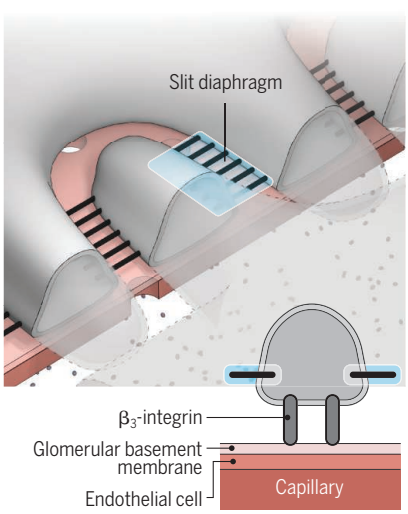
In one scenario for a severe form of kidney disease, a blood-borne molecule called soluble urokinase plasminogen activator receptor (suPAR) disrupts the organ's filtration units, or glomeruli, which remove waste and fluid from the bloodstream. Other molecules may intensify this attack.



A dangerous immune response
Animal models suggest immature immune cells in the bone marrow release more suPAR when an organism is under attack. The molecule, an all-purpose marker of ill health, may be directly toxic to the kidney.

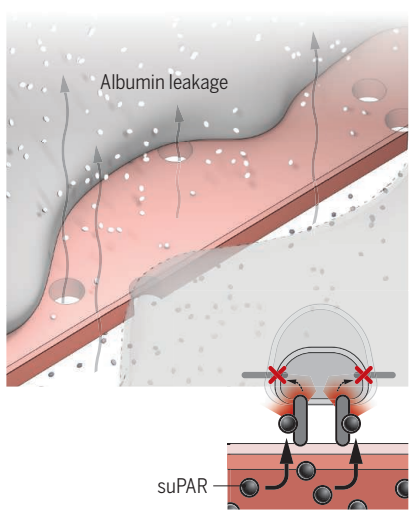
A HEALTHY FILTER

In each glomerulus, the footlike extensions of cells called podocytes wrap around capillaries, fitting together tightly to create narrow "slit diaphragms." The slits form a fine mesh that allows only small molecules to escape from the bloodstream into the urine.



KIDNEY DISEASE

By binding to β_3-integrin, a surface receptor on the podocyte feet, suPAR is thought to change their shape, disrupting the kidney's filter. That allows large proteins such as albumin to leak into the urine—a sign of kidney disease.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

The Emory-based group quickly agreed to conduct follow-up renal examinations in more than 1300 patients who had no evidence of kidney dysfunction when they enrolled. The team found a strong link between high suPAR levels and the later development of kidney disease. For patients with the highest levels of suPAR, the risk was three times that of patients in the lowest group, and suPAR levels could predict kidney disease up to 5 years before the first symptoms appeared. "The effect was huge," Hayek recalls.

The association was so robust, he says, that when the group first submitted its findings for publication, "the first response we got from [*NEJM*] was: 'How is that association so strong? Is that real? Something is wrong with your cohort.'" But after Hayek and Reiser found the same association in a second, unrelated cohort—the Women's Interagency HIV Study—*NEJM* published their findings in 2015. "In that paper," Reiser says, "we could show that suPAR is the strongest risk factor known in healthy people for new chronic kidney disease. Even stronger than hypertension, diabetes, black race—all of these risk factors that are known to be strong. When you adjust for those, suPAR had the strongest risk."

Jochen Reiser has spent years amassing evidence that suPAR mounts a powerful assault on the kidney.

In the latest piece of evidence, published last summer in *Nature Medicine*, Reiser collaborated with researchers at the African American Study of Kidney Disease and Hypertension, based at Johns Hopkins University School of Medicine in Baltimore, Maryland, to compare the influence of suPAR and two gene mutations known to predispose African-Americans to kidney disease. A study of about 600 participants revealed that if suPAR levels remain low, "no notable differences" in kidney dysfunction were apparent between people who had the high-risk "disease genes" and people who did not. Conversely, high levels of suPAR strongly predicted kidney disease in African-Americans, regardless of whether the individual had the genetic variants.

YET NEPHROLOGISTS are still divided about whether suPAR actually attacks the kidneys—and if so, how aggressively. Doubters point to the conflicting clinical results and the slow progress of kidney damage in Reiser's mice with elevated suPAR levels. The original 2011 animal and clinical data are "as complete as you can get," Vincenti says. "But at some point, there has to be independent duplication of that data."

Several unresolved issues might explain the discrepancies. Different forms of suPAR can circulate in the blood, and some variants might be more pathogenic than whole suPAR. And a team led by Minnie Sarwal of UCSF, Dany Anglicheau of Necker Hospital in Paris, and Reiser has shown that in FSGS, a second blood-borne factor, an anti-CD40 autoantibody, works with suPAR to attack podocytes. "Everyone agrees it's more complicated" than the initial findings in 2011 suggested, Reiser concedes. "But meanwhile, the data gets stronger and stronger that suPAR is the worst toxin you can have for the kidneys."

The controversy may not be resolved to everyone's satisfaction until a human trial indisputably shows that removing suPAR cures or slows the progression of kidney disease. Several groups are trying to develop a monoclonal antibody drug that would remove suPAR from the blood. One such group is Trisaq, a company Reiser and Sever founded in 2011. Vincenti said his group also has developed monoclonal antibodies to suPAR for clinical testing. "I was excited to try it in patients," he says. "But we could not demonstrate, at least in our samples, that suPAR was a biomarker for either FSGS or recurrent FSGS. [That's] held it back."

The first human proof may come not from a drug, but from a medical device. Miltenyi Biotec, a company in Bergisch Gladbach, Germany, makes apheresis devices, which remove substances from plasma, and it is developing a technology that would selectively scrub suPAR out of patients' blood. "The key question," notes CEO Stefan Miltenyi, "is if suPAR is the cause [of] renal diseases or just a bystander molecule." Miltenyi hopes to launch a clinical trial in 2019.

For FSGS patients such as Henry, who relies on 8-hour overnight sessions of dialysis to stay alive, a breakthrough therapy can't come soon enough. But suPAR is already beginning to influence clinical decisions. Singh, Henry's transplant physician, has used suPAR levels to manage the care of several kidney patients. And since 2013, every patient arriving at the emergency department at Copenhagen University Hospital Hvidovre has undergone suPAR testing to help physicians make triage and discharge decisions.

REISER OFTEN LIKENS suPAR to cholesterol—a key marker and disease-associated molecule that can be monitored and, perhaps, ultimately controlled. But the main lesson of suPAR, he believes, is cautionary in an age of genomics and personalized medicine. Although a huge amount of attention (and government coin) has been devoted to identifying genes associated with disease, the environment can sometimes trump them. "I think that the gene adds to the risk profile—it's part of the picture," he concedes. "But the environment is a way-underestimated modifier that becomes way more important, quite frankly, than the underlying gene event. And this is … a beautiful illustration of exactly that principle." ■

Stephen S. Hall is a science journalist in New York City.

PHOTO: RUSH PRODUCTION GROUP

BIOSAFETY AND BIOSECURITY

Risk-based reboot for global lab biosafety

New WHO guidance could expand access to lab facilities

By **Kazunobu Kojima,[1] Catherine Makison Booth,[1,2] Kathrin Summermatter,[3] Allan Bennett,[4] Marianne Heisz,[5] Stuart D. Blacksell,[6,7] Michelle McKinney[8]***

Laboratory biosafety is fundamental to controlling exposure to pathogens, protecting the laboratory workforce and the wider community against inadvertent exposures or releases. Since 1983, the World Health Organization (WHO) *Laboratory Biosafety Manual* (*LBM*) has encouraged countries to implement basic concepts in biological safety and to develop national codes of practice for the safe handling of pathogenic microorganisms in laboratories. But as technologies continue to evolve, and with them potential threats and benefits to laboratory safety, so too must approaches to biosafety. With revision toward the fourth edition of the *LBM* under way, we propose a shift in focus to a risk-based, technology-neutral, and cost-effective approach to biosafety, making sure that laboratory facilities, safety equipment, and work practices are locally relevant, proportionate, and sustainable. This will allow more flexibility in laboratory design, reduce focus on pathogen risk groups and biosafety levels as the de facto starting point of laboratory considerations, and place more emphasis on human factors and worker training. Improved sustainability of laboratory oper-

PHOTO: XINHUA/ALAMY STOCK PHOTO

Published by AAAS

A researcher dons a protective suit at China's National Biosafety Laboratory in Wuhan, Hubei Province, China.

ations through lower construction and operating costs, particularly in resource-limited settings, may pave the way for equitable access to clinical and public health laboratory tests and biomedical research opportunities, without compromising safety.

The *LBM* has been in broad use at all levels of laboratories and biomedical sectors globally, serving as a de facto standard that represents best practices, especially where resources are less plentiful and regulatory frameworks are less well developed. The first publication of the *LBM* coincided with the year during which polymerase chain reaction (PCR) was invented. Since then, over the course of many decades and three editions of the *LBM*, many common biosafety concepts and practices have been applied without assuming the availability of technologies such as molecular testing, which, because it does not require propagation of the pathogen, poses considerably less likelihood for exposure compared to conventional culture-based diagnostics.

PRECAUTIONARY AND PRESCRIPTIVE

Under earlier editions of the *LBM*, local risk assessment was not regularly and extensively practiced, which necessitated a more precautionary and prescriptive approach to defining safety criteria for pathogens and laboratory facilities. Biological agents that have minimal or no available preventative or therapeutics, such as Ebola virus, have typically been handled only in high-containment facilities, irrespective of the procedure, which might have contributed to unnecessary proliferation of costly high-containment laboratories.

Local risk assessment was encouraged in the third edition of the *LBM* in 2004, but, at the same time, it did not adequately address what had become overly simplified implementation of biosafety programs born of the tendency to equate pathogen risk groups (RGs, or hazard groups) with laboratory biosafety levels (BSLs, or containment levels). RGs and BSLs, which both have the same numerical nomenclature, have incorrectly been assumed to be essentially equated—for example, all work with RG2 pathogens must be conducted at BSL2, RG3 at BSL3, RG4 at BSL4, and so on.

The consequence of this equation is that pathogen and local risk assessments are not routinely carried out, missing opportunities to optimize the means of risk control proportionate to the assessed risks. Although RGs may be a useful tool for national regulatory regimes, applying them universally does not take into account national epidemiological profile, such as endemicity, nor does it necessarily consider local application of procedures and techniques.

When the concept of RGs was introduced in the *LBM* for national oversight regimes, there was clear guidance that RG determination was to be specific to a particular jurisdiction, based on local and/or national conditions. But this approach was lost over time as many applied this concept internationally. A major source of confusion is inconsistency in the description of pathogen RGs and BSLs among countries (*1*).

ENGINEERED AND UNSUSTAINABLE

At higher BSLs, there has been an emphasis on strict application of detailed engineering controls, but much less consideration has typically been given to procedural and human factors, despite the fact that the same pathogen could pose considerably different risks according to the procedures to be undertaken. Consequently, structural systems such as complicated and costly centralized ventilation units, complex waste disposal, and building automation systems tend to take precedence over risk assessment, workforce development, and training on good microbiological practices and procedures.

Yet a review of recent laboratory-acquired infections (*2*) demonstrated that most were caused not by malfunctions of engineering controls but by human factors such as improper use of personal protective equipment, disregard for or inadequate risk assessments, and lack of standard operating procedures and/or properly trained staff. The best-designed and most engineered laboratory is only as good as its least competent worker.

For routine diagnostics, and pathogen research not involving high concentrations and large volumes of infectious materials, there is no evidence that using complex biological containment facilities will provide a safer working environment, though reliance on such facilities will most certainly increase costs. Facilities of complex design are costly to build, operate, and maintain, particularly those with centralized ventilation systems or complex waste-disposal systems. Day-to-day running and maintenance costs of high-containment laboratories are often not prioritized during budgeting and planning.

Most laboratories, especially those in resource-limited countries, cannot reasonably afford to build, run, and maintain such facilities without long-term donor or external support but often strive to acquire them on the basis of past global trends emphasizing building of containment laboratories, where local risk assessment was not a routine practice. Where there is a lack of financial resources for laboratory services, running complex containment facilities can take valuable and limited resources away from day-to-day laboratory operations, potentially leaving lower-profile laboratories more vulnerable to adverse occurrences. The lack of technical resources, such as appropriately trained and locally available engineers, further complicates these challenges.

[1]World Health Organization, Geneva, Switzerland. [2]Health and Safety Laboratory, Buxton, UK. [3]Institute of Virology and Immunology, Mittelhäusern, Switzerland. [4]Public Health England, Salisbury, UK. [5]Public Health Agency of Canada, Ottawa, Canada. [6]Mahidol Oxford Tropical Medicine Research Unit, Bangkok, Thailand. [7]Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Churchill Hospital, Oxford, UK. [8]Centers for Disease Control and Prevention, Washington, DC, USA. *Present address: National Institutes of Health, Bethesda, MD, USA. Email: kojimak@who.int

HUMAN-FOCUSED AND FLEXIBLE

In the new *LBM*, all risk-mitigation measures should be selected using a risk- and evidence-based, rather than BSL checklist, approach. They should be informed by the consequence and likelihood of exposure to a pathogen. The same principles apply when considering emerging technologies in the context of dual-use research of concern and synthetic biology; biosecurity considerations should be proportionate to the assessed risks to complement biosafety effectively, sharing the same goal to prevent unnecessary release and exposure of biological materials, be it accidental or deliberate (*3*–*5*). A risk-based approach also underpins the revised World Organisation for Animal Health standard for managing biological risk in veterinary laboratories and animal facilities (*6*).

This multifactorial assessment includes criteria such as route(s) of infection; pathogenicity and infectious dose; prophylaxis or vaccine availability; disease severity and mortality; contagiousness; endemicity; high-risk laboratory procedures (that is, work with aerosols, high titers or volumes, sharps, animals); laboratory workforce competency; highly susceptible individuals and/or hosts; volume and titer of the biological agent(s) being produced and/or handled; and biosecurity (potential for misuse and/or weaponization).

In lieu of defined biosafety levels, "core requirements" are proposed in the revised *LBM*, a combination of common biosafety elements to be implemented and used and a minimum requirement for safe work during most laboratory procedures. These include not only the laboratory facility and equipment but also procedures and appropriately trained and competent staff, which too often received disproportionately less attention.

Where the local risk assessment necessitates, heightened control measures such as the use of biological safety cabinets, extra personal protective equipment, and, in rare cases, high containment would be used to mitigate the identified risk(s) of exposure to the biological agent(s) being handled. This reinvented definition is intended to be more continuous than the discrete classification of BSL. The intent is to choose the right set of safety measures to control the identified risk(s) and not rely on a predefined "universal" set that may be an overdesign, depending on the pathogen and/or activities proposed.

Trainees exercise safe glove removal, using fluorescent tracers to visualize potential contamination.

PHOTO: KAZUNOBU KOJIMA

GLOBALIZING BIOSAFETY

The need to update international laboratory biosafety guidance is part of a broader initiative to globalize biosafety, emphasizing principles and approaches that are accessible to countries spanning a broad range of financial, technical, and regulatory resources. The WHO revised the International Health Regulations (IHR) in 2005 "to help the international community prevent and respond to acute public health risks that have the potential to cross borders and threaten people worldwide." This requires that all 196 States Parties to the IHR are well prepared for potential outbreaks and new diseases, including prompt early diagnosis to enable infection prevention and control that should entail safe and secure handling and transport of infectious agents. The importance of better recognition of biosafety and biosecurity among all health professionals is reflected by their inclusion within the 19 technical areas in the IHR Joint External Evaluation tool (*7*).

Countries that have or are developing national oversight regimes for biosafety and biosecurity should approach oversight with a goal to be as flexible as possible to allow for multiple solutions and mitigation measures. If RGs have been developed at the national level and embedded within legal frameworks, there should be an opportunity for periodic updates so that new evidence that supports changes to the risk profile of a pathogen are easily incorporated into an RG change. Oversight systems should consider the procedures being conducted in facilities, and how they contribute to or reduce risk, and be addressed accordingly.

Whether prescribed or recommended, specific mitigation measures may best be applied through mechanisms with greater flexibility than legislation, such as guidelines or standards. Oversight bodies should actively engage with biosafety practitioners within their jurisdictions, to ensure consideration of practical input from the field and maintain local and national relevance.

The process of revising the *LBM* and preparing accompanying technical guidance is ongoing. A critical element is an extensive consultation process of a broad range of stakeholders. It is our hope that our proposal here can inspire some readers to contribute constructively to this process. ■

REFERENCES AND NOTES

1. D. Tian, T. Zheng, *PLOS ONE* **9**, e101163 (2014).
2. N. Wurtz *et al.*, *Eur. J. Clin. Microbiol. Infect. Dis.* **35**, 1247 (2016).
3. WHO, "Biorisk management—Laboratory biosecurity guidance" (WHO, Geneva, Switzerland, 2006).
4. European Committee for Standardization (CEN), "Laboratory biorisk management" (CEN Workshop Agreement, CWA 15793, CEN, Brussels, Belgium, 2011).
5. The biorisk management reference document in (*4*) expired in 2014. Efforts have been ongoing to convert it to International Standard 35001 of the International Organization for Standardization (ISO), but this has yet to be formally approved by ISO member countries.
6. World Organisation for Animal Health (OIE), in *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* (OIE, Paris, France, 2015), chap. 1.1.4.
7. WHO, "Joint External Evaluation tool: International Health Regulations" (WHO, Geneva, Switzerland, 2016).

ACKNOWLEDGMENTS

The authors thank relevant WHO staff for their support of this revision project, in particular, S. Cognat, F. Fuchs, and G. Rodier. We are grateful for financial assistance from the Biosecurity Engagement Program of the U.S. Department of State and the Global Partnership Program of Global Affairs Canada. S.D.B. is supported by the Wellcome Trust of the United Kingdom. The authors alone are responsible for the views expressed in this article, and they do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

10.1126/science.aar2231

PERSPECTIVES

ECOLOGY

Plant responses to CO_2 are a question of time

Long-term experiments show unexpected plant responses to elevated CO_2 concentrations

By **Mark Hovenden**[1] *and* **Paul Newton**[2]

Rising carbon dioxide (CO_2) concentrations in the atmosphere as a result of fossil fuel burning are expected to fertilize plants, resulting in faster growth. However, this change is not expected to be the same for all plants. Rather, scientists believe that differences in photosynthetic mechanism favor one plant group—the C_3 plants—over the other, the C_4 plants. On page 317 of this issue, Reich *et al.* (*1*) show that, although this expectation is met in the first few years of a long-term experiment, the situation reverses after 15 to 20 years, with important implications for future crop production and ecosystems.

In 1966, Hatch and Slack (*2*) found that some plants have a distinct mechanism for assimilating CO_2 from the atmosphere. This has profound ecological consequences. The ancestral method of photosynthesis combines CO_2 with a five-carbon molecule to produce two identical three-carbon molecules. However, the enzyme that catalyzes this reaction also combines the same five-carbon compound with dioxygen, thus reducing the carbon assimilation rate and, hence, growth. Hatch and Slack discovered that some plants can avoid this by first combining CO_2 from the atmosphere with a three-carbon molecule, producing a four-carbon molecule as the first stable product in photosynthesis. Plants with this pathway are known as C_4 plants, distinguishing them from those with the ancestral pathway, termed C_3 plants.

Although only about 3% of the global plant species are C_4 plants (*3*), they play a crucial role in many ecosystems, particularly savannas and grasslands (*4*, *5*). C_4 species contribute 25% of land biomass globally (*3*), provide forage for animals in both natural (for example, Serengeti) and managed (for example, Great Plains) grasslands, and contribute 14 of the world's 18 worst weeds (*6*). It is clearly important, therefore, to predict the future distribution and abundance of C_4 plants.

Reich *et al.* report findings from a 20-year experiment (see the photo) in the state of Minnesota, USA, where they exposed C_3 and C_4 plants to an elevated concentration of atmospheric CO_2—arguably the most predictable global change. The C_4 method of photosynthesis appears to have evolved during a period of declining atmospheric CO_2 concentration and allows the plants to use CO_2 more efficiently than C_3 species (*4*). Today, the atmospheric CO_2 concentration is higher than at any other time in the past 500,000 years and continues to rise (*7*). Most scientists expect C_3 plants to benefit from this additional CO_2 and outcompete C_4 species, because C_3 photosynthesis increases in efficiency with increasing CO_2 concentration to a far greater extent than does C_4 photosynthesis.

For 20 years, Reich *et al.* have exposed C_3 and C_4 plants to elevated CO_2 concentrations at their site in Minnesota.

Over the initial years of Reich *et al.*'s experiment, the results conformed to this expectation: C_3 plants responded strongly to elevated CO_2, whereas C_4 plants were unresponsive. However, as the experiment progressed, the response of the C_3 plants declined, and the response of the C_4 plants became stronger. These changes were so marked and consistent that after 15 to 20 years, the response of C_3 plants to CO_2 was negligible but that of C_4 plants was strong. Indeed, the growth stimulation from the CO_2 in the C_4 plants during the final 5 years of the experiment exceeded that in C_3 plants at any stage of the experiment.

Differences in photosynthesis rates or interannual variation in temperature and rainfall could not explain this result. What was important was how soil nitrogen dynamics changed in response to elevated CO_2 and how this differed for the C_3 and C_4 plants. The growth response to elevated CO_2 for both C_3 and C_4 species was directly and positively correlated to the effect of elevated CO_2 on the net mineralization rate (that is, the rate at which soil nitrogen was converted from organic to inorganic forms).

The availability of soil nitrogen to plants can have a strong influence on whether plants respond positively to elevated CO_2 (*8*). Mineralization is a key factor in this availability. That the nitrogen mineralization rate increased over time in soil under C_4 plants, but declined in soil under C_3 plants, could therefore explain the growth responses. Thus, given sufficient time, the biogeochemical response to the rise in CO_2 concentration appeared to overwhelm the initial, physiology-driven growth response.

[1]*School of Biological Sciences, University of Tasmania, Hobart, Tasmania 7001, Australia.* [2]*Land and Environmental Management, AgResearch Grasslands, Palmerston North, New Zealand. Email: mark.hovenden@utas.edu.au*

The different growth response of C_3 and C_4 species to elevated CO_2 concentrations has important consequences. Earth system models assume a positive relationship between C_3 plant growth and CO_2 concentration, an assumption that Reich *et al.*'s results challenge. As a result, Earth's future carbon-sequestration potential could be far lower than predicted by these models, because most vegetation is C_3 dominated. Furthermore, agricultural production involving C_3 plants such as wheat, rice, and temperate pasture species might not increase with the rising CO_2 concentration. It is also likely that introduced plant species, many of which are C_4 plants, will increase in vigor and become an even greater problem than they are now. Finally, ecosystems in which C_3 and C_4 species grow together are likely to change as the competitive balance alters, with consequences for forage supply and its nutritional quality (*5*). Increased growth rates of C_4 species are even more likely in the future because of global warming; the efficiency of C_3 photosynthesis declines with increasing temperature, but this does not occur in C_4 plants, making them more competitive in warmer conditions (*9*).

Reich *et al.* were only able to make their discoveries because their experiment ran uninterrupted for two decades. This is extremely rare globally, showing that funding for long-term global-change experiments is a necessity. The experiment relied on a concerted effort to continually apply for funding, given the largely short-term nature of funding cycles. Because most funding agencies place a value on innovation and novelty, scientists are forced to come up with new reasons and new measurements to keep existing experiments running. The tenacity of Reich *et al.* and their ability to keep their experiment running has overturned existing theory and should lead to changes in how we think about and prepare for Earth's future. Who knows how many processes remain undiscovered because of the unwillingness of funding agencies to support long-term experiments? ■

REFERENCES

1. P. B. Reich, S. E. Hobbie, T. D. Lee, M. A. Pastore, *Science* **360**, 317 (2018).
2. M. D. Hatch, C. R. Slack, *Biochem. J.* **101**, 103 (1966).
3. R. F. Sage, M. Stata, *J. Plant Physiol.* **172**, 104 (2015).
4. R. F. Sage, *J. Exp. Bot.* **68**, e11 (2017).
5. D. A. Wedin, in *Warm-Season (C4) Grasses*, L. E. Moser, B. L. Burson, L. E. Sollenberger, Eds. (American Society of Agronomy, Madison, WI, 2004), vol. 45, pp. 15–50.
6. L. G. Holm, D. L. Plucknett, J. B. Pancho, J. P. Herberger, *The World's Worst Weeds: Distribution and Biology* (Univ. Press of Hawaii, Honolulu, 1977).
7. D. Lüthi *et al.*, *Nature* **453**, 379 (2008).
8. R. E. McMurtrie *et al.*, *Funct. Plant Biol.* **35**, 521 (2008).
9. R. F. Sage, D. S. Kubien, *Photosyn. Res.* **77**, 209 (2003).

10.1126/science.aat2481

MATERIALS

On the quest for the strongest materials

Diamond nanoneedles have strength approaching the theoretical maximum

By **Javier LLorca**[1,2]

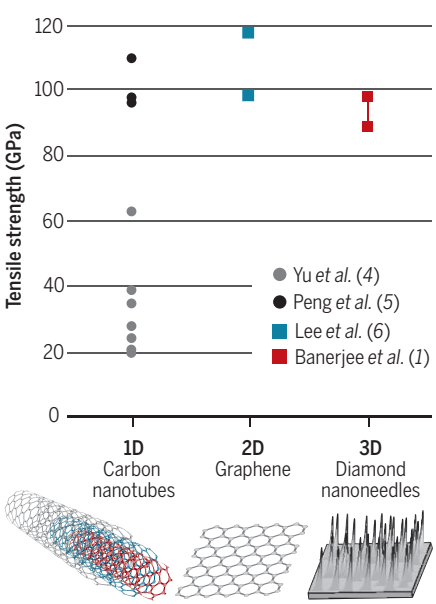
The strength of a material is a measure of its ability to withstand a load without breaking. Scientists in search of the strongest materials have recently turned their attention to nanomaterials, which have few of the defects that typically reduce a material's strength. On page 300 of this issue, Banerjee *et al.* (*1*) show that when nanoscale single-crystal diamond needles are elastically deformed, they fail at a maximum local tensile stress of ~89 to 98 GPa, which is very close to the theoretical limit for this material.

The maximum possible strength that a material can have in either tension or shear is controlled by the fracture of the interatomic bonds and is on the order of 10% of the elastic or shear moduli, respectively. However, it is difficult to achieve these strengths in practice because defects in the solid will lead to inelastic relaxation or brittle fracture well before the atomic bond can be stretched to the theoretical limit. Maximum elastic tensile strains supported by bulk solids are between 0.2 and 0.4%, whereas tensile strains of up to 4% have been measured in micrometer-size whiskers (*2*). Recent progress in nanomaterial synthesis and nanomechanical testing has opened the possibility of probing the strength of material systems that are practically free of defects (*3*). In parallel, atomistic simulations based on density-functional theory and molecular dynamics can predict accurately the fracture strength of perfect crystals and allow the influence of defects and free surfaces on this property to be explored.

Because the carbon-carbon bond is the strongest in nature, the search for the strongest material has focused on one-dimensional carbon nanotubes and two-dimensional graphene nanoscale objects (see the figure). Experimental results and ab initio calculations indicate that the elastic modulus of carbon nanotubes and graphene is ≈1 TPa (*4*–*6*). The strengths measured in both types of nano-objects are thus very close to the theoretical limit (see the figure). By contrast, the reported tensile strength of bulk cubic diamond is much smaller (<10 GPa) (*7*). These differences in strength have been partially attributed to brittle fracture from defects during tensile deformation of bulk samples. Higher tensile strengths (up to 20 GPa) for diamond have been reported from Hertzian indentation tests (*8*), but these values must be treated with caution because of the uncertainties associated with the spherical indentation to determine the tensile strength. Regardless of the experimental technique, diamond fractures by cleavage along the (111) plane, which has the lowest fracture energy. The strengths reported by Banerjee *et al.* (which correspond to a tensile strain of ≈9%) are very close to the theoretical limit for diamond and to the maximum strength values reported for carbon nanotubes and gra-

Strong, stronger...

Experimental measurements have shown similarly high tensile strengths for multiwall carbon nanotubes, graphene, and diamond nanoneedles.

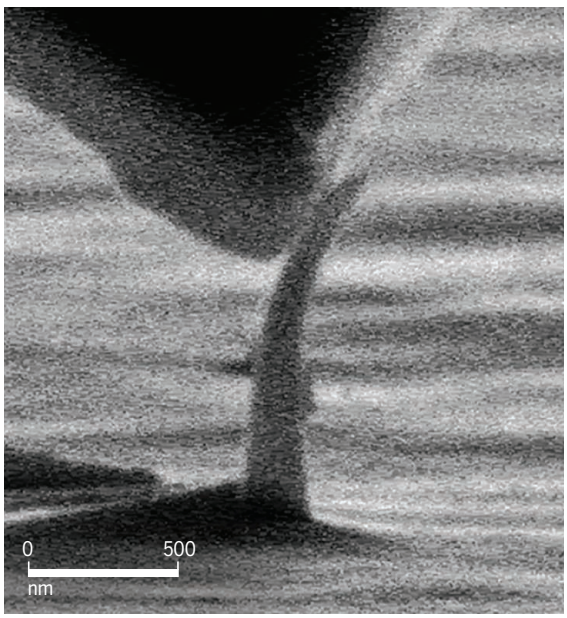


GRAPHIC: ADAPTED BY N. CARY/*SCIENCE*

[1]*IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain.* [2]*Department of Materials Science, Polytechnic University of Madrid, E. T. S. de Ingenieros de Caminos, 28040 Madrid, Spain. Email: javier.llorca@imdea.org*

phene (see the figure). Thus, the strength of bulk diamond, which has sp^3 bonds, is equivalent to that reported for carbon nanotubes and graphene with sp^2 bonds.

The authors applied these large elastic strains by bending the nanoneedles (see the image); fracture was not triggered at smaller strains because of the small volume of the nanoscale needles, the paucity of defects, and the smoothness of the surface, which was produced with careful reactive ion etching from a <111>-oriented diamond film produced by means of chemical vapor deposition. Fracture occurred through cleavage along (111) planes, and the strength measured was supported by combined density-functional and molecular-dynamics simulations at 300 K. The latter predicted a maximum critical strain of 13%, corresponding to a tensile strength of 130 GPa.



A nanoindenter tip bends a diamond nanoneedle.

These results open the possibility of modifying fundamental properties of diamond through elastic strain engineering (*9*). The properties of any crystalline material depend on the lattice parameters, which are dictated by the electronic structure. Large elastic deformations (around 10%) modify the electronic structure, leading to changes in electronic, magnetic, catalytic, and other properties that could be tuned as a function of the applied strain tensor. This concept has been used to enhance the drive current in complementary metal-oxide semiconductor technology by improving the electron and hole mobility through lattice straining (*10*). Other applications include the transformation of paramagnetic materials into ferromagnetic ones (*11*) and the manipulation of mechanochemical coupling so as to increase catalytic activity (*12*).

However, the elastic strains applied to modify the properties in all these cases were much lower (below 5%) than those reported for carbon nanotubes, graphene, and diamond needles, which were closer to 10%. Changes in properties should increase with the elastic strain and, as postulated by Gilman (*13*), the electronic structure should be drastically modified near the critical bond-breaking strain, leading to unusual or singular chemical and physical properties. Thus, exploration of the effect of very large elastic strains on the properties of crystalline solids may lead to the discovery of new or unexpected behaviors.

Many applications of elastic strain engineering require the elastic strains to be distributed over a substantial area or volume. This can be achieved by means of epitaxial growth (*10*) or severe plastic deformation (*11*) in the case of bulk materials. However, this task is more challenging in the case of nanoscale objects because it is difficult to maintain the mechanical continuity between these objects. For instance, the maximum strengths reported for carbon nanotube fibers, which are made up of bundles of nanotubes, are ~2.5 GPa, which is well below the values reported for single carbon nanotubes (see the figure) (*14*–*15*). Elastic straining techniques that can be applied to bulk materials, such as the elastic bending of nanoneedles in Banerjee *et al.*, can potentially be extended to large surfaces and different materials. Thus, this strategy can be used to explore the effect of large elastic strains on chemical and physical properties of diamond and other solids. ■

PHOTO: A. BANERJEE *ET AL.*

REFERENCES

1. A. Banerjee *et al.*, *Science* **360**, 300 (2018).
2. S. S. Brenner, *Science* **128**, 569 (1958).
3. T. Zhu, J. Li, *Prog. Mater. Sci.* **55**, 710 (2010).
4. M.-F. Yu *et al.*, *Science* **287**, 637 (2000).
5. B. Peng *et al.*, *Nat. Nanotechnol.* **3**, 626 (2008).
6. G.-H. Lee *et al.*, *Science* **340**, 1073 (2013).
7. J. E. Field, *Rep. Prog. Phys.* **75**, 126505 (2012).
8. K. Miyoshi, *Solid Lubrication Fundamentals and Applications* (CRC Press, 2001).
9. J. Li, Z. Shan, E. Ma, *MRS Bull.* **39**, 108 (2014).
10. Y. Sun, S. E. Thompson, T. Nishida, *J. Appl. Phys.* **101**, 104503 (2007).
11. C. M. Cepeda-Jiménez *et al.*, *Acta Mater.* **123**, 206 (2017).
12. B. Yildiz, *MRS Bull.* **39**, 147 (2014).
13. J. J. Gilman, *Electronic Basis of the Strength of Materials* (Cambridge Univ. Press, 2008).
14. A. M. Beese *et al.*, *ACS Nano* **8**, 11454 (2014).
15. C. González *et al.*, *Prog. Mater. Sci.* **89**, 194 (2017).

10.1126/science.aat5211

NEUROSCIENCE

Whispering neurons fuel cortical highways

Synaptic communication accelerates neuronal migration in the developing brain

By **Alejandro F. Schinder**[1] *and* **Guillermo M. Lanuza**[2]

The mammalian neocortex is one of the most intricate entities found in nature, both in terms of structure and function. It is the brain region responsible for the execution of high-order functions, including sensory perception, motor control, cognition, and speech. Its development is equally complex because it requires that millions to billions (depending on the species) of neurons assemble in distinct layers and connect with exquisite precision to perform complicated information processing operations. During embryonic development, formation of the cerebral cortex involves the migration of excitatory neurons generated in the ventricular zone toward the cortical plate, where they establish their final position in six well-defined horizontal layers consisting of different types of neurons and architecture. Along this migratory phase, developing neurons undergo a morphological transition from multipolar shape to bipolar morphology. Bipolar neurons exhibit faster locomotion, quickly reaching their final destination. On page 313 of this issue, Ohtaka-Maruyama *et al.* (*1*) reveal that this important switch to bipolar neurons is influenced by glutamate release from neurons located at the subplate, just beneath the cortical plate. Subplate neurons trigger this transformation by making transient synaptic contacts with multipolar neurons in transit to the corti-

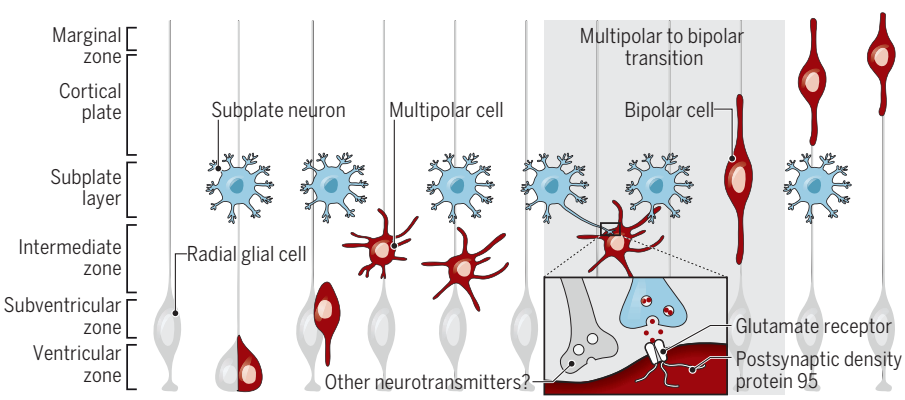
[1]*Laboratorio de Plasticidad Neuronal, Fundación Instituto Leloir, Instituto de Investigaciones Bioquímicas de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Avenida Patricias Argentinas 435, Buenos Aires C1405BWE, Argentina.* [2]*Laboratorio de Genética del Desarrollo Neural, Fundación Instituto Leloir, Instituto de Investigaciones Bioquímicas de Buenos Aires, CONICET, Avenida Patricias Argentinas 435, Buenos Aires C1405BWE, Argentina. Email: aschinder@leloir.org.ar*

cal laminae. Understanding this process is important because disruption of neocortical migration results in several human neurodevelopmental diseases.

During cortical development, excitatory neurons are derived from radial glial cells located at the ventricular zone (*2*–*4*) (see the figure). The earlier-born neurons form the preplate, which later splits into the more superficial marginal zone and the deeper subplate (SP). The cortical plate then emerges in between these two layers, fueled by the accumulation of neurons migrating from the ventricular zone. Layers are established by the subsequent addition of neurons migrating in an inside-out manner; the most recently generated neurons migrate through older cells toward the marginal zone, giving rise to the neocortex (*3*–*5*). Consequently, the deeper layers are built early and the superficial layers arise toward the end of cortical development.

The migratory journey of cortical neurons

During the multipolar stage at the intermediate zone, neurons receive transient glutamatergic synaptic inputs from subplate neurons, which influence their conversion into fast-locomoting bipolar neurons that ultimately form the cortical layer.



Developing neurons undergo a series of choices regarding their specification (what type of neuron?), migration (what cortical layer?), growth (where to project dendrites and axon?), and selection of their synaptic partners (input and output cells). These choices are made according to a combination of intrinsic programs and extrinsic signaling (*3*–*5*). Cortical lamination involves the formation of transient cell layers that direct the migration and connectivity of new neurons and disappear in the mature brain. The SP is an early transient layer, consisting of heterogeneous neuronal populations exhibiting mature morphological and functional properties (*6*). SP neurons receive functional inputs from thalamic projections and are also connected among themselves and with other cortical neurons through chemical and electrical synapses. Activity of the SP network is relevant for guiding the growth of thalamic axons toward cortical layer IV and for the maturation of excitatory pyramidal cells in that layer (*7*, *8*). Because new neurons navigate across the SP to reach the cortical plate, Ohtaka-Maruyama *et al.* examined whether SP neurons might influence radial migration.

The process of radial migration (*2*, *5*, *9*) is characterized by three phases: Multipolar neurons move slowly, in a winding manner, almost pausing at the intermediate zone before reaching the SP; stalling neurons convert to bipolar morphology; bipolar neurons migrate at faster speed along radial glial fibers through the cortical plate, until reaching their final cortical lamina (see the figure). The authors characterized neuronal migration in brain slices obtained from mouse embryos. The use of fluorescence microscopy techniques revealed that SP neurons project axons toward the intermediate zone. Some of those axons express the vesicular glutamate transporter VGLUT2 and contact multipolar neurons, which suggests that they might form glutamatergic synapses. Further analyses revealed that highly active synapses form in the area where SP axons mingle with multipolar cells. Such activity is likely driven by excitatory thalamocortical projections that directly activate SP networks (*6*, *7*). The authors showed that neuronal migration is halted when electrical activity in SP neurons is reduced or when synaptic release from their axons is blocked. Together, these results suggest that the switch from multipolar cells to bipolar, fast-migrating neurons requires glutamate release from SP neurons, consistent with the role of glutamatergic signaling in neuronal migration during brain development (*10*).

However, it is also conceivable that additional players contribute to radial migration in a manner that relies on synaptic communication. Electron microscopy revealed contacts with heterogeneous structure formed onto multipolar cells. Those synapse-like structures do not resemble canonical glutamatergic synapses (*11*): Most terminals lacked synaptic vesicles, with these seeming to be present in low numbers and in only 20 to 30% of presynaptic terminals; synaptic shape was irregular and bearing atypical ultrastructural morphology. These features might allow multipolar cells to respond to the transient nature of the putative glutamatergic synapses, but this view prompts additional questions regarding the activity-dependent control of radial migration. Is glutamate the main trigger for the switch? Do multipolar cells integrate signals arising from additional neurotransmitters, such as γ-aminobutyric acid (GABA), exerting depolarizing effects that commonly promote neuronal maturation (*12*)? Do multipolar cells receive synaptic signals from other neuronal populations within or outside the developing cortex? It will also be critical to interrogate the mechanisms that link synaptic activity to cytoskeletal reorganization, which ultimately executes the multipolar to bipolar transformation (*13*, *14*).

The experimental approach used by Ohtaka-Maruyama *et al.* bears the potential to explore structure and function of synaptic connections between the SP network and multipolar cells in depth. For instance, optogenetic and electrophysiological recordings would readily allow combining the restricted and specific stimulation of SP neuron terminals with recordings of postsynaptic responses in multipolar cells to unambiguously establish the nature of the chemical and electrical synapses that influence neuronal migration. Ultimately, it will be crucial to develop complete loss-of-function models to unearth the extent to which neurotransmitter release from SP neurons shapes the rules of cell migration and stratification in the mammalian neocortex. ■

REFERENCES

1. C. Ohtaka-Maruyama *et al.*, *Science* **360**, 313 (2018).
2. S. C. Noctor *et al.*, *Nature* **409**, 714 (2001).
3. L. Custo Greig *et al.*, *Nat. Rev. Neurosci.* **14**, 755 (2013).
4. S. Lodato *et al.*, *Annu. Rev. Cell Dev. Biol.* **31**, 699 (2015).
5. A. Gupta *et al.*, *Nat. Rev. Genet.* **3**, 342 (2002).
6. P. O. Kanold *et al.*, *Annu. Rev. Neurosci.* **33**, 23 (2010).
7. K. L. Allendoerfer *et al.*, *Annu. Rev. Neurosci.* **17**, 185 (1994).
8. P. O. Kanold *et al.*, *Science* **301**, 521 (2003).
9. H. Tabata *et al.*, *J. Neurosci.* **23**, 9996 (2003).
10. H. Komuro *et al.*, *J. Neurobiol.* **37**, 110 (1998).
11. K. M. Harris *et al.*, *Cold Spring Harb Perspect. Biol.* **4**, a005587 (2012).
12. D. F. Owens *et al.*, *Nat. Rev. Neurosci.* **3**, 715 (2002).
13. C. Conde *et al.*, *Nat. Rev. Neurosci.* **10**, 319 (2009).
14. J. A. Cooper, *Front. Cell. Neurosci.* **8**, 386 (2014).

10.1126/science.aat4587

GRAPHIC: N. DESAI/*SCIENCE*

NEURODEGENERATION

Parkin function in Parkinson's disease

Models of Parkin-mediated ubiquitination lend insight into the role of pathological mutations

By **Connor Arkinson** *and* **Helen Walden**

Parkinson's disease (PD) is the second most common neurodegenerative disease, and is characterized by involuntary shaking, muscle rigidity, and the progressive loss of dopaminergic neurons. In ~5 to 10% of PD cases there is a genetic association, with almost 20 genes attributed to date. One example is early-onset autosomal recessive PD (ARPD), for which the majority of cases are linked to mutations in the Parkin gene (*PRKN*; also known as *PARK2*). *PRKN* encodes the E3 ubiquitin ligase Parkin, which plays important roles in mitochondrial quality control and turnover. Parkin, although localized to the mitochondria under certain conditions, is primarily cytosolic (*1*). A second ARPD-associated gene, *PINK1* (*PTEN-induced putative kinase 1*), encodes a mitochondrially tethered kinase that regulates Parkin activity through phosphorylation events. Mutations in *PINK1*, although rare, are associated with a phenotype similar to that of ARPD patients with *PRKN* mutations. Numerous mutations throughout *PRKN* are linked to ARPD, making the functional examination of Parkin crucial to understanding ARPD pathogenesis. A wealth of structural studies have transformed our knowledge of Parkin regulation and catalytic mechanisms. However, the current picture is incomplete, leading to several possible models of Parkin catalysis, which has implications for understanding how the ARPD-associated mutations affect the protein and thus PD pathogenesis.

The major function of Parkin is to ligate ubiquitin to lysine residues, an essential posttranslational modification involved in almost every cellular pathway. Ubiquitination occurs via the sequential action of three enzymes: an E1-activating enzyme, an E2-conjugating enzyme, and an E3 ligase. Parkin belongs to the RBR [RING–in-between-RING (IBR)–RING] family of E3 ligases, which contain a RING1 domain for E2 recruitment, a catalytic domain (RING2) through which transfer of ubiquitin is mediated, and a linking IBR domain (*2*). There are 14 RBR family members, including HOIP (HOIL-1–interacting protein), HHARI (human homolog of Ariadne), and Dorfin (double ring-finger). Owing to its relevance in ARPD, Parkin is the most studied RBR ligase and thus also serves as an important model for understanding RBR mechanism.

Parkin comprises five domains: ubiquitin-like domain (Ubl), RING0, RING1, IBR, and RING2. The Ubl and RING0 domains are unique to Parkin, with the RBR module common to all RBR ligases. ARPD-associated *PRKN* mutations are distributed across all five domains and are also found within domain-domain interfaces and linker regions. This broad distribution of mutations across all of Parkin suggests the structural integrity of all domains is important in ARPD pathogenesis. Parkin adopts an autoinhibited conformation mediated by its amino-terminal Ubl domain (*3*). The Ubl is joined to the carboxyl-terminal RING0-RBR (R0RBR) domains by a flexible 65–amino acid linker. Crystal structures of R0RBR Parkin, with both the Ubl and linker removed, reveal a rigid core formed of RING0, RING1, and RING2 domains, with the IBR domain showing degrees of flexibility (*4*, *5*). Structures of Parkin that include the Ubl domain show that the Ubl is associated with RING1 (*4*, *6*, *7*). Each structure highlights several autoinhibitory features: The RING0 partially occludes the catalytic cysteine (Cys^{431}); a small helical element [repressor element of Parkin (REP)], along with the Ubl domain, blocks the predicted E2 binding site on RING1; and finally, the distance between the RING2 Cys^{431} and predicted E2 catalytic cysteine is over 30 Å, which is too far to allow transfer of ubiquitin. Thus, for full Parkin function, not only are activators required to disrupt the autoinhibitory mechanisms, but large conformation changes are predicted to bridge the distance between catalytic sites.

A patient with advanced Parkinson's disease: Mutations in Parkin are associated with early-onset Parkinson's disease.

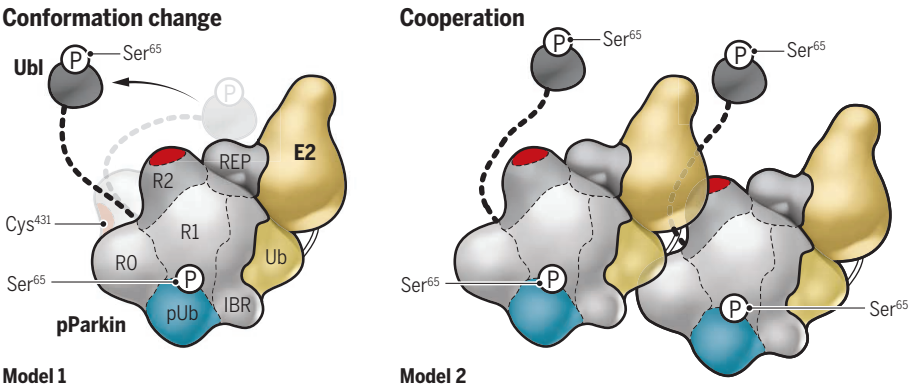
Multiple ARPD-associated proteins, such as Parkin, PINK1, and DJ-1 (also known as PARK7), play roles in mitochondrial homeostasis, suggesting that dysfunction in this process may be a feature of ARPD. PINK1 phosphorylates Parkin at Ser^{65} of the Ubl domain, which increases its ubiquitin ligase activity (*1*). In addition to directly phosphorylating Parkin (pParkin), PINK1 phosphorylates ubiquitin at the equivalent Ser^{65} (pUb), which leads to activation of Parkin (*1*). Allosteric binding of pUb to Parkin weakens Ubl association with the R0RBR domains in trans (*6*–*8*) and results in straightening of helix 3 within RING1 to facilitate movement of the IBR domain (*8*, *9*). Furthermore, PINK1 phosphorylates Parkin-pUb more efficiently than Parkin alone, suggesting that displacement of the Ubl is required for full kinase activity (*10*). In pParkin, the pUbl is released to form an extended conformation, and only in the presence of pUb (*11*). Point mutations that disrupt individual autoinhibitory features of Parkin show that upon activation with pUb, Parkin undergoes conformation changes that release both REP and the Ubl domain (*12*).

E2s transfer ubiquitin to Parkin, yet interaction between E2 and autoinhibited Parkin is weak (*6*). However, phosphorylation of Parkin and ubiquitin renders the E2 interaction measurable (*6*, *7*). Although the interaction of pParkin-pUb and E2s is low, it is increased when the E2 is loaded with ubiquitin (E2~Ub) (*7*). Finally, computational analyses predict that Ubl phosphorylation initiates an unwinding of domains (*13*). Thus, the prevailing view of PINK1-mediated Parkin activation is one where both phosphorylation events are needed to drive conformational

Institute of Molecular, Cell and Systems Biology, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK. Email: helen.walden@glasgow.ac.uk

Models of ubiquitin transfer to Parkin

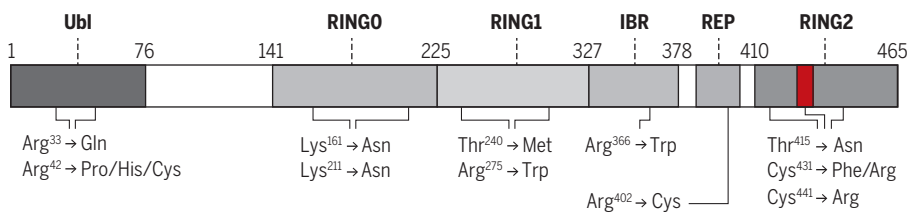
Structures of pParkin and the E2~Ub complex indicate that Cys^{431} of the RING2 (R2) domain is too far from the E2~Ub binding site for ubiquitin transfer to pParkin. Two models, involving conformational change or cooperation, have been suggested to explain how transfer might occur. P, phosphate.



Model 1
Conformational changes release the REP-Ubl and R2 domains of pParkin for E2~Ub binding and exposure of Cys^{431}.

Model 2
With the ubiquitin binding site revealed through conformation changes, this model illustrates how pParkin cooperates with another E2~Ub-pParkin complex to transfer ubiquitin from the E2.

ARPD-associated mutations in Parkin
The broad distribution of mutations across Parkin suggests the structural integrity of all domains is important in ARPD pathogenesis.



changes to release the Ubl domain, the REP, and potentially the RING0 domain to allow binding of E2~Ub and ubiquitin transfer. However, the structure of pUb bound to the R0RBR domains shows no movement of the REP (*8*). Furthermore, a recent crystal structure of all five human Parkin domains (Ubl-R0RBR) bound to pUb shows no displacement of the Ubl domain or the REP, nor alteration of the catalytic cysteine environment (*9*). One caveat is that the Ubl-RING0 linker is missing. This linker is not conserved in Parkin orthologs, but does influence in vitro activity (*7*). Furthermore, Parkin is not phosphorylated in this structure and there is no E2~Ub present. The straightening of RING1 helix 3 observed in the pUb-R0RBR structure is also observed in the presence of the Ubl domain (*8, 9*). In this context, however, the IBR is displaced, revealing a hidden ubiquitin binding site on RING1 helix 3 that is proposed to interact with the donor ubiquitin from E2~Ub. Thus, pUb and ubiquitin bind on opposing sides of the now-straightened RING1 helix 3, and mutation of either face leads to reduced affinity for pUb as well as E2~Ub (*13, 14*).

This bipartite binding of ubiquitin molecules is also observed in the crystal structure of another RBR ligase, HOIP, where ubiquitin binds in the equivalent position as pUb, and E2~Ub binding corresponds to the hidden ubiquitin binding site (*14*). Both of these structures show that the catalytic cysteine and the donor ubiquitin are still too far for ubiquitin transfer from E2 to Cys^{431} in cis. Does the E2-to-RING2 transfer of ubiquitin require further conformational changes of the RBR modules, or are other mechanisms possible? This conundrum is reminiscent of the allegory of the long spoons, where the inability to feed oneself with such unwieldy tools in Hell is circumvented in Heaven by feeding one another. Interestingly, self-association of E3 molecules has been reported to support Parkin activity both in vitro and in cells (*15*), thus raising the possibility of trans-cooperation for catalytic function (see the figure). Further support for this comes from the crystal structure of HOIP RBR-E2~Ub complex, where the RING2 of a neighboring molecule is in proximity to E2~Ub bound to a different RBR module (*14*). This "domain-swapped" dimer suggests the possibility of a trans-based mechanism. Intriguingly, HOIP functions biologically in concert with another RBR ligase, HOIL-1L (heme-oxidized IRP2 ubiquitin ligase 1 homolog), suggesting that other RBR ligases may function in collaboration, akin to two molecules of Parkin.

Although these recent studies raise the possibility of a cooperative model for Parkin function, there remain several important caveats. One is that a cooperative model does not explain E2 interaction as the REP and Ubl remain attached at the E2 binding site. In addition, the absence of the Ubl-RING0 linker, removed to aid high-resolution crystallography, necessarily reduces interdomain flexibility. Further, the lack of a structure of pParkin makes it challenging to fully understand the role of Parkin phosphorylation, as phosphorylation of the Ubl and pUb binding releases the Ubl domain and greatly enhances E2~Ub interaction. Finally, Parkin mutants can rescue each other by functioning in trans, but does wild-type Parkin cooperate for ubiquitin transfer? To distinguish these models, rescuing Parkin mutants deficient in one part of the transfer cascade with others deficient in another would help decipher at which stage cooperation can occur.

Currently, the data supporting each model do not exclude the other model, and therefore, both modes of ubiquitin transfer—through conformational change or through cooperative binding—could occur. Parkin is mainly cytoplasmic, yet PINK1 resides at the mitochondria. Does this mean that Parkin is only activated at the mitochondria, or are there alternative mechanisms dependent on subcellular localization? For example, ubiquitin-interacting motifs on Parkin substrates have been shown to increase Parkin activity (*3*). Therefore, Parkin may employ distinct mechanisms of activation in different cellular contexts, which could include structural rearrangement and/or cooperation between molecules. Distinguishing between these models will facilitate understanding how and when Parkin recognizes substrates, and how ARPD-associated mutations disrupt Parkin function. Characterizing Parkin's catalytic cycle will be important for targeting Parkin for therapeutic intervention. It will be fascinating to see what further conformations and mechanisms emerge as the understanding of the catalytic cycle of Parkin-mediated ubiquitin transfer deepens. ■

REFERENCES

1. N. Panicker *et al.*, *Biochem J.* **474**, 3075 (2017).
2. D. M. Wenzel *et al.*, *Nature* **474**, 105 (2011).
3. V. K. Chaugule *et al.*, *EMBO J.* **30**, 2853 (2011).
4. J.-F. Trempe *et al.*, *Science* **340**, 1451 (2013).
5. K. K. Dove, R. E. Klevit, *J. Mol. Biol.* **429**, 3363 (2017).
6. V. Sauvé *et al.*, *EMBO J.* **34**, 2492 (2015).
7. A. Kumar *et al.*, *EMBO J.* **34**, 2506 (2015).
8. T. Wauer *et al.*, *Nature* **524**, 370 (2015).
9. A. Kumar *et al.*, *Nat. Struct. Mol. Biol.* **24**, 475 (2017).
10. A. Kazlauskaite *et al.*, *EMBO Rep.* **16**, 939 (2015).
11. J. D. Aguirre *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 298 (2017).
12. M. Y. Tang *et al.*, *Nat. Commun.* **8**, 14697 (2017).
13. T. R. Caulfield *et al.*, *PLOS Comput. Biol.* **10**, e1003935 (2014).
14. B. C. Lechtenberg *et al.*, *Nature* **529**, 546 (2016).
15. M. Lazarou *et al.*, *J. Cell Biol.* **200**, 163 (2013).

10.1126/science.aar6606

GRAPHIC: C. BICKEL/*SCIENCE*

GENETICS

If two deletions don't stop growth, try three

Triple mutations in yeast reveal broad scope of genetic interactions

By **Albertha J. M. Walhout**

How many and which genes does an organism need at a minimum to sustain growth? With the identification of essential genes in a variety of organisms such as yeast (*1*), worms (*2*), and human cultured cells (*3*–*5*), the answer may seem obvious: Remove all nonessential genes, and the organism should be able to continue to survive and grow. In the yeast *Saccharomyces cerevisiae*, the workhorse of genetic interaction mapping, ~1000 of the ~6000 (~17%) protein-coding genes are essential for viability (*1*). So theoretically, one might think that one could remove the other ~5000 nonessential genes and still have a growing yeast. The reason this does not work is that genes interact with each other to drive different traits, including viability and growth. On page 283 of this issue, Kuzmin *et al.* (*6*) provide the first systematic analysis of trigenic interactions in yeast in which three gene deletions are combined in individual strains in order to ascertain the importance of multigene interactions for growth. This study shows that higher-order genetic interactions are abundant, which has implications for the interpretation of human gene variants that may interact to affect health and disease.

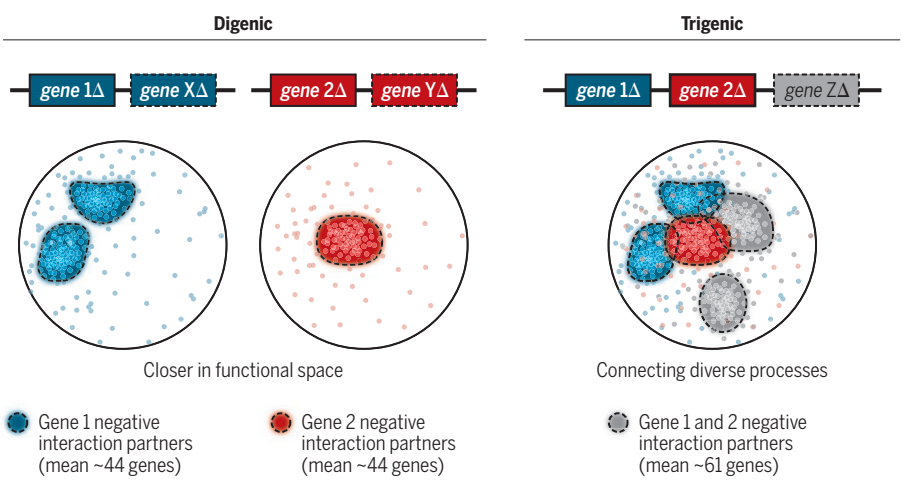
A comprehensive study of digenic interactions—that is, of interactions between pairs of deleted nonessential genes—defined 10,000 synthetic lethal interactions in yeast (*7*). Synthetic lethal interactions occur when deletion of each individual gene maintains viability and growth, whereas deletion of both genes combined kills the cells. These 10,000 synthetic lethal interactions among nonessential genes involve more than 3000 genes, thus revealing their "hidden essentiality." Synthetic lethal interactions are only one type of genetic interaction. Generally, genetic interactions are defined when a double mutant confers a phenotype that is different than expected based on the phenotype from mutants of each individual gene. Such interactions come in two basic flavors: positive and negative, where the phenotype of strains harboring positively interacting genes is less severe than expected, whereas the phenotype of strains containing negatively interacting gene pairs is more severe, with synthetic lethality being the most extreme. It is becoming clear that genetic interactions are widespread. For instance, ~550,000 (~3%) of all gene pairs [18 million, including hypomorphic (partial loss-of-function, temperature-sensitive) alleles of essential genes] display a negative genetic interaction in yeast (*7*). However, the prevalence and attributes of more complex genetic interactions remained poorly characterized.

Kuzmin *et al.* present the first large-scale trigenic interactions that affect colony growth. The growth of triple-deletion strains was compared with the growth of the strains harboring single or double deletions in the relevant genes. Systematically mapping trigenic interactions is a daunting endeavor even in yeast, with a total of 36 billion combinations, ~2000 times more than the number of gene pairs that can be tested for digenic interactions (*7*). In this study, a set of query strains carrying 302 single gene mutations and 151 double gene mutations were selected as a starting point. These query strains were tested for genetic interactions versus an array of 1182 strains, each carrying a mutation in an informative gene, so that most general biological processes were included (see the figure). Thus, in total close to 200,000 trigenic interactions (~0.0006% of all possible combinations) were tested. The focus in this study was on negative genetic interactions, which are easier to capture and quantify and usually more informative of gene function.

Overall, trigenic interactions were found to be ubiquitous. More than 3000 trigenic interactions were detected, versus almost 10,000 digenic interactions. When extrapolated to all possible pairs and triplets, it was estimated that 3% of all 18 million digenic pairs exhibit a negative genetic interaction, whereas a lower ratio was inferred for trigenic genetic interactions, which is estimated to be 100 million of all 36 billion possible triplets (~0.3%). Interestingly, about two-thirds of the trigenic negative interactions occur between genes in which at least one pair exhibits a digenic interaction. This indicates that genetic interactions themselves are heavily modified by genetic background. Further, trigenic interactions are overall

Complex gene interactions

Triple mutants in yeast reveal abundance of higher-order genetic interactions. Understanding these interactions has implications for the interpretation of human gene variants that may interact to affect health and disease.



	DIGENIC	TRIGENIC
Potential number of yeast strains	~18 million	~36 million
Negative genetic interactions	~550,000	~100 million

GRAPHIC: A. KITTERMAN/*SCIENCE*

Program in Systems Biology and Program in Molecular Medicine, University of Massachusetts Medical School, MA 01605, USA. Email: marian.walhout@umassmed.edu

weaker than digenic interactions and, like digenic interactions, are enriched for essential genes. This suggests diminishing returns for the mapping of even higher-order genetic interactions. However, trigenic interactions, although enriched among functionally related genes, tend to span across a wider distance in the overall functional network mapped previously (*7*) and may provide broader insights into the connections between bioprocesses at a higher level.

Aside from functionally annotating and connecting genes, this study has broad implications for the true understanding of the relationships between genotype and phenotype, which is highly relevant for understanding human traits, including diseases and the response to therapeutic treatments. The yeast study "only" uses deletion alleles in nonessential genes and hypomorphic alleles for essential genes, studies a single trait (growth), and examines one experimental condition. In the wild, organisms harbor many different alleles, from gain of function to partial or total loss of function or even neomorphic function (in which the allele confers a new function not associated with the wild-type gene), and in many different combinations. It is clear that many of these will interact in unexpected ways to affect every conceivable trait, including modifier alleles in genes that only exhibit an effect when combined with additional, severe mutations. For instance, inborn errors of human metabolism, which were classically viewed as Mendelian diseases, are increasingly appreciated to be responsive to genetic background and are thus complex (*8*). Given the complexity of multigenic traits, reading one's genome sequence to devise strategies for precision and personalized medicine may not be a feasible means of diagnosis and treatment for years to come. ■

"...this study has broad implications for the true understanding of the relationships between genotype and phenotype..."

REFERENCES

1. G. Giaever *et al.*, *Nature* **418**, 387 (2002).
2. R. S. Kamath *et al.*, *Nature* **421**, 231 (2003).
3. T. Hart *et al.*, *Cell* **163**, 1515 (2015).
4. T. Wang *et al.*, *Science* **350**, 1096 (2015).
5. V. A. Blomen *et al.*, *Science* **350**, 1092 (2015).
6. E. Kuzmin *et al.*, *Science* **360**, eaao1729 (2018).
7. M. Costanzo *et al.*, *Science* **353**, aaf1420 (2016).
8. C. A. Argmann, S. M. Houten, J. Zhu, E. E. Schadt, *Cell Metab.* **23**, 13 (2016).

10.1126/science.aat4667

HYPOTHESIS

A unifying concept in vascular health and disease

Interventions to restore blood vessel stability could improve health outcomes

By **Martin A. Schwartz,[1,2,3] Dietmar Vestweber,[4] Michael Simons[1,2]**

Not unlike Tolstoy's remark about happy versus unhappy families, current wisdom in vascular biology holds that healthy blood vessels are mostly similar, whereas vessels in different vascular diseases are mostly different. But is this really the case? An evaluation of the literature suggests that unresolved vascular remodeling may be a key element of virtually all vascular diseases. This commonality raises the possibility of unifying principles that govern vascular remodeling and the possibility that methods to restore normal remodeling could effectively treat multiple disease states.

Postnatal physiological vascular remodeling encompasses the sprouting of existing blood vessels to form new ones (angiogenesis), and the growth and remodeling of small vessels into larger arteries that support higher blood flow (arteriogenesis). Importantly, both processes are self-limited. Angiogenesis is stimulated by tissue demand for oxygen and nutrients due to tissue growth, exercise, or wound healing (*1*). Secretion of vascular endothelial growth factor (VEGF) by cells in ischemic tissue is the main driver, but once blood flow is established in the new vessels, oxygen and nutrient transport into the tissue decreases VEGF expression, shuts down angiogenesis, and restores vascular stability. Arteriogenesis is triggered mainly by increased blood flow and hence fluid shear stress through small vessels. This induces a local inflammatory response, leading to VEGF production (*2*). The resultant outward remodeling, vessel enlargement, and de novo formation of arterial vasculature restore normal levels of shear stress, which resolves vascular wall inflammation and restores normal function. These two processes are of relatively short duration, on the order of days to weeks in mouse models, and fully restore normalcy.

By contrast, prolonged or chronic stimulation of angiogenesis or arteriogenesis results in pathological remodeling and leads to morphologically and functionally abnormal vessels. One example is atherosclerotic plaques that form as a result of converging biomechanical (disturbed blood flow and high vessel wall stress), metabolic (hyperlipidemia, hyperglycemia), and inflammatory and thrombotic factors (*3, 4*). Plaque formation appears to be driven by altered endothelial cell (EC) phenotype, including increased permeability that allows lipoproteins to enter the vessel wall, increased expression of proteins that recruit monocytes, and other immune cells; activated ECs also stimulate smooth muscle cell migration and growth to form a neointima that narrows the vessel lumen. Further, vascular wall inflammation increases sensitivity of ECs to transforming growth factor–β (TGF-β), leading to endothelial-mesenchymal transition (EndMT) (*5*). The resultant EC-derived mesenchymal cells drive further increases in permeability and inflammation, thus perpetuating the incomplete repair state.

A similar process occurs in vascular malformations such as cerebral cavernous malformation (CCM) and hereditary hemorrhagic telangiectasia (HHT) (*6, 7*). In the inherited form of these diseases, patients have a heterozygous loss-of-function mutation in a key gene (*CCM1*, *2*, or *3* for CCM; *Alk1*, *endoglin*, or *Smad4* for HHT). A "second hit" mutation plus angiogenic or inflammatory stimuli lead to clonal expansion and formation of lesions with poorly formed vessel walls that are prone to rupture and bleeding. Again, the pattern of elevated permeability, abnormal extracellular matrix (ECM), inflammation, and EndMT can form positive feedback circuits that contribute to lesion progression. Yet another example is chronic inflammation due to persistent infection, injury, or autoimmune processes. In these conditions, angiogenic blood vessels remain immature with increased permeability, ECM remodeling, and expression of genes for leukocyte recruitment, thus perpetuating inflammation (*8*).

There are several well-conserved features that distinguish stable from remodeling vas-

[1]*Yale Cardiovascular Research Center, 300 George Street, New Haven, CT 06511, USA.* [2]*Departments of Medicine (Cardiology) and Cell Biology, Yale School of Medicine, New Haven, CT, USA.* [3]*Department of Biomedical Engineering, Yale University, New Haven, CT, USA.* [4]*Max Planck Institute for Molecular Biomedicine, Münster, Germany.*
Email: martin.schwartz@yale.edu

culature. In stable vessels, ECs rest on a basement membrane of mainly laminins and type IV collagen with associated glycoproteins but low amounts of provisional ECM proteins such as fibronectin and fibrin (*4*). By contrast, all forms of instability are associated with metalloprotease secretion, ECM degradation, assembly of provisional matrices, and alterations in key EC signaling and gene expression pathways. Stable vessels produce high amounts of angiopoietin 1, which binds its EC receptor Tie2 to stabilize the vessels; expression of angiopoietin 2 (Ang2), an antagonist of Tie2, is low (*9*). Junctional permeability is low in stable vessels and elevated during inflammation and remodeling (*8*). This rule has exceptions because fenestrated endothelium allows passage of fluids, and the high endothelial venules that facilitate leukocyte movement into tissues during immune surveillance allow passage of cells; however, these forms of permeability have distinct mechanisms. All forms of remodeling are also associated with a shift toward a more oxidative state, or in the case of pathologies, to oxidative stress (*10*).

Whereas stable vessels exhibit normal EC fate markers, vascular instability is associated with EndMT, in which endothelial markers are reduced and mesenchymal markers increase (*11*). Functional consequences include increased permeability, increased expression of leukocyte adhesion molecules, enhanced motility, and expression and deposition of provisional ECM. EndMT has been observed in physiological and pathological angiogenesis, and in atherosclerosis, chronic inflammation, and vascular malformations.

How do physiological and pathological remodeling differ? In physiological remodeling, the initiating stimulus disappears once adaptation is completed, leading to restoration of normalcy. By contrast, in pathological remodeling, the initial triggers generally persist. Even when initiating triggers are removed, the vasculature does not return to its normal state. Such persistence involves positive feedback loops that maintain the endothelium in the remodeling or disease state, leading to disease progression. Several examples illustrate this point.

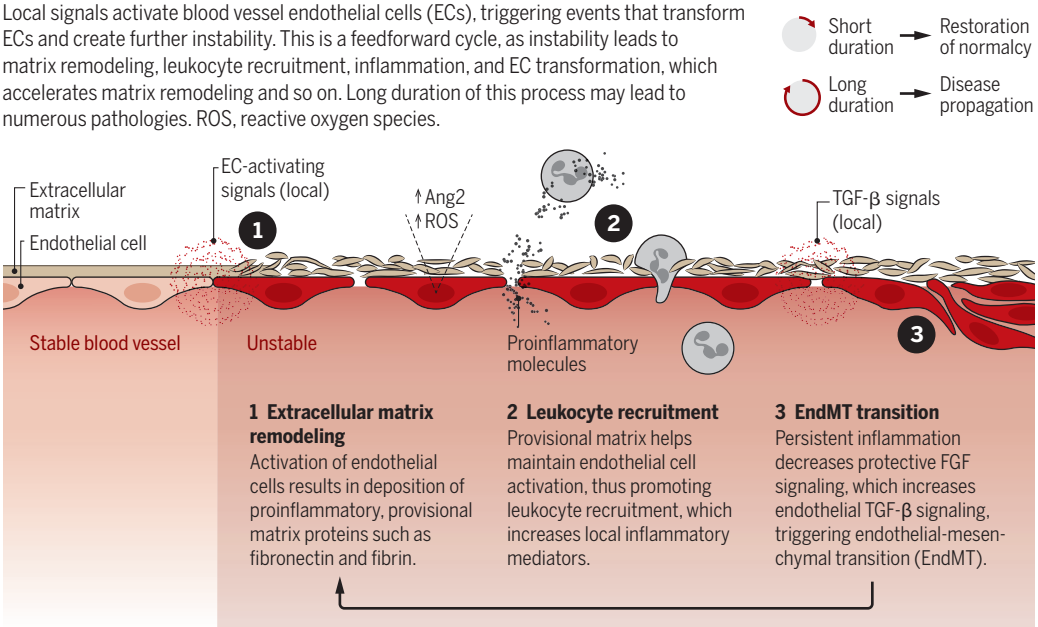
In atherosclerosis, statins slow progression but do not stop or reverse the disease even when plasma lipids are drastically lowered; ECs overlying atherosclerotic plaques remain activated. Some positive feedback loops involve increased vascular permeability and expression of leukocyte adhesion molecules. Entry of circulating ECM proteins such as fibronectin and fibrinogen into the vessel wall contributes to deposition of proinflammatory provisional ECM. Leaky endothelium also allows entry of lipoproteins into inflamed vessel wall under oxidative stress, resulting in production of proinflammatory oxidized lipoproteins that further increase lipid uptake and inflammatory mediators. Activation of adaptive immune responses to damaged proteins within the plaque and enhanced entrance of circulating inflammatory cells due to increased expression of adhesion molecules accelerate disease (*3*). Similar processes take place in transplant arteriopathy, characterized by low-grade, persistent vascular inflammation and incomplete repair (*12*). CCMs also exhibit this constellation of increased permeability, altered ECM, inflammatory activation, and EndMT that likely mediate disease progression (*7*).

These findings imply that persistent instability is central to multiple vascular disorders (see the figure). EndMT is likely an important component due to conversion of ECs to migratory mesenchymal cells that cause irreversible changes in tissue architecture. EndMT is induced by inflammation-driven decreases in protective fibroblast growth factor (FGF) signaling in ECs, which increases sensitivity to TGF-β (*11*). Thus transformed, ECs increase expression of leukocyte adhesion molecules and altered ECM, and have high vascular permeability, further driving inflammation. These events establish a positive feedback loop in which EndMT drives inflammation, which leads to further EndMT. This inflammation–TGF-β circuit also promotes fibrosis, which is largely irreversible.

This hypothesis implies that interventions to restore vascular stability might improve outcomes in multiple disease states. This might be achieved by agonists for Tie2, FGF receptors, or other pathways that promote stability, or by antagonists for inflammatory or endothelial TGF-β pathways that promote instability. The key requirement is to break the feedback loops that maintain the remodeling state and thereby restore tissue homeostasis. Although this notion flies in the face of current trends toward personalized medicine, the search for unifying principles that govern health and disease is as old as medicine. Much remains to be done before we achieve the deep understanding of biological processes needed to design effective interventions in vascular disease, yet our current state of knowledge offers a glimpse of the path forward. ■

Persistent instability and disease

Local signals activate blood vessel endothelial cells (ECs), triggering events that transform ECs and create further instability. This is a feedforward cycle, as instability leads to matrix remodeling, leukocyte recruitment, inflammation, and EC transformation, which accelerates matrix remodeling and so on. Long duration of this process may lead to numerous pathologies. ROS, reactive oxygen species.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES

1. R. H. Adams, K. Alitalo, *Nat. Rev. Mol. Cell Biol.* **8**, 464 (2007).
2. M. Simons, A. Eichmann, *Circ. Res.* **116**, 1712 (2015).
3. I. Tabas *et al.*, *J. Cell Biol.* **209**, 13 (2015).
4. A. Yurdagul Jr. *et al.*, *Biochem. J.* **473**, 1281 (2016).
5. P.-Y. Chen *et al.*, *J. Clin. Invest.* **125**, 4514 (2015).
6. S. Tual-Chalot *et al.*, *Front. Genet.* **6**, 25 (2015).
7. M. G. Lampugnani *et al.*, *Curr. Opin. Hematol.* **24**, 256 (2017).
8. J. A. Nagy *et al.*, *Angiogenesis* **11**, 109 (2008).
9. S. M. Parikh, *Curr. Opin. Hematol.* **24**, 432 (2017).
10. F. Santilli *et al.*, *Vascul. Pharmacol.* **74**, 23 (2015).
11. E. Dejana *et al.*, *Nat. Commun.* **8**, 14361 (2017).
12. R. N. Mitchell, *Cardiovasc. Pathol.* **13**, 33 (2004).

10.1126/science.aat3470

BOOKS *et al.*

The BICEP2 telescope (left) was deployed to search for evidence of cosmic inflation in 2009.

COSMOLOGY

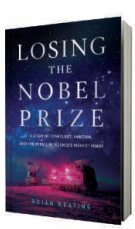
Reconsidering the Nobel Prize

After dust stymies a quest to confirm cosmic inflation, a physicist questions science's most prestigious award

By **Paul Halpern**

In March 2014, the BICEP2 (Background Imaging of Cosmic Extragalactic Polarization) collaboration, a team of astrophysicists mapping the cosmic background radiation left over from the Big Bang, excitedly announced an unexpectedly strong signal in its data. The researchers reported apparent evidence that primordial gravitational waves from a hypothesized rapid stretching of space, called the inflationary era, skewed the radiation's profile in a particular way.

Although ordinary light is unpolarized, twisting equally in clockwise and counterclockwise directions—a bit like a balanced collection of left-handed and right-handed screws—BICEP2 revealed distinct polarization patterns that beautifully matched some of inflation theory's predictions. Months later, however, hopes were dashed when further analysis pointed to a more prosaic cause for the signal: interstellar dust.



Losing the Nobel Prize
A Story of Cosmology, Ambition, and the Perils of Science's Highest Honor
Brian Keating
Norton, 2018. 350 pp.

Losing the Nobel Prize, by astrophysicist Brian Keating of the University of California, San Diego, offers a riveting account of the rise and fall of the seeming confirmation of the cosmological theory of inflation. Keating, who played a major role in the BICEP2 project and in related previous endeavors, wonderfully captures the drama behind major (and expensive) cosmological experiments that have the potential to revolutionize science and bolster careers. He nicely explains the impetus behind the inflationary-era hypothesis and describes why many researchers are eager to confirm it.

Keating offers vivid profiles of the personalities involved in shaping our modern view of the universe, including the pioneers of inflation. One of those figures, Princeton cosmologist Paul Steinhardt, receives special attention in the book for his controversial decision to abandon the inflation approach, which he helped mold, and his subsequent efforts to craft a competing model, called the cyclic universe. Keating cogently explains the philosophy behind Steinhardt's decision, placing it in the context of earlier visions, such as the steady-state cosmological model.

Keating describes how the BICEP2 signal originally seemed to favor inflation and imperil cyclic models. As a bonus, it seemed to provide a way of probing gravitational waves at the quantum level, potentially revealing new physics. He well captures the joy felt by inflationary theorists, such as Andrei Linde, when their models seemed proved at last and their grave disappointment when the purported evidence vanished into dust.

Occluding Keating's account, however, are his personal beefs about "losing" the Nobel Prize. Unlike the Olympics or marathon races, the Nobel Prize is not a competition between vying players who train for a shot at a trophy. These days, physics has so many researchers and subfields that it is not always clear who and which are in contention.

Normally, scientists don't base their careers on what will snag them the "Nobel gold," but rather on a passion for scientific inquiry and the enjoyment that comes from learning more about the universe. It is therefore jolting to read passages in which he recalls calculating what he needed to do to maximize his chances of being a laureate and in which he recounts slights from his colleagues that threatened to block his way.

Such gripes taint Keating's otherwise thoughtful analysis of ways the Nobel committee's selection process could be improved. That is a shame, because many of his ideas, considered dispassionately, are quite worthy.

Keating suggests, for example, opening up the physics prize to larger groups than the current maximum of three—a change that certainly makes sense given the sizable collaborations in today's megaprojects. He also urges establishing new prizes for emerging scientific disciplines and correcting past omissions. He supports allowing posthumous awards, within reason, such as to those who die right before their colleagues are honored for the same result. Finally, he suggests awarding more prizes for serendipitous findings, rather than through confirmation.

Losing the Nobel offers an important lesson that not everything in science goes according to plan. There are often many false starts before progress is made. Researchers can spend their careers planning and carrying out experiments that lead nowhere.

Yet true scientific satisfaction comes with the journey, not just the destination. Even if the Nobel Prize vanished tomorrow, we are curious beings and would undoubtedly carry on extending the limits of knowledge. Losing the Nobel would be a shame, but hardly the end of science. ■

The reviewer is in the Department of Mathematics, Physics, and Statistics at the University of the Sciences, Philadelphia, PA 19104, USA. Email: p.halper@usciences.edu

10.1126/science.aat0374

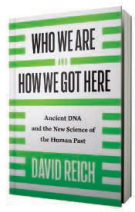
ANCIENT DNA

Wielding new genomic tools wisely

Troubling traces of biocolonialism undermine an otherwise eloquent synthesis of ancient genome research

By **María C. Ávila Arcos**

Who We Are and How We Got Here
Ancient DNA and the New Science of the Human Past
David Reich
Pantheon, 2018. 362 pp.



In *Who We Are and How We Got Here,* David Reich gracefully describes how recent advances in genomics have enabled the study of ancient genomes and how this, in turn, has significantly affected the study of the evolutionary and demographic history of our species. With a pleasant narrative style that immediately engages both scientists and nonscientists, the book describes the technological and statistical advances that have allowed researchers to read the genomes of humans, past and present.

The author is in a privileged position to convey the progress and major milestones achieved in the burgeoning field of ancient genomics because his research group has been a main player in this arena since the beginning. This means, of course, that the book disproportionately highlights work from his lab, as Reich himself freely acknowledges. (Much of this progress has been made possible by a custom DNA capture kit that, as of now, cannot simply be ordered by other researchers. Access to such kits, at a reasonable cost, would greatly enhance the ability of other groups to produce methodologically consistent data.)

Combining personal views and "hard science," *Who We Are and How We Got Here* takes the reader on a journey from the dawn of our species to more recent events that have shaped our current population structure, all in the context of the knowledge generated by comparing the DNA of ancient and modern human populations. Throughout the book, Reich illustrates how continuous gene flow is the norm, not the exception, in the history of our species. "Mixture is fundamental to who we are," he writes. "[W]e need to embrace it, not deny it occurred."

In our quest to understand human origins, we must remain vigilant with regard to the ethical implications of such research.

The book eloquently addresses what ancient DNA has revealed about these mixing events, from the interbreeding between our ancestors and the Neandertals and Denisovans 50,000 to 60,000 years ago to the effects of European colonization of the Americas, which gave rise to admixed African American and Latino populations. It describes the migrations and mixing events that formed the population structure observed in Europe, India, and East Asia over the course of 10,000 years and explains how some current genetic variation can only be explained by past mixing with yet uncharacterized ancient populations, which Reich refers to as "ghost" populations.

Reich also proposes a controversial theory in which he suggests that modern human ancestors actually may have lived in Eurasia, not in Africa, calling Eurasia a "hothouse of human evolution." According to this theory, ancestral humans, including Neandertals and Denisovans, descended from the African *Homo erectus*, left Africa for Eurasia, and later returned to serve as the founders of what would eventually evolve into modern humans. This provocative proposal is sure to generate debate.

Although scientifically solid and comprehensive, the book raises some concerns. Even though Reich devotes a substantial part of the book to discussing the implications of ancient DNA research, he overlooks some of the most profound issues inherent in this type of work, specifically with regard to the practices surrounding the collection and processing of human DNA samples.

Early in the book, for example, he writes that his goal is to build "an American-style genomics factory" to investigate ancient DNA. When one considers the social and historical context of the human populations that will be studied—many of which have been historically marginalized, colonized, and exploited—this statement becomes problematic. Such intentions could easily be perceived as a continuation of exploitation or biocolonialism.

An unfortunate analogy further highlights the problem. "We are … like explorers in the late eighteenth century, sailing to every corner of the globe," he writes. During the era to which Reich refers, European adventurers indeed collected samples from around the world, but these specimens were usually taken without the consent of, or regard for, the communities to whom they rightfully belonged.

On several occasions throughout the book, Reich reveals his frustration with policies and legislation that prohibit the exportation of samples or that restrict the study of human remains, but he fails to entertain the idea that such scenarios could represent opportunities to train local scientists and members of the communities from whence the samples came to conduct their own studies. Building local capacities would allow this kind of research to be truly democratic.

There is no doubt that Reich has changed the game when it comes to our understanding of human history. His book, however, misses critical opportunities to highlight some of the major ethical issues surrounding human genomes and ancient DNA research. ■

The reviewer is at the International Laboratory for Human Genome Research, Juriquilla, Querétaro, Mexico. Email: mavila@liigh.unam.mx

10.1126/science.aat3780

PHOTO: JOE MCNALLY PHOTOGRAPHY 2018

Corrected 20 April 2018. See full text.

Elephant tusks in Kenya.

PHOTO: ALISSA EVERETT/ALAMY STOCK PHOTO

Edited by **Jennifer Sills**

Ivory crisis: Growing no-trade consensus

In their Perspective, "Breaking the deadlock on ivory" (15 December 2017, p. 1378), D. Biggs *et al.* propose steps to enhance unity around the African elephant poaching crisis. We support their recommendations for dialogue among African elephant range states. However, the Perspective misrepresents the evidence-driven rationale of the no-trade approach to ivory, promotes a counterproductive geographically divided approach to wildlife trade, and understates the growing worldwide policy consensus to end ivory trade.

By asserting that a no-trade approach is motivated by "sacred" values and misidentifying animal rights as central to this position, Biggs *et al.* imply that the no-trade approach is not pragmatic. In fact, the no-trade position and the pro-trade position differ not in core values or objectives, but in interpretations of evidence on the relative usefulness of improved governance, markets, and sociocultural change in addressing poaching. It would be ideal if only ivory from naturally deceased elephants could be used to fund conservation sustainably. However, the evidence suggests that this cannot be practically achieved for elephants. Economic models supporting ivory sales ignore elephants' low population and productivity (*1*). Thus, new demand for ivory will likely outpace new legal supply, increase black market prices (*2*), and further incentivize elephant poaching in countries struggling to patrol vast areas (*3*). Legal trade also makes it more difficult to detect contraband (*4*) and fails to address the escalating levels of criminality driving most ivory shipments over the past decade (*5*). A one-time legal ivory sale to China and Japan permitted by the Convention on International Trade in Endangered Species (CITES) in 2008 corresponded with an abrupt increase in poaching (*2*). In contrast, the 1989 ban on international ivory trade and decisions to restrict legal domestic trade from 2015 onward were each accompanied by at least a halving of the price of ivory (*6*, *7*).

Elephant conservation would suffer under Biggs *et al.*'s proposal for further regional differentiation of ivory trade policies. Any legal trade in ivory undermines efforts to reduce elephant poaching everywhere (*2*). All 37 African elephant range states have expressed shared conservation objectives in the 2010 African Elephant Action Plan (*8*) and should be regarded as equal stakeholders. That 76% of African elephants live in transboundary populations (*9*) necessitates cooperation between neighbors and continent-wide management approaches. These approaches could include identifying revenue sources to replace ivory sales for pro-trade countries, as Biggs *et al.* suggest. CITES' legitimacy would be undermined by devolving its authority or making decisions "outside of the public's view," as proposed by Biggs *et al.* Decreasing public scrutiny during negotiations could increase vulnerability to commercial interests or assertive governments focused on short-term benefits (*10*).

In contrast to the "deadlock" portrayed by Biggs *et al.*, a global consensus is growing for a complete ban on trade in ivory to combat elephant poaching. Biggs *et al.* themselves recognized current near-total domestic bans on ivory trade (in the United States, China, and the United Kingdom) and the motion to stop all legal domestic sales adopted at the 2016 IUCN World Conservation Congress. Additionally, since 2010, Parties to CITES have dismissed proposals for sales of stockpiled ivory and rejected a decision-making mechanism that could reestablish trade (*11*).

Instead of perpetuating demand for ivory through sales, we suggest that demand be minimized through a combination of regulatory instruments (domestic trade bans) and sociocultural interventions (behavior change campaigns). Other strategies include dismantling supply chains using intelligence-driven law enforcement; strengthening judicial systems; and encouraging cross-border cooperation, human-elephant coexistence projects, and alternative economic opportunities for poachers and traders. Combining nature-compatible livelihoods with strengthened revenue streams from elephant-oriented tourism could—if well-governed—promote equitable development and participatory conservation across rural Africa (*12*).

Nitin Sekar,[1]* William Clark,[2] Andrew Dobson,[1] Paula Cristina Francisco Coelho,[3] Phillip M. Hannam,[1] Robert Hepworth,[4] Solomon Hsiang,[5,6] Paula Kahumbu,[7] Phyllis C. Lee,[8,9] Keith Lindsay,[9] Carlos Lopes Pereira,[10] Samuel K. Wasser,[11,12] Katarzyna Nowak[13,14]

[1]Princeton University, Princeton, NJ 08544, USA. [2]Great Falls, VA 22066, USA. [3]Ministry of Environment, Luanda, Angola. [4]David Shepherd Wildlife Foundation, Shalford, Guildford, Surrey, GU4 8JU, UK. [5]Global Policy Laboratory, Goldman School of Public Policy, University of California, Berkeley, Berkeley, CA 94720, USA. [6]National Bureau of Economic Research, Cambridge, MA 02138, USA. [7]WildlifeDirect, Nairobi, Kenya. [8]Behaviour and Evolution Research Group, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK. [9]Amboseli Trust for Elephants, Nairobi, Kenya. [10]National Administration of Conservation Areas, Ministry of Land, Environment and Rural Development, Maputo, Mozambique. [11]Center for Conservation Biology, University of Washington, Seattle, WA 91895, USA. [12]Conservation Science Group, University of Cambridge, Cambridge, CB2 3QZ, UK. [13]Department of Zoology and Entomology, University of the Free State, Phuthaditjhaba, 9866, South Africa. [14]The Safina Center, Setauket, NY 11733, USA.
*Corresponding author. Email: nitin.sekar@gmail.com

REFERENCES

1. D. Lusseau, P. C. Lee, *Curr. Biol.* **26**, 2951 (2016).
2. S. M. Hsiang, N. Sekar, "Does legalization reduce black-market activity? Evidence from a global ivory experiment and elephant poaching data," *National Bureau of Economic Research* (Working Paper no. 22314, 2016).
3. S. K. Wasser *et al.*, *Science*, **349**, 84 (2015).
4. Environmental Investigation Agency, "Vanishing Point" (2014); https://eia-international.org/wp-content/uploads/EIA-Vanishing-Point-lo-res1.pdf.
5. C. May, "Transnational crime and the developing world," Global Financial Integrity (2017); www.gfintegrity.org/report/transnational-crime-and-the-developing-world/.
6. M. O'Connell, M. Sutton, "The effects of trade moratoria on international commerce in African elephant ivory: A preliminary report," TRAFFIC USA (WWF and the Conservation Foundation, 1990).
7. L. Vigne, E. Martin, "Decline in the legal ivory trade in China in anticipation of a ban" (Save the Elephants, 2017); www.savetheelephants.org/wp-content/

uploads/2017/03/2017_Decline-in-legal-Ivory-trade-China.pdf.
8. CITES, "African elephant action plan" (2010); www.iucn.org/sites/dev/files/import/downloads/e15i_68.pdf.
9. K. Lindsay *et al.*, *Biol. Conserv.* **215**, 260 (2017).
10. A. Gupta, *Glob. Environ. Polit.* **8**, 1 (2008).
11. S. Korwin, K. Lindsay, R. Reeve, *Pachyderm* **58**, 140 (2017).
12. R. Naidoo *et al.*, *Nat. Commun.* **7**, 13379 (2016).

10.1126/science.aat1105

Ivory crisis: Role of bioprinting technology

We agree with D. Biggs *et al.* ("Breaking the deadlock on ivory," Perspectives, 15 December 2017, p. 1378) that to prevent the extinction of elephants, we must recognize that different values influence stakeholders' perspectives. Poaching is increasingly driven by demand from China, with its growing number of wealthy consumers and investors and its traditions of ivory usage (*1*). Moreover, ivory prices have increased considerably since 2000 (*1*–*3*), indicating that it is a good investment. Altering consumer preferences alone through changing mental models, as Biggs *et al.* propose, is unlikely to reduce the demand for ivory and prevent the killing of elephants. Policy-makers must take a broader systems-thinking approach, whereby they consider not only how people value ivory psychologically but also how technology can be used to influence underlying economic demand and supply levers of the ivory trade, and ultimately its price.

Developments in three-dimensional (3D) bioprinting have made it possible to produce an indistinguishable substitute of elephant ivory and rhino horn. Using a small sample of tissue, the machine can replicate the species' DNA precisely in the printed version (*4*–*6*). Bioprinting offers several potential options for substantially reducing the market price for ivory. For example, if large volumes of bioprinted ivory were successfully introduced into African ivory markets (mixed, without detection, with the genuine ivory), then this ivory would likely pass further through the supply chain and into the black markets of Asia. A recent experiment shows that such an intervention is indeed possible (*7*). Mixing substantial volumes of bioprinted ivory into the supply chain, would not only increase supply and reduce the price, but also create information uncertainty among investors as to whether they are buying genuine ivory. This strategy has been shown to be effective in combating the shark fin trade (*8*).

Magdalena Lenda,[1,2,3]* Piotr Skórka,[2] Błażej Mazur,[4] Adrian Ward,[1,3] Kerrie Wilson[1,3]
[1]School of Biological Sciences, The University of Queensland, Brisbane, QLD, 4072, Australia. [2]Institute of Nature Conservation, Polish Academy of Sciences, Mickiewicza 33, 31-120 Kraków, Poland. [3]ARC Centre of Excellence for Environmental Decisions, The University of Queensland, St. Lucia, Brisbane, QLD, 4072, Australia. [4]Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland. *Corresponding author. Email: m.lenda@uq.edu.au

REFERENCES
1. Y. Gao *et al.*, *Biol. Cons.* **180**, 23 (2014).
2. G. Wittemyer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13117 (2014).
3. M. 't Sas-Rolfes, B. Moyle, D. Stiles, *Pachyderm* **55**, 62 (2014).
4. A. Klarmann, "3D printed rhino horns to deter ivory poaching?," All3DP (2017); https://all3dp.com/3d-printed-rhino-horns-poaching/.
5. S. V. Murphy, A. Atala, *Nat. Biotechnol.* **32**, 773 (2015).
6. Z. Corbyn, "Can we save the rhino from poachers with a 3D printer?," *The Guardian* (2015); www.theguardian.com/environment/2015/may/24/artificial-3d-printed-fake-rhino-horn-poaching.
7. B. Christy, "How killing elephants finances terror in Africa," *National Geographic* (2015); www.nationalgeographic.com/tracking-ivory/article.html.
8. M. Fabinyi, N. Liu, *Cons. Soc.* **12**, 218 (2014).

10.1126/science.aat0925

Response

Sekar *et al.* argue that there is unequivocal evidence that ivory trade bans are necessary for conserving elephants, and that a growing consensus removes the need to consider or incorporate alternative values in this debate. In doing so, they overlook relevant literature [e.g., (*1*–*3*)] and do not account for marginalized voices from key range states (*4*). Their response illustrates why the current impasse is unlikely to be resolved without a new structured process, underpinned by recognition that interpretation of scientific information on both sides of any contentious debate is influenced by values (*5*, *6*).

Sekar *et al.* describe apparently compelling evidence for the consequences of different policies relating to ivory, but there is much literature that contests their conclusions (*1*–*3*, *7*). For example, before the 17th Conference of the Parties (CoP), a technical advisory group of the Convention on International Trade in Endangered Species (CITES) took the unusual step of issuing a formal statement about the methodological shortcomings of one of the key working papers Sekar *et al.* cite (*3*, *8*). Sekar *et al.* also cite the motion adopted at IUCN's 2016 World Conservation Congress to prohibit legal domestic ivory sales as illustrating consensus. However, this debate was so adversarial that a diverse group of 30 prominent individuals, spanning 20 countries and including people from eight nongovernmental organizations and seven governments, publicly highlighted its pitfalls and urged for this approach to be avoided in the future (*9*).

This lack of consensus explains why different range states, all of which are committed to conserving elephants and all with access to the same data sets, continue to take opposing positions, illustrating how interpretation of evidence often reflects underlying assumptions, value systems, and mental models (*10*, *11*). Our paper acknowledged that there is policy momentum toward a trade ban, but called for a new process because polarized debates persist between and among range states and researchers. If this polarization continues, it will undermine policy implementation.

Despite Sekar *et al.*'s claim, our proposals would not undermine the legitimacy of CITES. Instead, we highlighted that experience from other contentious issues, such as negotiating climate change policy and the end to armed conflict, shows that progress is more likely through iterations of discussion in small groups by key stakeholders, rather than in adversarial public environments such as CITES CoPs (*12*, *13*). Such an approach could be facilitated by CITES and feed into CITES processes, as happened with the African Elephant Range State Dialogues (*14*). We also reiterate that range states, which are the ultimate custodians of Africa's elephants, should own and lead this process to develop policies that navigate the trade-offs their societies face.

We did not advocate a particular policy position, but instead called for a structured process to overcome the barriers to evidence-based decision-making. This should account for the different values and mental models that influence this debate, building consensus on how the available evidence is interpreted and what research is needed to tackle uncertainties and data gaps.

We agree with Sekar *et al.* that any process to build consensus must incorporate the need for sustainable financing to protect elephants from poaching and other threats like habitat loss. Such financing also needs to provide economic benefits to communities that live with elephants. Critically, to strengthen sustainability, policy processes on ivory must give more of a voice to those responsible for and affected by policy decisions than to those who suffer none of the costs of living with elephants. Overcoming this long-standing deadlock requires a new approach; conservation can learn from successes in other polarized debates to achieve lasting positive outcomes for elephants and other iconic taxa threatened by illegal trade.

We agree with Lenda *et al.* that synthetic ivory might provide new solutions, but it could also have unintended negative consequences by changing the nature and size of the ivory trade in unpredictable ways. There is already an effort within CITES to address the issue of synthetically produced wildlife products (*15*), which would benefit from

adopting the structured process we propose.

Duan Biggs,[1]* Robert J. Smith,[2] Vanessa M. Adams,[3] Henry Brink,[2] Carly N. Cook,[4] Rosie Cooney,[5] Matthew H. Holden,[6] Martine Maron,[6] Jacob Phelps,[7] Hugh P. Possingham,[8] Kent H. Redford,[9] Robert J. Scholes,[10] William J. Sutherland,[11] Fiona M. Underwood,[12] E. J. Milner-Gulland[13]

[1]Griffith University, Nathan, QLD 4111, Australia. [2]University of Kent, Canterbury, CT2 7NR, UK. [3]Macquarie University, Sydney, NSW 2109, Australia. [4]Monash University, Clayton, VIC 3800, Australia. [5]IUCN SULi, 1196, Gland, Switzerland. [6]University of Queensland, St. Lucia, QLD, 4072, Australia. [7]Lancaster University, Lancaster, UK. [8]The Nature Conservancy, Arlington, VA 22203, USA. [9]Archipelago Consulting, Portland, ME 04112, USA. [10]University of Witwatersrand, Johannesburg, South Africa. [11]University of Cambridge, Cambridge, CB2 3EJ, UK. [12]Taunton, UK. [13]Oxford University, Oxford, OX2 6GG, UK. *Corresponding author. Email: ancientantwren@gmail.com

REFERENCES

1. M. 't Sas-Rolfes, B. Moyle, D. Stiles, *Pachyderm* **55**, 62 (2014).
2. R. B. Martin *et al.*, "Decision-making mechanisms and necessary conditions for a future trade in African elephant ivory" (CITES, 2012).
3. CITES, "A statement from the MIKE and ETIS Technical Advisory Group on recent claims that the CITES-approved ivory sales in 2008 caused a spike in poaching levels" (2017); https://cites.org/sites/default/files/eng/cop/17/InfDocs/E-CoP17-Inf-42.pdf.
4. R. Duffy, *Environ. Planning C Gov. Pol.* **31**, 222 (2013).
5. D. Kahan, *Nature* **463**, 296 (2010).
6. D. M. Kahan *et al.*, *Nat. Clim. Change* **2**, 732 (2012).
7. B. Moyle, *Ecol. Econ.* **107**, 259 (2014).
8. S. M. Hsiang, N. Sekar, "Does legalization reduce black-market activity? Evidence from a global ivory experiment and elephant poaching data," National Bureau of Economic Research (Working Paper no. 22314, 2016).
9. S. N. Stuart *et al.*, *Oryx*, 10.1017/S0030605317001557 (2017).
10. CITES, "Closure of domestic markets for elephant ivory" (COP 17 Doc 57.2, 2016); https://cites.org/sites/default/files/eng/cop/17/WorkingDocs/E-CoP17-57-02.pdf.
11. CITES, "Consideration of proposals for amendment of Appendices 1 and 2" (2016); https://cites.org/sites/default/files/eng/cop/17/E-CoP17-Prop-14.pdf.
12. A. Kahane, K. Van Der Heijden, *Transformative Scenario Planning: Working Together to Change the Future* (Berrett-Koehler Publishers, 2012).
13. D. H. Cole, *Nat. Clim. Change* **5**, 114 (2015).
14. CITES, "African range States agree a restricted ivory trade from four southern African countries" (2002); https://cites.org/eng/news/pr/2002/021102_eledialogue.shtml.
15. CITES 17.89-17.91, "Specimens produced from synthetic or cultured DNA" (2017); www.cites.org/eng/dec/valid17/81839 .

10.1126/science.aat1596

Insurance coverage for genomic tests

On 16 March, the Centers for Medicare and Medicaid Services (CMS) announced that Medicare will cover Food and Drug Administration (FDA)–approved or cleared genomic tests that encompass broad gene panels for advanced cancer patients (*1*). The final policy does not include the initial draft's proposed "coverage with evidence development" (CED)—i.e., coverage of tests run as part of clinical trials and registries—which some had argued should be applied to develop a stronger evidence base for these tests (*2*). Instead, tests not already approved in the national coverage determination can be reviewed for coverage by local Medicare Administrative Contractors (MACs). The new policy reflects a substantial shift in determining how genomic tests are evaluated for coverage, which provides a needed "roadmap" for coverage. However, to develop effective and efficient policies, stakeholders should support further research to address how the new policy will affect ongoing cancer research as well as the access to and affordability of next-generation sequencing testing for cancer patients.

The new policy has the potential to increase access to testing, but it may remain out of reach for many patients. Private payers may not follow the CMS policy for covered tests, as there are myriad reasons that payers have limited their coverage for broad genomic tests

(*3*), and private payers often do not use Medicare policies as precedents (*4*). The tests that remain uncovered by CMS may not be covered by local MACs either (*5*). The numerous laboratories that offer their own tests that do not currently meet the coverage requirements in the new policy may have trouble finding the trial participants and funding they need to obtain the evidence required. Although CMS's policy may spur these laboratories to develop evidence even without a CED requirement, this process may take several years, and its outcome is uncertain (*4*). The CMS policy is binding only on Medicare; it is uncertain whether states will cover these tests for Medicaid patients (*6*).

Likewise, the policy may increase affordability and equity for these tests, but with caveats. Benefit-cost tradeoffs were not examined as they are outside the scope of CMS. CMS is caught in an ongoing dilemma: Coverage policies are determined irrespective of cost, yet there is a constant drumbeat of calls to reduce Medicare expenditures (*7*). Lastly, it will be important to understand the implications of the new policy for genomic tests for patients with other types of cancer and with other conditions, which face similar challenges to coverage (*8*).

Given today's challenging health policy environment, CMS should work with stakeholders, including other federal agencies, to carefully evaluate the benefits and risks of this novel coverage approach and to consider what additional policy mechanisms will be needed to ensure that the necessary evidence is generated. We must address the substantial uncertainty about the impact of coverage policies on the health outcomes of Medicare beneficiaries.

Kathryn A. Phillips,[1,2,3] Julia R. Trosman,[4] Patricia A. Deverka,[1,5] Bruce Quinn,[6] Sean Tunis,[7] Peter J. Neumann,[8] James D. Chambers,[8] Louis P. Garrison Jr.,[9] Michael P. Douglas,[1] Christine B. Weldon[4]

[1]Department of Clinical Pharmacy, Center for Translational and Policy Research on Personalized Medicine (TRANSPERS), University of California San Francisco, San Francisco CA 94143, USA. [2]Philip R. Lee Institute for Health Policy, San Francisco, CA 94118, USA. [3]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA. [4]Center for Business Models in Healthcare, Northwestern University Feinberg School of Medicine, Chicago, IL 60611 USA. [5]American Institutes for Research, Chapel Hill, NC 27517, USA. [6]Bruce Quinn Associates, Los Angeles, CA 90036, USA. [7]Center for Medical Technology Policy (CMTP), Baltimore, MD 21202, USA. [8]The Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA 02111, USA. [9]The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, Department of Pharmacy, University of Washington, Seattle, WA 98195, USA. *Corresponding author. Email address: kathryn.phillips@ucsf.edu

REFERENCES

1. CMS, Decision Memo for Next-Generation Sequencing (NGS) for Medicare Beneficiaries with Advanced Cancer (CAG-00450N) (2018); www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=290&DocID=CAG-00450N&bc=AAAAAAAAQAAA&.
2. R. Eisenberg, H. Varmus, *Science* **358**, 1133 (2017).
3. K. A. Phillips *et al.*, *Nat. Biotechnol* **35**, 614 (2017).
4. J. D. Chambers *et al.*, *Health Aff. (Millwood)* **34**, 1376 (2015).
5. D. R. Levinson, "Local coverage determinations create inconsistency in Medicare coverage" (2014); https://oig.hhs.gov/oei/reports/oei-01-11-00500.pdf.
6. E. Schattner, "Medicare will cover genetic cancer testing, with caveats," *Forbes* (2018); www.forbes.com/sites/elaineschattner/2018/03/16/medicare-will-cover-genetic-cancer-testing-with-caveats/#d78931d9e556.
7. P. J. Neumann, J. D. Chambers, *N. Engl. J. Med.* **367**, 1775 (2012).
8. J. R. Trosman *et al.*, *J. Natl. Compr. Canc. Netw.* **15**, 219 (2017).

ACKNOWLEDGMENTS

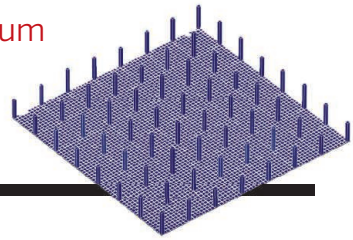
This letter was funded by grants from the National Human Genome Research Institute (R01 HG007063 and U01 HG009599) and by the Helen Diller Family Comprehensive Cancer Center Support Grant of the National Institutes of Health under Award Number P30 CA082103-18.

10.1126/science.aas9268

RESEARCH

An integrated quantum optics platform

Wang et al., *p. 285*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

Megafauna, such as these rhinos, have been under threat from humans for millennia.

EXTINCTION

Megafaunal loss

Today, it is well known that human activities put larger animals at greater risk of extinction. Such targeting of the largest species is not new, however. Smith *et al*. show that biased loss of large-bodied mammal species from ecosystems is a signature of human impacts that has been following hominin migrations since the Pleistocene. If the current trend continues, terrestrial mammal body sizes will become smaller than they have been over the past 45 million years. Megafaunal mammals have a major impact on the structure of ecosystems, so their loss could be particularly damaging. —SNV

Science, this issue p. 310

EARLY EARTH

A strongly oxidizing Paleoproterozoic era

Two billion years ago, marine sulfate concentrations were around one-third as high as modern ones, constituting an oxidizing capacity equivalent to more than 20% of that of the modern ocean-atmosphere system. Blättler *et al*. found this by analyzing a remarkable evaporite succession more than 1 billion years older than the oldest comparable deposit discovered to date. These quantitative results, for a time when only more qualitative information was previously available, provide a constraint on the magnitude and timing of early Earth's response to the Great Oxidation Event 2.3 billion years ago. —HJS

Science, this issue p. 320

NEURODEVELOPMENT

Transient instruction changes migration

The brain neocortex is built by waves of neurons migrating from deep within the brain to the surface layers. Ohtaka-Maruyama *et al*. found that a layer of neurons that multipolar neurons encounter on their travels instructs the migrating neurons to change phenotype and direction (see the Perspective by Schinder and Lanuza). These subplate neurons form transient glutamatergic synapses with the immature migrants. This results in the migrating multipolar neurons becoming bipolar, more directed, and faster in their final migrations. —PJH

Science, this issue p. 313; see also p. 265

AUTISM GENOMICS

Inherited variation contributes to autism

About one-quarter of genetic variants that are associated with autism spectrum disorder (ASD) are due to de novo mutations in protein-coding genes. Brandler *et al*. wanted to determine whether changes in noncoding regions of the genome are associated with autism. They applied whole-genome sequencing to ~2600 families with at least one affected child. Children with ASD had inherited structural variants in noncoding regions from their father. Regulatory regions of some specific genes were disrupted among multiple families, supporting the idea that a component of autism risk involves inherited noncoding variation. —LMZ

Science, this issue p. 327

DRUG DEVELOPMENT

An innovative approach for a rare disease

Charcot-Marie-Tooth disease type 2A (CMT2A) is a rare, inherited neurodegenerative condition. Affected individuals develop severe progressive muscle weakness, motor deficits, and peripheral neuropathy. Although defects in the gene encoding mitofusin 2 (MFN2) are known to cause CMT2A, the disease remains incurable. Rocha *et al*. identified specific MFN2 residues contributing to the disease and developed a class of MFN2-agonist drugs. The small

CREDITS: (TOP TO BOTTOM) J. WANG *ET AL.*; RYAN RIGGS

Published by AAAS

molecules restored mitochondrial fusion and activity in the sciatic nerves of mice; they may also help in other diseases linked to mitochondrial trafficking. —PNK

Science, this issue p. 336

HIV

Taking an active interest in HIV latency

HIV cure efforts have been thwarted by an inability to target the latent reservoir, which is thought to be largely composed of resting CD4 T cells. A recent report suggested that the Fcγ receptor CD32 might be a marker of latently infected CD4 T cells. Abdel-Mohsen *et al.* meticulously examined T cells from treated HIV patients across the world. $CD32^+$ HIV-infected T cells had an activated phenotype and HIV RNA, indicating active HIV transcription. In contrast, the majority of HIV DNA resided in $CD32^-$ cells. Thus, targeting $CD32^+$ cells is unlikely to hit the latent HIV reservoir. —LP

Sci. Transl. Med. **10**, eaar6759 (2018).

MATERIALS SCIENCE

Reserving the right to stretch

Retractable antennae or certain spider silks can stretch well beyond their apparent length because they have a reserve of material that lets them expand and contract over much longer distances. Grandgeorge *et al*. made nonwoven fibrous membranes by electrospinning a block copolymer with varying ratios of two components. They infused these membranes with a liquid that let the fibers buckle and fold without changing the apparent surface area. When the membranes were stretched, this material could unbuckle and slide along the membrane surface, allowing it to extend without breakage. —MSL

Science, this issue p. 296

CANCER

Earlier detection of ovarian cancer

Ovarian cancer is the fifth leading cause of cancer-related deaths among females in the United States, owing in part to the late stage at which it is often diagnosed. Survival rates increase dramatically when it is detected early, and new methods for advanced detection are greatly needed. Williams *et al.* developed a carbon nanotube–based sensor that optically detects the U.S. Food and Drug Administration–approved ovarian cancer biomarker HE4. When implanted into live cancer-bearing mice, distinct wavelength responses from individual nanotubes in the device rapidly and repeatedly differentiated mice with ovarian cancer from controls. The same was true in tests using samples from ovarian cancer patients. —PJB

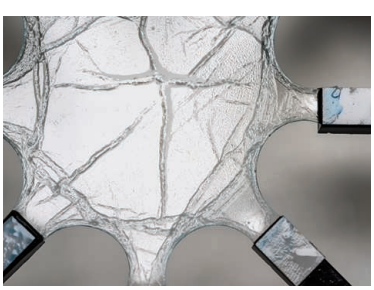
Sci. Adv. 10.1126/sciadv.aaq1090 (2018).

QUANTUM GASES

Recurring coherence

A finite isolated system should return almost to its initial state if it evolves for long enough. For a large system, "long enough" is often unfeasibly long. Rauer *et al*. found just the right conditions to observe the recurrence of the initial state in a system of two one-dimensional superfluids with thousands of atoms in each. The superfluids were initially coupled—locking their quantum mechanical phases together—and then allowed to evolve independently. After the uncoupling, the researchers observed their phases regaining coherence two more times. —JS

Science, this issue p. 307

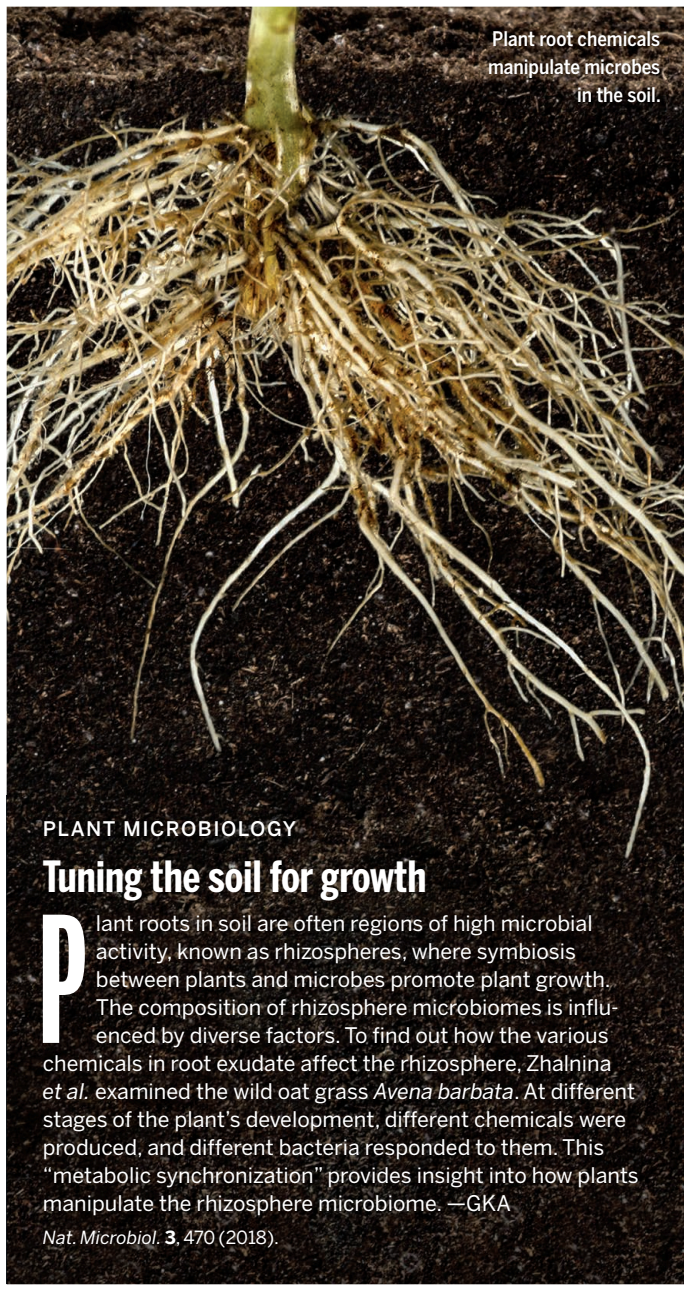


A designed polymer can stretch like a spring.

PHOTOS: (FROM LEFT) SORBONNE UNIVERSITÉ/CNRS; PHIL DEGGINGER/SCIENCE SOURCE

IN OTHER JOURNALS

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



Plant root chemicals manipulate microbes in the soil.

PLANT MICROBIOLOGY

Tuning the soil for growth

Plant roots in soil are often regions of high microbial activity, known as rhizospheres, where symbiosis between plants and microbes promote plant growth. The composition of rhizosphere microbiomes is influenced by diverse factors. To find out how the various chemicals in root exudate affect the rhizosphere, Zhalnina *et al.* examined the wild oat grass *Avena barbata*. At different stages of the plant's development, different chemicals were produced, and different bacteria responded to them. This "metabolic synchronization" provides insight into how plants manipulate the rhizosphere microbiome. —GKA

Nat. Microbiol. **3**, 470 (2018).

SINGLE-CELL GENOMICS

Tying genotype to phenotype, cell by cell

RNA transcripts can now be sequenced within single cells. Such studies identify variants that affect gene expression at a level that may not be detected by bulk sequencing. Van der Wijst *et al*. sequenced ~25,000 individual blood cells from 45 individuals. Expression quantitative trait loci that had previously been identified in bulk studies were found, which helped validate the approach. An additional 287 genes were identified that showed differences in expression in single cells resulting from genetic variation, 48 of which differed among cell types. Furthermore, single-cell sequencing detected variants that affect regulatory networks and revealed

MICROBIOLOGY

Weathering life after death

Life is tough on rock exposed to ultraviolet radiation and extreme desiccation, and food is scarce. But microbial life does get a grip, and it contributes substantially to rock weathering by harvesting minerals for metabolism. Brewer and Fierer sampled 149 gravestones from Europe and the Americas and used marker-gene and shotgun metagenomic sequencing to uncover what was living on them. Geography, climate, and rock type were the main determinants of the microbial communities. Granite-based organisms were genetically geared for acid tolerance and mobility, whereas limestone-based communities tended to live in lichen associations, fix carbon, and resist radiation. Many of the communities were symbiotic or endolithic, indicating that some recourse to food and shelter is available even on the smoothest slab. —CA

Environ. Microbiol. **20**, 958 (2018).

Tombstone encrustation is influenced by rock chemistry.

personalized coexpression patterns. —LMZ

Nat. Genet. 10.1038/s41588-018-0089-9 (2018).

ICE SHEETS

Explaining uneven mass loss

The Greenland Ice Sheet, along with the Antarctic Ice Sheet and glaciers worldwide, is melting at an accelerating rate. This melting is not uniform, however, with adjacent fjords often exhibiting quite different behaviors. What can account for those differences? Millan *et al.* present oceanographic observations and bathymetric data from the vicinity of 20 major glaciers in southeast Greenland, which show that retreating glaciers occupy deep valleys exposed to warm Atlantic water, whereas stable ones rest on sills away from warm water. These observations can explain the complex pattern of ice-front positions from the 1930s to the present. —HJS

Geophys. Res. Lett. 10.1002/2017GL076561 (2018).

MAGNETISM

Manipulating an antiferromagnet

Magnetic materials are routinely used in electronic devices, and the ability to change their magnetic state using electric fields is highly desirable. Most devices use ferromagnets—materials in which individual atomic spins all point in one direction—but there are important advantages to developing analogous devices with antiferromagnetic (AFM) materials. Liu *et al.* made thin films of the AFM $MnPt_3$, a material that transitions from a noncollinear to a collinear AFM state a bit above room temperature. The thin films were grown on a ferroelectric substrate; applying an electric field to the substrate near the transition temperature changed the strain in the $MnPt_3$ film, which in turn caused the film to change its spin structure from collinear to noncollinear. —JS

Nat. Electron. **1**, 172 (2018).

METASURFACES

Reconfigurable metasurfaces

The design flexibility in patterning metasurfaces allows many bulk optical components to be replaced with elements just a fraction of their thickness. For integrated optical devices or miniaturized lightweight components, metasurfaces offer a clear advantage over glass-based optics. By combining metasurface designs with arrays of electromechanically actuated microcantilevers, Zhao *et al.* demonstrate the ability to actively control the optical properties of a metasurface. Operating in the terahertz regime, they show that they can manipulate the polarization of transmitted light in real time to form a waveplate. Such active control of metasurface properties will be useful for developing imaging and sensing terahertz technologies. —ISO

Optica **5**, 303 (2018).

CANCER

Even more genes control cell growth

A full understanding of cancer evolution needs a systematic approach. Screening for genes that drive unrestrained proliferation of human cells could answer questions about the relative roles of mutations, gene dose, and tissue specificity in cancer development. To find those for which dosage changes could promote or inhibit cell proliferation, Sack *et al.* screened 16,000 genes in mammary, fibroblast, and pancreatic cells. They found nearly 400 genes that drove cell growth and more than 1000 that suppressed proliferation. Many expected genes that control the cell cycle were detected, but most of the identified genes were not previously known to regulate proliferation. Alterations in the somatic copy number of these genes in cancers indicate that they may contribute to tumorigenesis. Proliferation control depended strongly on cell type. Specific genetic-network architecture may be created during development in different cell types, perhaps through epigenetic control. —LBR

Cell 10.1016/j.cell.2018.02.037 (2018).

PHOTO: STAN PRITCHARD/ALAMY STOCK PHOTO

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

YEAST GENOMICS

Trigenic interactions in yeast link bioprocesses

To dissect the genotype-phenotype landscape of a cell, it is necessary to understand interactions between genes. Building on the digenic protein-protein interaction network, Kuzmin *et al*. created a trigenic landscape of yeast by using a synthetic genetic array (see the Perspective by Walhout). Triple-mutant analyses indicated that the majority of genes with trigenic associations functioned within the same biological processes. These converged on networks identified in the digenic interaction landscape. Although the overall effects were weaker for trigenic than for digenic interactions, trigenic interactions were more likely to bridge biological processes in the cell. —LMZ

Science, this issue p. 283; see also p. 269

ADVANCED IMAGING

Continuing the resolution revolution

The living cell contains dynamic, spatially complex subassemblies that are sensitive to external perturbations. To minimize such perturbations, cells should be imaged in their native multicellular environments, under as gentle illumination as possible. However, achieving the spatiotemporal resolution needed to follow three-dimensional subcellular processes in detail under these conditions is challenging: Sample-induced aberrations degrade resolution and sensitivity, and high resolution usually requires intense excitation. Liu *et al*. combined noninvasive lattice light-sheet microscopy with aberration-correcting adaptive optics to study a variety of delicate subcellular events in vivo, including organelle remodeling during mitosis and growth cone dynamics during spinal cord development. —SMH

Science, this issue p. 284

QUANTUM OPTICS

Large-scale integrated quantum optics

The ability to pattern optical circuits on-chip, along with coupling in single and entangled photon sources, provides the basis for an integrated quantum optics platform. Wang *et al*. demonstrate how they can expand on that platform to fabricate very large quantum optical circuitry. They integrated more than 550 quantum optical components and 16 photon sources on a state-of-the-art single silicon chip, enabling universal generation, control, and analysis of multidimensional entanglement. The results illustrate the power of an integrated quantum optics approach for developing quantum technologies. —ISO

Science, this issue p. 285

NANOPHOTONICS

Light confined to a single atomic layer

The development of nanophotonic technology is reliant on the ability to confine light to spatial dimensions much less than the wavelength of the light itself. Typically, however, in metal plasmonic approaches, there is a trade-off between confinement and losses. Alcaraz Iranzo *et al*. fabricated heterostructures comprising monolayers of graphene and hexagonal boron nitride (hBN) and an array of metallic rods. The light was confined vertically (as propagating plasmons) between the metal and the graphene, even when the insulating hBN spacer was just a single monolayer. Such heterostructures should provide a powerful and versatile platform for nanophotonics. —ISO

Science, this issue p. 291

MATERIALS SCIENCE

Small, smooth, and bendable diamonds

If you manage to deform a diamond, it usually means you have broken it. Diamonds have very high hardness, but they do not deform elastically. This limits their usefulness for some applications. However, Banerjee *et al*. discovered that diamond nanoneedles can deform elastically after all (see the Perspective by LLorca). The key was in their small size (300 nm), which allowed for very smooth-surfaced, defect-free diamonds. The deformation was close to the theoretical limit for diamond, which opens up the potential for applications in microelectronics and drug delivery. —BG

Science, this issue p. 300; see also p. 264

COLLOIDS

Images frozen in time

Freezing processes involving multiple phases are common, as seen in the zone refining used to remove impurities from a crystal through the heating of a narrow region. However, it is difficult to visualize the processes as they happen. Dedovets *et al*. studied the freezing of an oil-in-water emulsion and the interaction of the oil droplets with the leading edge of the freezing water. In situ visualization of the freezing front was achieved through confocal microscopy of samples on a moving sample holder. The authors tracked the effects of solution concentration on the droplet positions and observed premelting at the interface of the droplets by using fluorescence markers. —MSL

Science, this issue p. 303

PLANT ECOLOGY

A short-term trend reversed

Theory and empirical data both support the paradigm that C_4 plant species (in which the first product of carbon fixation is a four-carbon molecule) benefit less from rising carbon dioxide (CO_2) concentrations than C_3 species (in which the first product is a three-carbon molecule). This is because their different photosynthetic physiologies respond differently to atmospheric CO_2 concentrations. Reich *et al*. document a reversal of this pattern in a 20-year CO_2 enrichment experiment using grassland plots with each type of plant (see the Perspective by Hovenden and Newton). Over the first 12 years, biomass increased with elevated CO_2 in C_3 plots but not C_4 plots, as expected. But over the next 8 years, the pattern reversed: Biomass increased in C_4 plots but not C_3 plots. Thus, even the best-supported short-term drivers of plant response to global change might not predict long-term results. —AMS

Science, this issue p. 317; see also p. 263

CANCER GENOMICS

The cellular composition of H3K27M gliomas

Diffuse midline gliomas with histone H3 lysine27-to-methionine mutations (H3K27M-glioma) are an aggressive type of childhood cancer with few options for treatment. Filbin *et al*. used a single-cell sequencing approach to study the oncogenic programs, genetics, and cellular hierarchies of H3K27M-glioma. Tumors were mainly composed of cells resembling oligodendrocyte precursor cells, whereas differentiated malignant cells were a smaller fraction. In comparison with other gliomas, these cancers had distinct oncogenic programs and stem cell–like profiles that contributed to their stable tumor-propagating potential. The analysis also identified a lineage-specific marker that

may be useful in developing therapies. —LMZ

Science, this issue p. 331

STRUCTURAL BIOLOGY

First steps of translocation elucidated

Ribosomes synthesizing membrane or secretory proteins are targeted to the endoplasmic reticulum (ER) in eukaryotic cells by the signal recognition particle (SRP). Upon reaching the ER, the SRP interacts with its receptor to promote transfer of the signal sequence to the protein-conducting channel or translocon. Kobayashi *et al*. studied the ribosomal complex that forms on the ER, in which the SRP and its receptor interact to transfer the newly synthesized protein to the translocon. The observed organization of the assembly reveals the roles of multiple eukaryotic-specific protein components present in the SRP and its receptor in stabilizing the conformation that facilitates signal sequence handover. —SMH

Science, this issue p. 323

NEURODEGENERATION

Conformation changes or cooperativity?

Parkin is a commonly mutated protein ubiquitin ligase implicated in Parkinson's disease. Understanding how Parkin functions is important to understanding its pathogenic role. In a Perspective, Arkinson and Walden discuss recent structural insights into the activities of Parkin. These have led to the interesting proposal that Parkin, and other ubiquitin ligases in the same family, may function cooperatively—so that two proteins are needed to carry out its enzymatic functions. Although conformational changes of Parkin are not ruled out, this model may help explain the Parkinson's disease–associated mutation profile that spans the Parkin protein. —GKA

Science, this issue p. 267

HYPOTHESIS

Restoring blood vessel stability

What if a wide array of diseases that afflict blood vessels—from atherosclerosis to transplant vasculopathy and vascular malformations—had a common causal element? In a Perspective, Schwartz *et al.* propose that fundamental aspects of normal vascular remodeling promote disease if prolonged or exaggerated. They also suggest that unchecked vascular endothelial cell activation and extracellular matrix remodeling are common to multiple pathologies. If this is the case, interventions to restore vascular stability could improve health outcomes in multiple disease states. —LC

Science, this issue p. 270

HORMONES

Estrogen accentuates autoimmunity

More than any other risk factor, being female confers the greatest risk of developing an autoimmune disorder. Mohammad *et al*. identified a role for estrogen in the development of autoimmune T cell responses. Deletion of estrogen receptor α (ERα) in T cells reduced disease burden in a mouse model of colitis. ERα-expressing T cells were more activated after stimulation and produced more proinflammatory cytokines than T cells lacking this receptor. Conversely, ERα-deficient T cells were more likely to differentiate into regulatory T cells, which suppress the development of autoimmune disorders. —ERW

Sci. Signal. **11**, eaap9415 (2018).

ALLERGY

Potent platelets

Anaphylaxis results from inappropriate immune responses to allergens. Human platelets express the immunoglobulin G (IgG) receptor FcγRIIA/CD32A and release inflammatory mediators in response to their engagement. However, the contribution of platelets to anaphylaxis is not well understood. To address this, Beutier *et al*. developed mouse models that express either human FcγRIIA/CD32A alone or the full human IgG receptor complexity. Anaphylaxis induced a marked decrease in platelet levels, but preventive platelet depletion reduced anaphylaxis severity. A clinical study of patients with drug-induced anaphylaxis showed that a severe reaction was likewise associated with fewer circulating platelets. Activated platelets released serotonin, which contributed to anaphylaxis severity. Thus, platelets play a critical role in IgG-mediated anaphylaxis. —CNF

Sci. Immunol. **3**, eaan5997 (2018).

RESEARCH ARTICLE SUMMARY

YEAST GENOMICS

Systematic analysis of complex genetic interactions

Elena Kuzmin,* Benjamin VanderSluis,* Wen Wang, Guihong Tan, Raamesh Deshpande, Yiqun Chen, Matej Usaj, Attila Balint, Mojca Mattiazzi Usaj, Jolanda van Leeuwen, Elizabeth N. Koch, Carles Pons, Andrius J. Dagilis, Michael Pryszlak, Jason Zi Yang Wang, Julia Hanchard, Margot Riggi, Kaicong Xu, Hamed Heydari, Bryan-Joseph San Luis, Ermira Shuteriqi, Hongwei Zhu, Nydia Van Dyk, Sara Sharifpoor, Michael Costanzo, Robbie Loewith, Amy Caudy, Daniel Bolnick, Grant W. Brown, Brenda J. Andrews,† Charles Boone,† Chad L. Myers†

INTRODUCTION: Genetic interactions occur when mutations in different genes combine to result in a phenotype that is different from expectation based on those of the individual mutations. Negative genetic interactions occur when a combination of mutations leads to a fitness defect that is more exacerbated than expected. For example, synthetic lethality occurs when two mutations, neither of which is lethal on its own, generate an inviable double mutant. Alternatively, positive genetic interactions occur when genetic perturbations combine to generate a double mutant with a greater fitness than expected. Global digenic interaction studies have been useful for understanding the functional wiring diagram of the cell and may also provide insight into the genotype-to-phenotype relationship, which is important for tracking the missing heritability of human health and disease. Here we describe a network of higher-order trigenic interactions and explore its implications.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aao1729

RATIONALE: Variation in phenotypic outcomes in different individuals is caused by genetic determinants that act as modifiers. Modifier loci are prevalent in human populations, but knowledge regarding how variants interact to modulate phenotype in different individuals is lacking. Similarly, in yeast, traits including conditional essentiality—in which certain genes are essential in one genetic background but nonessential in another—often result from an interplay of multiple modifier loci. Because complex modifiers may underlie the genetic basis of physiological states found in natural populations, it is critical to understand the landscape of higher-order genetic interactions.

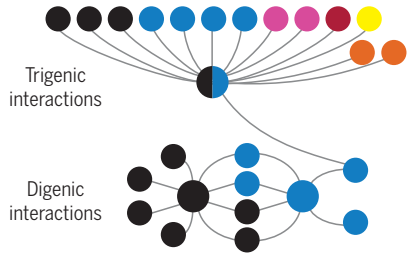
RESULTS: To survey trigenic interactions, we designed query strains that sampled key features of the global digenic interaction network: (i) digenic interaction strength, (ii) average number of digenic interactions, and (iii) digenic interaction profile similarity. In total, we tested ~400,000 double and ~200,000 triple mutants for fitness defects and identified ~9500 digenic and ~3200 trigenic negative interactions. Although trigenic interactions tend to be weaker than digenic interactions, they were both enriched for functional relationships. About one-third of trigenic interactions identified "novel" connections that were not observed in our digenic control network, whereas the remaining approximately two-thirds of trigenic interactions "modified" a digenic interaction, suggesting that the global digenic interaction network is important for understanding the trigenic interaction network. Despite their functional enrichment, trigenic interactions also bridged distant bioprocesses. We estimate that the global trigenic interaction network is ~100 times as large as the global digenic network, highlighting the potential for complex genetic interactions to affect the biology of inheritance.

CONCLUSION: The extensive network of trigenic interactions and their ability to generate functionally diverse phenotypes suggest that higher-order genetic interactions may play a key role in the genotype-to-phenotype relationship, genome size, and speciation. ■

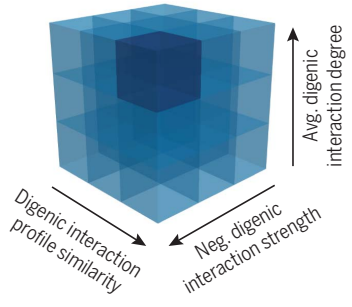
Systematic analysis of trigenic interactions. We surveyed for trigenic interactions and found that they are ~100 times as prevalent as digenic interactions, often modify a digenic interaction, and connect functionally related genes as well as genes in more diverse bioprocesses (multicolored nodes). PPI, protein-protein interaction.

Mapping digenic and corresponding trigenic interaction networks

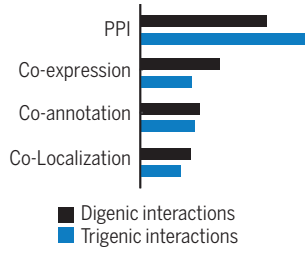
Digenic and trigenic interactions were scored in parallel for the corresponding single mutant and double mutant query strains (large nodes).



Global digenic network properties of query genes enriched for trigenic interactions

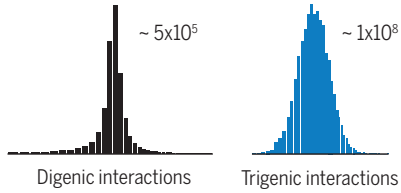


Trigenic interactions are functionally informative



Trigenic interactions often overlap digenic interactions

Estimate of global negative genetic interaction networks



The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: chadm@umn.edu (C.L.M.); brenda.andrews@utoronto.ca (B.J.A.); charlie.boone@utoronto.ca (C.B.)
Cite this article as: E. Kuzmin *et al.*, *Science* **360**, eaao1729 (2018). DOI: 10.1126/science.aao1729

RESEARCH ARTICLE

YEAST GENOMICS

Systematic analysis of complex genetic interactions

Elena Kuzmin,[1,2]*† Benjamin VanderSluis,[3]* Wen Wang,[3] Guihong Tan,[1] Raamesh Deshpande,[3] Yiqun Chen,[1] Matej Usaj,[1] Attila Balint,[1,4]‡ Mojca Mattiazzi Usaj,[1] Jolanda van Leeuwen,[1] Elizabeth N. Koch,[3] Carles Pons,[3] Andrius J. Dagilis,[5] Michael Pryszlak,[1] Jason Zi Yang Wang,[1,2] Julia Hanchard,[1,2] Margot Riggi,[6,7,8,9] Kaicong Xu,[3] Hamed Heydari,[1,2] Bryan-Joseph San Luis,[1] Ermira Shuteriqi,[1] Hongwei Zhu,[1] Nydia Van Dyk,[1] Sara Sharifpoor,[1] Michael Costanzo,[1] Robbie Loewith,[6,8,9] Amy Caudy,[1,2] Daniel Bolnick,[5] Grant W. Brown,[1,4] Brenda J. Andrews,[1,2]§ Charles Boone,[1,2,10]§ Chad L. Myers[3]§

To systematically explore complex genetic interactions, we constructed ~200,000 yeast triple mutants and scored negative trigenic interactions. We selected double-mutant query genes across a broad spectrum of biological processes, spanning a range of quantitative features of the global digenic interaction network and tested for a genetic interaction with a third mutation. Trigenic interactions often occurred among functionally related genes, and essential genes were hubs on the trigenic network. Despite their functional enrichment, trigenic interactions tended to link genes in distant bioprocesses and displayed a weaker magnitude than digenic interactions. We estimate that the global trigenic interaction network is ~100 times as large as the global digenic network, highlighting the potential for complex genetic interactions to affect the biology of inheritance, including the genotype-to-phenotype relationship.

Genetic interactions occur when a combination of mutations in different genes leads to an unexpected phenotype that deviates from a model incorporating the combined effects of the corresponding single-mutant phenotypes. In humans, each individual carries thousands of different variants that may modulate gene function, which means that there is incredible potential for combinatorial genetic interactions to determine our personal phenotype (*1*, *2*). Indeed, genetic interactions are thought to represent a substantial component of the missing heritability associated with current genome-wide association studies (GWAS) (*3*). However, the statistical limitations associated with GWAS data sets preclude the detection of specific genetic interactions, and thus potential genetic networks underlying inherited traits, including diseases, remain elusive (*3*–*5*). To address the role of genetic interactions in the genotype-to-phenotype relationship, we have been exploring their general principles through systematic analysis of genetic networks underlying cellular fitness in a genetically tractable model system, the budding yeast *Saccharomyces cerevisiae* (*6*). Our previous studies focused predominantly on genetic interactions involving two genes (digenic interactions) (*7*). In this study, we analyzed a series of single-, double-, and triple-mutants by quantifying their colony size, as a proxy for fitness, to systematically explore complex genetic interactions.

There are two basic types of fitness-based genetic interactions. A negative genetic interaction refers to a combination of mutations that results in a fitness defect that is more severe than expected (*8*). Synthetic lethality is an extreme example of a negative genetic interaction and occurs when two mutations, neither of which is lethal on its own, combine and lead to an inviable double-mutant phenotype (*9*, *10*). Conversely, a positive genetic interaction occurs when a combination of genetic perturbations generates a double mutant with a greater fitness than expected; one example is genetic suppression, in which the fitness defect of a query mutant is alleviated by a mutation in a second gene (*11*). To map a global digenic interaction network for yeast, we constructed millions of double mutants and identified hundreds of thousands of negative and positive genetic interactions (*7*). To put these results in perspective, although only ~1000 of the ~6000 total yeast genes are individually essential and cause lethality when deleted (*12*) and an equivalent number of nonessential genes cause a slow-growth defect under standard laboratory conditions (*13*), ~550,000 different yeast gene pairs display a combinatorial negative genetic interaction, including a subset of ~10,000 extreme synthetic lethal interactions involving nonessential gene pairs (*7*). Thus, there are numerous potential ways to generate extreme lethal phenotypes through negative digenic interactions of nonessential gene pairs.

The set of digenic interactions for a query gene, its genetic interaction profile, provides a quantitative measure of function, because genes with similar roles have overlapping profiles (*14*, *15*). Genes belonging to the same biological pathway or protein complex display highly similar genetic interaction profiles. Moreover, a global network based on digenic interaction profile similarity reveals a hierarchy of functional modules, which includes detailed pathways and complexes, that in turn cluster into larger modules corresponding to bioprocesses. Those larger units subsequently assemble into modules corresponding to cellular compartments to outline the functional architecture of a cell (*7*).

A complete understanding of the role of genetic interactions in the genotype-to-phenotype relationship requires that we also investigate complex, higher-order genetic interactions involving more than two genes. Because there are ~2000 times as many triple gene combinations as gene pairs (~18 million), it is possible that there is a substantially greater number of trigenic than digenic interactions and that higher-order interactions may be important for driving inherited traits. In this study, we surveyed yeast trigenic interactions, sampling quantitative features of the digenic network, and explored the implications of the higher-order genetic interaction network.

Mapping trigenic interactions quantitatively and surveying the global trigenic landscape

To explore the trigenic interaction landscape, we designed query strains that sampled three key quantitative features of our global digenic interaction network (*7*). We designed query strains carrying mutations in two genes spanning a range of the following features: (1) digenic interaction strength, (2) number of digenic interactions (average digenic interaction degree), and (3) digenic interaction profile similarity (Fig. 1A and table S1). Gene pairs were selected to fill

[1]The Donnelly Centre, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. [2]Department of Molecular Genetics, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. [3]Department of Computer Science and Engineering, University of Minnesota–Twin Cities, 200 Union Street, Minneapolis, MN 55455, USA. [4]Department of Biochemistry, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. [5]Department of Integrative Biology, 1 University Station C0990, University of Texas at Austin, Austin, TX 78712, USA. [6]Department of Molecular Biology, University of Geneva, Geneva 1211, Switzerland. [7]Department of Biochemistry, University of Geneva, 1211 Geneva, Switzerland. [8]iGE3 (Institute of Genetics and Genomics of Geneva), 1211 Geneva, Switzerland. [9]Swiss National Centre for Competence in Research Programme Chemical Biology, 1211 Geneva, Switzerland. [10]RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan.
*These authors contributed equally to this work. †Present address: Rosalind and Morris Goodman Cancer Research Centre, McGill University, 1160 Avenue des Pins Ouest, Montreal, Quebec H3A 1A3, Canada. ‡Present address: Center for Chromosome Stability, Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen N, Denmark.
§Corresponding author. Email: chadm@umn.edu (C.L.M.); brenda.andrews@utoronto.ca (B.J.A.); charlie.boone@utoronto.ca (C.B.)

bins of varying digenic interaction attributes and to cover all major biological processes in the cell, thus producing a sample that would provide a diverse survey of the trigenic interaction landscape. We largely focused on unambiguous singletons because duplicated genes represent a relatively small subset of genes and thus can only represent a small fraction of the global trigenic interaction network. For this survey, we constructed 151 double-mutant query strains and 302 single-mutant strains, encompassing 47 temperature-sensitive alleles of different essential genes and 255 deletion alleles of nonessential genes. The query strains in this set were selected to span the different digenic attribute bins according to predefined thresholds (table S1). An additional 31 double-mutant queries fell outside of the defined thresholds but were included for validation and comparison purposes (data S1 to S3) (*16*). The fitness of the resulting query strains was measured using a quantitative growth assay, and the behavior of the single- and double-mutant query strains showed strong agreement with our previously published data set (figs. S1 and S2 and data S4) (*7*, *15*).

Trigenic interaction screening required development and implementation of three operational components. First, synthetic genetic array (SGA) analysis—an automated form of yeast genetics that is often used to cross a query gene mutation into an array of single mutants to generate a defined set of haploid double mutants (*6*)—was adapted such that a double-mutant query strain could be crossed into an array of single mutants to generate triple mutants for trigenic interaction analysis (Fig. 1B). Because the identification of a trigenic interaction requires comparison with the corresponding double mutants, we also conducted screens in which the individual mutants of the query gene pair were scored for digenic interactions (Fig. 1B). Second, for experimental feasibility, we assembled a diagnostic array of 1182 strains, comprising 990 nonessential gene deletion mutants and 192 essential gene mutants carrying temperature-sensitive alleles, which combine to span ~20% of the yeast genome (data S5). The diagnostic array was designed to be highly representative of the rest of the genome in terms of exhibited genetic interaction profiles (fig. S3). Briefly, array strains were selected from a larger genetic interaction data set for their ability to represent different regions of the global network in a minimally redundant way. This was accomplished by iteratively selecting strains to maximize the performance of profile similarities when predicting coannotations to a functional gold standard (*17*). Third, we developed a scoring method, the τ-SGA score, which combines double- and triple-mutant fitness estimates derived from colony size measurements to identify trigenic interactions quantitatively (Fig. 1C). The τ-SGA score differs from the MinDC score reported previously (*18*), because it accounts for all cases in which two of the genes are not independent, resulting in an expectation that contains digenic interaction effects scaled by the fitness of the noninteracting genes (fig. S4) (*16*). The final trigenic τ-SGA interaction score then accounts for digenic effects but also enables detection of trigenic interactions in which digenic effects of insufficient explanatory power can be found.

We focused exclusively on the analysis of deleterious negative trigenic interactions for two reasons. First, quantitative scoring of negative

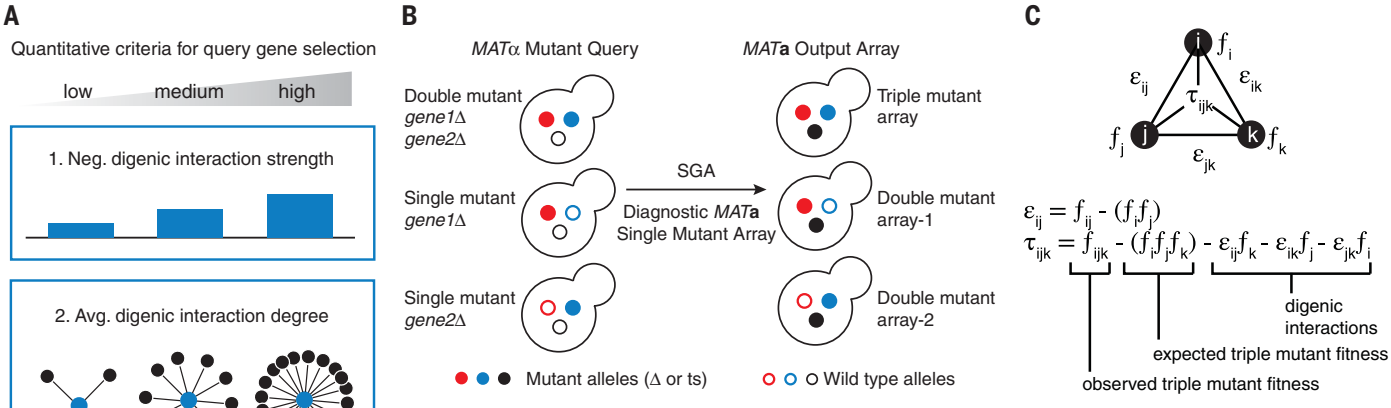


Fig. 1. Triple-mutant synthetic genetic array (SGA) analysis. (**A**) Criteria for selecting query strains for sampling trigenic interaction landscape of singleton genes in yeast. The gene pairs were grouped into three general categories based on a range of features: (1) Digenic interaction strength. Gene pairs were directly connected by zero to very weak (digenic interaction score: 0 to −0.08, n = 74 strains), weak (−0.08 to −0.1, n = 32), or moderate (<−0.1, n = 45) negative digenic interactions. (2) Number of digenic interactions. Gene pairs had a low (10 to 45 interactions, n = 50), intermediate (46 to 70, n = 53), or high (>71, n = 48) average digenic interaction degree (denoted by the number of black edges of each node). (3) Digenic interaction profile similarity. Gene pairs had low (score: −0.02 to 0.03, n = 46; represented by genes A and B, which show a relatively low overlap of genetic interactions with genes K to R), intermediate (0.03 to 0.1, n = 59; represented by genes C and D, which display an intermediate overlap of genetic interactions), or high (>0.1, n = 46, represented by genes E and F, which display a relatively high level of overlap of genetic interactions) functional similarity, as measured by digenic interaction profile similarity and coannotation to the same GO term(s). Query mutant genes were either nonessential deletion mutant alleles (Δ) or conditional temperature-sensitive (ts) alleles of essential genes. (**B**) Diagram illustrating the triple-mutant SGA experimental strategy. To quantify a trigenic interaction, three types of screens are conducted in parallel. To estimate triple-mutant fitness, a double-mutant query strain carrying two desired mutated genes of interest (red and blue filled circles) is crossed into a diagnostic array of single mutants (black filled circle). Meiosis is induced in heterozygous triple mutants, and haploid triple-mutant progeny is selected in sequential replica pinning steps. In parallel, single-mutant control query strains are used to generate double mutants for fitness analysis. (**C**) Triple-mutant SGA quantitative scoring strategy. The top equation shows the quantification of a digenic interaction, where ε_{ij} is the digenic interaction score, f_{ij} is the observed double-mutant fitness, and the expected double-mutant fitness is expressed as the product of single-mutant fitness estimates $f_i f_j$. In the bottom equation, the trigenic interaction score (τ_{ijk}) is derived from the digenic interaction score, where f_{ijk} is the observed triple-mutant fitness and $f_i f_j f_k$ is the triple-mutant fitness expectation expressed as the product of three single-mutant fitness estimates. The influence of digenic interactions is subtracted from the expectation, and each digenic interaction is scaled by the fitness of the third mutation.

genetic interactions is often more accurate than that for positive interactions because there is a greater signal-to-noise ratio for negative genetic interactions. Hence, negative genetic interactions are associated with lower false-positive and false-negative rates than positive interactions (*8*), a feature that is important for the robust statistical analysis necessary to differentiate true trigenic interactions from the extensive background digenic network. Second, negative digenic interactions are generally more functionally informative than positive digenic interactions (*8*), and thus the large-scale mapping of a negative trigenic interaction network is expected to provide the most mechanistic insight into gene function and pathway wiring.

Trigenic interactions are enriched for functionally related genes

To obtain sufficient precision, we carried out each analysis, which involved screening the individual query genes for digenic interactions and the double-mutant query for trigenic interactions, in at least two replicate screens with four colonies per screen (fig. S5). In total, we tested 410,399 double and 195,666 triple mutants for fitness defects, meeting a previously established intermediate magnitude cutoff (*15*) (data S2) and identified 9363 digenic and 3196 trigenic negative interactions. From detailed validation of trigenic interactions of our *CLN1-CLN2* double-mutant query, which was screened previously (*19*), we estimated a false-negative rate of ~40%, a false-positive rate of ~20%, and a true-positive rate between ~60 and ~75% (table S2 and fig. S6) (*16*), which is consistent with our previous global digenic network analysis (*8*).

The distribution of trigenic interaction degree for array strains shows that the majority of low-degree genes (70%) account for ~88% of all trigenic interactions, whereas highly connected genes contribute the remaining ~12% of interactions. Thus, the trigenic interactions are not associated with a small set of highly connected genes; rather, the interactions are distributed across many different genes (fig. S7). On the other hand, with a smaller, more biased set of double-mutant query genes, the distribution of trigenic interaction degree shows that ~22% of them accounted for 51% of trigenic interactions, indicating that a particular subset of the digenic queries were enriched for trigenic interactions (fig. S7). About one-third of the newly mapped trigenic interactions identified connections that were not observed in our digenic control network; we refer to these as "novel" trigenic interactions. The remaining approximately two-thirds of the trigenic interactions overlapped a digenic interaction while still exhibiting a stronger than expected fitness defect in the triple mutant; these we refer to as "modified" trigenic interactions (fig. S8A). Thus, although a substantial fraction of trigenic interactions elucidate totally new functional information, the majority of the trigenic interactions we mapped expand upon the digenic interaction network.

We first assessed the functional information embedded in the trigenic network by comparing the distributions of digenic and trigenic interactions across different biological processes. As observed previously (*15*), digenic interactions were enriched among genes annotated to the same biological process and, although the magnitude of trigenic interaction enrichment was somewhat lower, they were comparably enriched for genes within the same bioprocess (Fig. 2A). We also evaluated the enrichment of digenic and trigenic interactions across common functional standards, including annotation to the same Gene Ontology (GO) biological process, subcellular-localization pattern, protein-protein interaction, and gene coexpression (Fig. 2B). Like digenic interactions (*7*, *15*), genes involved in trigenic interactions were significantly enriched for all of these standards, with genes participating in the "modified" class of trigenic interactions exhibiting stronger functional relationships (fig. S8B) as well as a stronger magnitude of interactions (fig. S8C). Thus, trigenic interactions resemble digenic interactions in that they are rich in functional information, which means that genes participating in many trigenic interactions can be predicted from alternative data sets and general knowledge of cellular function.

Trigenic interactions expand functional connections mapped by the global digenic network

Our functional analysis revealed that trigenic interactions have some properties distinct from those of digenic interactions, which suggests that trigenic interactions may be useful for discovering previously unknown connections between genes and their corresponding pathways. As an illustrative example, we examined the *MDY2-MTC1* double-mutant query, which is a highly connected hub within the trigenic network. *MDY2* encodes a protein that interacts and functions with components of the GET (guided entry of tail-anchor) pathway (*20*), which is important for Golgi–to–endoplasmic reticulum (ER) trafficking and inserting tail-anchored proteins into ER membranes. *MTC1* encodes a protein of unknown function that localizes to the early Golgi apparatus. The *MTC1* digenic interaction profile is similar to that of *USO1*, which is involved in vesicle-mediated ER-to-Golgi transport (*21*), and *RUD3*, which encodes a Golgi matrix protein important for the structural organization of the cis-Golgi (*22*), suggesting that *MTC1* also has a role in the early secretory pathway.

As expected from these previous findings, the *MDY2* and *MTC1* digenic interactions identified in our screen were enriched for genes involved in cell polarity and the early secretory pathway (Fig. 3, fig. S9A, and data S6). However, the *MDY2-MTC1* double-mutant query profile encompassed a much more functionally diverse set of genes (Fig. 3). For example, although we observed novel trigenic interactions with ER-to-Golgi transport genes, we also observed trigenic interactions with genes involved in other modes of vesicle

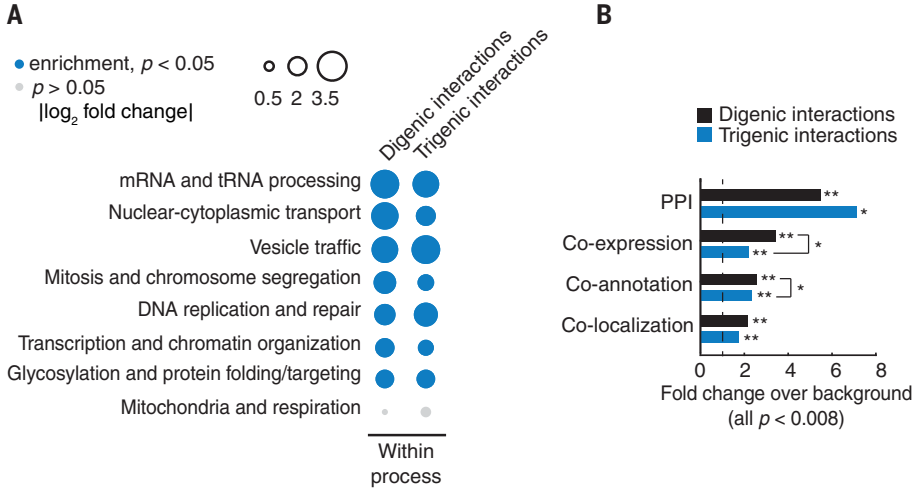


Fig. 2. Functional characterization of trigenic interactions. (**A**) Frequency of negative genetic interactions within biological processes. For our analysis, we used the fraction of screened query-array combinations exhibiting negative interactions belonging to functional gene sets annotated by SAFE (spatial analysis of functional enrichment) on the global genetic interaction network (*55*). The "within process" category received a count for any combination in which both genes for digenic interactions or all three genes for trigenic interactions were annotated to the same term. The size of the circle assigned to each "within process" element reflects the fold increase over the background fraction of interactions (digenic = 0.023, trigenic = 0.016). Significance was assessed with a hypergeometric test; $P < 0.05$. Blue circles represent significant enrichment; gray circles denote no significant change. (**B**) Enrichment of negative digenic and trigenic interactions across four functional standards. The dashed line indicates no enrichment. The functional standards are merged protein-protein interaction (PPI) (*56*–*60*), coannotation (based on SAFE terms) (*7*), coexpression (*61*), and colocalization (*62*). Significance was assessed with a hypergeometric test; * represents $10^{-4} \leq P < 0.01$, ** represents $P < 10^{-4}$.

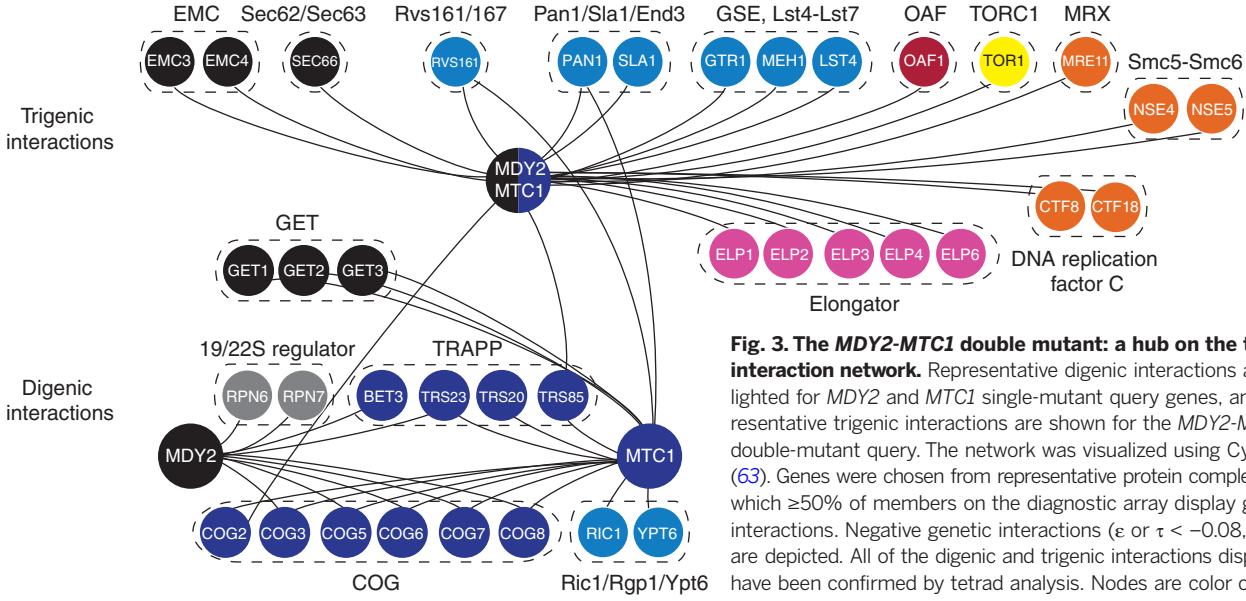


Fig. 3. The *MDY2-MTC1* double mutant: a hub on the trigenic interaction network. Representative digenic interactions are highlighted for *MDY2* and *MTC1* single-mutant query genes, and representative trigenic interactions are shown for the *MDY2-MTC1* double-mutant query. The network was visualized using Cytoscape (*63*). Genes were chosen from representative protein complexes (*8*) in which ≥50% of members on the diagnostic array display genetic interactions. Negative genetic interactions (ε or $\tau < -0.08$, $P < 0.05$) are depicted. All of the digenic and trigenic interactions displayed have been confirmed by tetrad analysis. Nodes are color coded on the basis of their biological roles and are labeled with gene names. Genes are grouped according to specific protein complexes.

trafficking, including endocytosis and peroxisome biology. Moreover, *MDY2-MTC1* trigenic interactions identified connections to genes with more diverse functions, such as components of the elongator complex (Fig. 3), which controls the modification of wobble nucleosides in tRNAs, and several genes involved in DNA replication and repair. Notably, the *MDY2-MTC1* query also showed a trigenic interaction with *TOR1* (*target of rapamycin 1*), which encodes the key kinase subunit of the TORC1 complex that is required for growth in response to nutrients by regulating ribosome biogenesis, nutrient transport, and autophagy (*23*). Consistent with this observation, the *MDY2-MTC1* trigenic interaction network captures a set of genes that have a dual role in TORC1 signaling and sorting of the general amino acid permease, including *GTR1*, *MEH1*, and *LST4* (*24*–*26*).

The spectrum of bioprocesses that are represented in genetic interaction profiles can be visualized by mapping functional enrichment within the context of the global yeast digenic interaction profile similarity network, which clusters genes into 17 distinct bioprocesses (*7*, *27*) (Fig. 4A). In comparison with the *MDY2* and *MTC1* digenic interaction profiles (Fig. 4, B and C), the *MDY2-MTC1* trigenic interactions were enriched not only for vesicle trafficking and cell polarity bioprocess regions of the network but also in regions encompassing genes annotated to the tRNA wobble modification bioprocess, DNA replication and repair, as well as mitosis and chromosome segregation (Fig. 4D). Thus, the *MDY2-MTC1* trigenic interaction profile exhibited a more expanded and functionally diverse set of connections than either of the corresponding *MDY2* or *MTC1* digenic interaction profiles.

We used a variety of assays to test three functional connections revealed by the *MDY2-MTC1* trigenic interaction profile. First, although the *MDY2-MTC1* double-mutant strain did not show an exaggerated cell biological phenotype associated with the early trafficking function (fig. S9B), it displayed a marked synthetic sick phenotype when combined with deletion of *SLA1*, which is involved in cortical actin assembly and endocytic vesicle formation, which translates into an extended Sla1 patch lifetime, reflecting a defect in endocytosis (Fig. 5A) (*28*). Second, given a negative trigenic interaction with *OAF1*, which encodes an oleate-activated transcription factor involved in peroxisome organization and biogenesis (*29*), we used fluorescence microscopy to explore peroxisome morphology. The *MDY2-MTC1* double mutant displayed an accumulation of relatively small peroxisomes (Fig. 5B), which may be indicative of a defect in ER-derived peroxisome membrane biogenesis (*30*). Third, the *MDY2-MTC1* double mutant showed pronounced sensitivity to hydroxyurea (HU) but not methyl methanesulfonate (MMS), which is consistent with a specific defect in DNA replication and reflects the negative genetic interactions we observed with a number of DNA replication and repair genes, including *NSE4* and *NSE5*, which encode components of the Smc5-Smc6 complex that mediates resolution of DNA structures spanning sister chromatids (Fig. 5C and fig. S10, A to C) (*31*). We suspect that the *MDY2-MTC1* double mutant may be primarily defective in trafficking functions that can modulate signaling or metabolic pathways and thereby influence DNA synthesis and repair pathways indirectly (fig. S10, D to H).

Trigenic interaction profiles are more functionally diverse than their corresponding digenic profiles

To test the generality of whether query genes connect to more functionally divergent genes through trigenic interactions than through digenic interactions, we compared digenic profile similarity of pairs of genes spanned by either digenic or trigenic interactions. Indeed, genes involved in trigenic interactions tend to show profiles that are less similar than those connected by digenic interactions, which suggests that they are less functionally related than those connected by digenic interactions (Fig. 6A and fig. S11A). We also found that trigenic interactions were more enriched than digenic interactions for connections that bridge several different biological processes, including mRNA and tRNA processing, vesicle trafficking, mitosis and chromosome segregation, and glycosylation and protein folding and targeting (Fig. 6B). Moreover, as we showed for the *MDY2-MTC1* double-mutant query (Fig. 4), trigenic interaction profiles were generally enriched for genes spanning more diverse bioprocesses than the corresponding digenic interaction profiles (Fig. 6C and fig. S11, B to D). Genes involved in vesicle trafficking were particularly enriched for trigenic interactions occurring between bioprocesses (Fig. 6B and figs. S7 and S11D). As we observed for *MDY2-MTC1*, other double-mutant queries carrying mutations in genes implicated in membrane trafficking were enriched for trigenic interactions with genes involved in DNA replication and repair machinery, which may indicate a general connection between these two bioprocesses. For example, the digenic query strain *MVP1-MRL1*, which carries mutations in genes required for sorting proteins to the vacuole (*32*, *33*), and the strain *SEC27-GET4*, which carries mutations in genes involved in ER-to-Golgi transport (*34*) and the insertion of tail-anchored proteins into ER membrane (*20*), both exhibited an enrichment of trigenic interactions with DNA replication and repair machinery (fig. S11D). In general, our findings show that trigenic interaction profiles are composed of connections involving genes

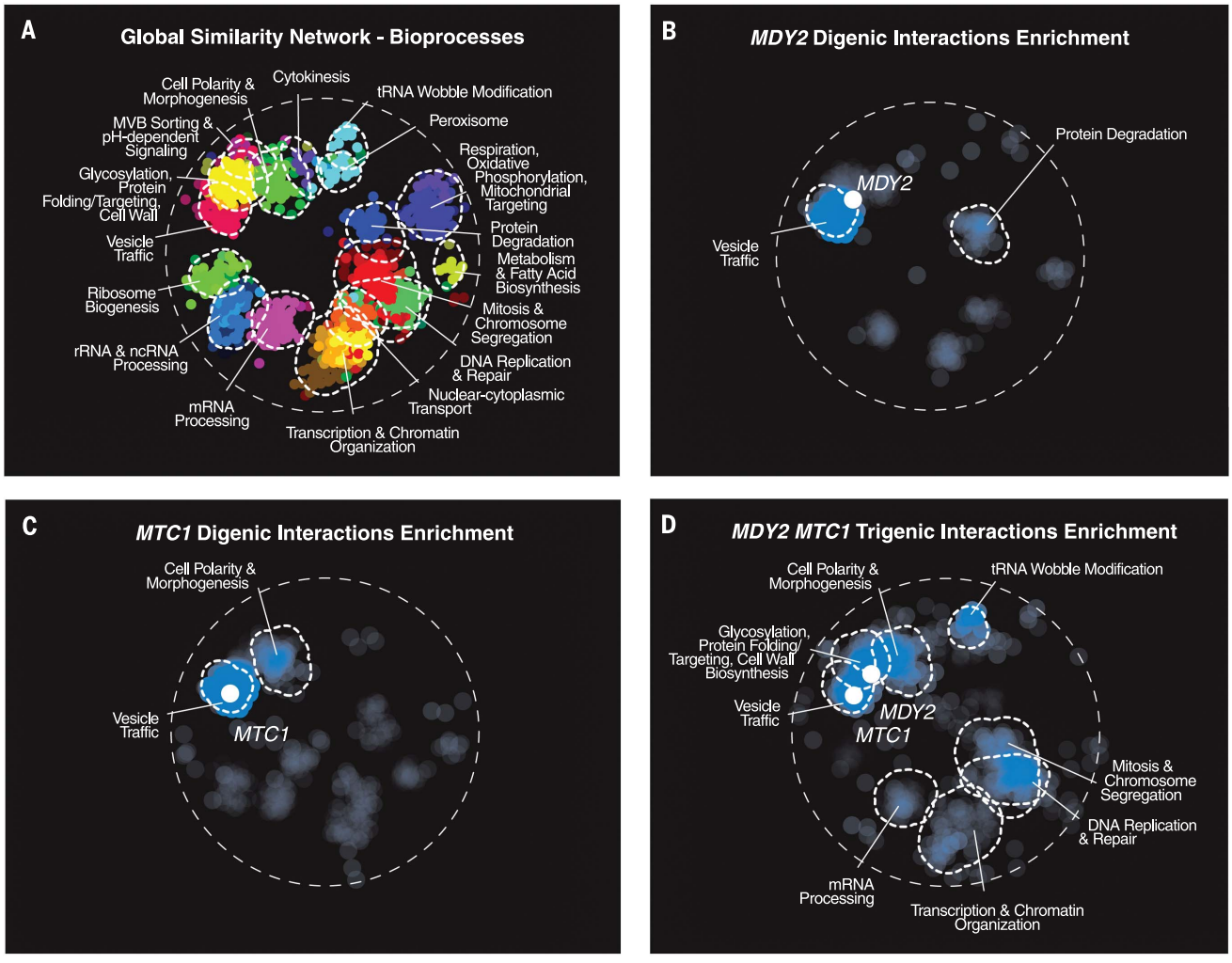


Fig. 4. Enrichment of genetic interactions within bioprocesses defined by a global network of digenic interaction profile similarities. (**A**) The global digenic interaction profile similarity network (*7*) was annotated using SAFE (*55*), identifying network regions enriched for similar GO biological process terms as outlined by dashed lines. rRNA, ribosomal RNA; ncRNA, noncoding RNA; MVB, multivesicular body. (**B**) *MDY2* digenic interactions showing bioprocess enrichments. (**C**) *MTC1* digenic interactions showing bioprocess enrichments. (**D**) *MDY2-MTC1* trigenic interactions showing bioprocess enrichments.

that are more functionally diverse than their corresponding digenic interaction profiles (Fig. 6C). However, despite their higher tendency to connect diverse processes, a significant fraction of trigenic interactions occurs among genes within the same bioprocess ($P < 1 \times 10^{-16}$; hypergeometric test) (Fig. 2A).

Gene features of trigenic interactions and the expanse of the global trigenic landscape

Having selected the query gene pairs based on the properties of the global digenic interaction network, we can assess how these properties relate to trigenic interaction frequency. The strongest correlation with the number of observed trigenic interactions for each gene pair was the digenic genetic interaction profile similarity (correlation coefficient $r = 0.41$, $P = 1.2 \times 10^{-7}$), followed by the average number of digenic interactions of the query genes ($r = 0.25$, $P = 1.9 \times 10^{-3}$) and the strength of a direct negative genetic interaction between the query gene pair ($r = 0.23$, $P = 5.4 \times 10^{-3}$) (fig. S12, A to C). Thus, numerous trigenic interactions were observed for functionally related query genes, which display overlapping profiles on the digenic similarity network and often show a digenic interaction with each other (*7*) (Fig. 7A). As observed for digenic interactions (*7*), the frequency of trigenic interactions was highly correlated with the fitness defect of the double-mutant query strain (fig. S13). Consistent with this observation, essential genes exhibited high connectivity on the trigenic interaction network. A double-mutant query that carries at least one temperature-sensitive allele of an essential gene, which is often associated with a fitness defect at the semipermissive screening temperature, exhibited more genetic interactions than a query deleted for a pair of nonessential genes ($P = 0.035$) (Fig. 7B). More generally, query genes that are highly connected on the digenic network are also highly connected on the trigenic network (fig. S12D).

Notably, trigenic interactions tend to be ~25% weaker than digenic interactions ($P < 1.7 \times 10^{-98}$) (Fig. 7C), which means the average digenic interaction often has a more profound phenotype than the average trigenic interaction. However, to fully understand the potential for trigenic interactions to drive fitness defects, we also need to estimate the frequency at which they occur. Because we have mapped digenic interactions comprehensively and we know the false-positive and false-negative rates associated with this analysis, we can estimate the number of digenic interactions within the yeast genome, revealing a distribution that centers on ~6×10^5 total negative interactions (Fig. 7D) (*16*). Furthermore, because digenic interaction properties are predictive of trigenic interaction degree, we can also extrapolate our findings to estimate the number of negative trigenic interactions across

the whole genome. As noted earlier, we selected gene pairs for trigenic analysis to fill bins of varying attributes, including double-mutant queries with either weak or strong interactions, as well as those with either sparse or rich genetic interaction profiles, as depicted schematically in Figs. 1A and 7A (table S1). For extrapolation to the whole genome, we used the mean trigenic interaction degree of double mutants in a given bin as the expected degree for any hypothetical pair from the genome with similar characteristics (data S7 and fig. S14). Integrating this value across the total number of gene pairs with the given characteristics, which preserves the double-mutant distribution across different digenic interaction features, and summing over all bins yielded an estimate of the total number of trigenic interactions.

In the yeast genome, there are 2000 times as many possible triple gene combinations (36 billion) as possible gene pairs (18 million), but the density of interactions (as both observed and extrapolated) is similar, reduced by only a factor of ~3 for trigenic interactions (table S3). We predict that ~10^8 trigenic combinations exhibit a negative genetic interaction, generating a conservative estimate of on the order of 100 times as many trigenic interactions as observed for the global digenic network (Fig. 7D and table S3). To establish confidence intervals (CIs) for the estimate, we repeated the extrapolation process with 10,000 bootstrapped samplings of the 151 double-mutant query pairs, keeping their associated trigenic interactions degrees and the corresponding digenic interaction features constant. The bin with the lowest digenic interaction degree encompasses a large fraction of the potential double mutants in the genome and is assigned a low trigenic interaction degree, which means that the summarized estimate provided is likely a conservative underestimate. Moreover, because our binning scheme restricts our extrapolation to ~25% of the potential trigenic interaction space (e.g., by omitting potential double-mutant queries that show a positive digenic interaction), we are underestimating its extent, and the true number of trigenic interactions is likely to be several times higher (fig. S15 and table S3) (*16*). The vast expanse of the global trigenic interaction network points to the potential for higher-order interactions to affect all aspects of the genetics of outbred populations, including the genotype-to-phenotype relationship.

Discussion

Systematic mapping of trigenic interactions revealed that their properties resemble those of digenic interactions because they often connect functionally related genes, which means that trigenic interactions have the potential to contribute to our understanding of the functional wiring diagram of the cell. The global digenic network is predictive of trigenic interactions because query gene pairs showing properties associated with shared functionality, such as overlapping digenic interaction profiles or a negative digenic genetic interaction, often map numerous trigenic interactions (Fig. 7A). Thus, if two query genes are in the same or similar bioprocess cluster on the global digenic profile similarity network (Fig. 4A), they will likely show a rich trigenic interaction profile, as we observed for the *MDY2-MTC1* double mutant query (Figs. 3 and 4D). Gene essentiality and the average digenic interaction degree of the query gene pair were also correlated with trigenic connectivity (Fig. 7B), indicating that highly connected hubs are consistent on both the digenic and trigenic interaction networks (fig. S12, C and D).

Many of the trigenic interactions we observed overlapped with at least one digenic interaction. In some cases, we chose query gene pairs displaying a negative genetic interaction and so all of the trigenic interactions in these profiles accentuated the query interaction (fig. S8). Moreover, a substantial proportion of trigenic interactions measured for noninteracting query pairs exacerbated a digenic interaction that was previously seen between one or both of the query genes and the third gene (fig. S8). Thus, our findings show that negative trigenic interactions often highlight the potential for a third mutation to amplify the phenotype of a digenic interaction. Analogously, in human genetics, the variation in an individual's genetic background can have profound influence on the penetrance of the phenotype associated with a disease gene (*35*).

Although we found that trigenic interactions tend to be slightly weaker than digenic interactions, they are ~100 times more numerous and are more functionally diverse than their digenic counterparts. The expanse of possible three-gene combinations makes exhaustive trigenic interaction mapping intractable with our current

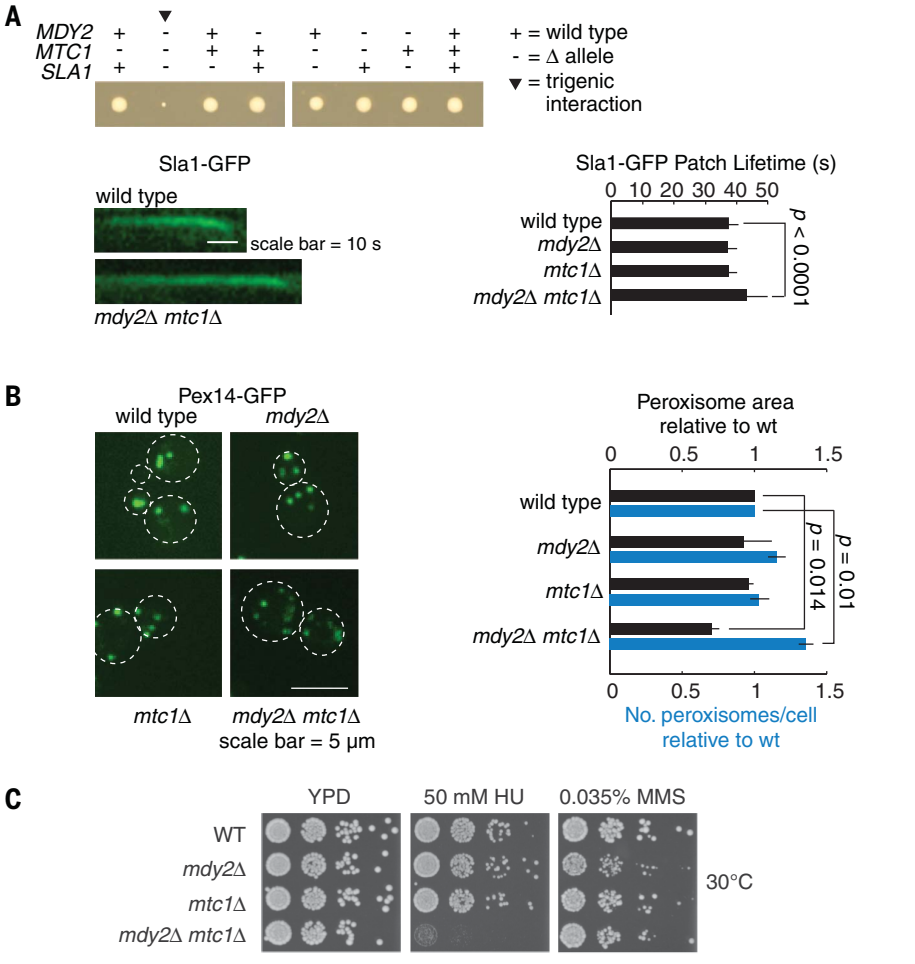


Fig. 5. Trigenic interactions reflect the physiology of the *MDY2-MTC1* double-mutant query strain. (**A**) Endocytic membrane trafficking is impaired in the *mdy2*Δ-*mtc1*Δ double-mutant query strain. (Top) Example of tetrad analysis confirmations for the *mdy2*Δ-*mtc1*Δ-*sla1*Δ triple-mutant strain. (Bottom left) Endocytic uptake dynamics were examined with the Sla1–green fluorescent protein (GFP) reporter. Representative kymographs are displayed for the wild type and the *mdy2*Δ-*mtc1*Δ double mutant. (Bottom right) Lifetime of Sla1-GFP endocytic vesicle formation was quantified across ~100 different patches in two independent experiments. Error bars represent SD. (**B**) Peroxisome biogenesis was monitored in the wild type (wt) and in *mdy2*Δ, *mtc1*Δ, and *mdy2*Δ-*mtc1*Δ mutants using Pex14p-GFP reporter. (**C**) Growth response to HU and MMS for the wild type and *mdy2*Δ, *mtc1*Δ, and *mdy2*Δ-*mtc1*Δ mutants. YPD, yeast extract, peptone, and dextrose.

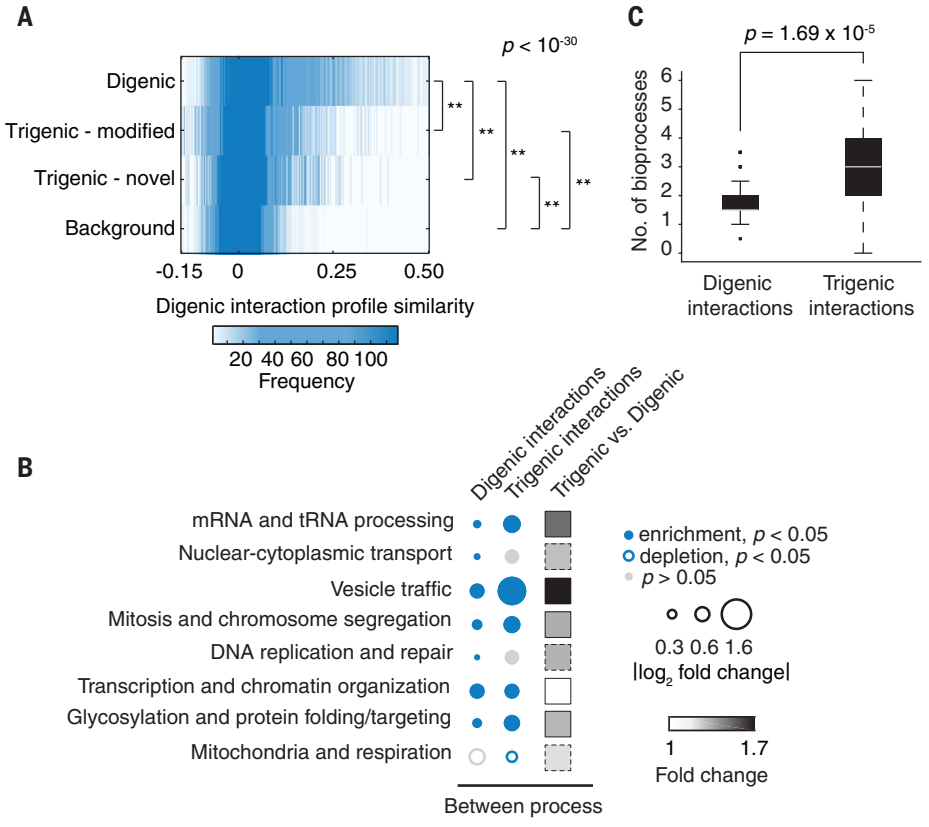


Fig. 6. Trigenic interactions are more functionally distant than digenic interactions. (**A**) Distribution of genetic interaction profile similarities of genes showing digenic and trigenic interactions. *P* values are based on a Wilcoxon rank sum test; $^{**}P < 10^{-30}$. (**B**) Frequency of negative genetic interactions between biological processes using SAFE annotations for digenic and trigenic interactions (*55*). The size of the circle assigned to each "between process" element reflects the fold increase over the background fraction of interactions (digenic = 0.023, trigenic = 0.016); $P < 0.05$ based on a hypergeometric test. The "between process" category received a count for any combinations that were not counted in the "within process" category shown in Fig. 2A. Filled blue circles represent significant enrichment, the open blue circle represents significant underenrichment, and gray circles denote no significant change. Trigenic versus digenic fold change (the ratio of trigenic interaction enrichment to digenic interaction enrichment) is represented by filled squares (black is maximal fold change; white is no fold change). In cases for which the "between process" enrichment was observed but is not significant ($P < 0.05$), the square is outlined with a dashed line. (**C**) Number of SAFE bioprocess clusters enriched for digenic or trigenic interactions.

methodology. However, the substantial overlap of the digenic and trigenic networks indicates that the genetic landscape of the cell expands with higher-order genetic interactions but does not change drastically in terms of its functional modularity. Thus, the global digenic network is highly informative of potential trigenic interactions and can be used effectively to predict candidate query gene pairs for efficient trigenic interaction analysis.

Trigenic interaction data may inform a wide variety of subjects within biology. For example, the number, magnitude, and properties of digenic and trigenic interactions clarifies aspects of speciation theory in evolutionary biology (*36*, *37*). Hybrids between species exhibit reduced fitness, which is usually attributed to negative (epistatic) interactions among genes that diverged in isolated populations. Each population may evolve fixed variants that are neutral or adaptive in their own genetic backgrounds. When these variants are brought together for the first time in hybrid genomes, they may cause deleterious genetic interactions, also termed Dobzhansky-Muller incompatibilities. As populations diverge from one another, the number of potential digenic interactions increases as the square of the number of substitutions, the so-called "snowball effect" (*36*, *38*, *39*). That is, each subsequent substitution in a distinct population has the potential to interact with any substitution from the other population (and vice versa), and thus the probability of a speciation event grows with each step. Most speciation genetics research has focused on these digenic interactions. However, the number of trigenic combinations accumulates exponentially faster than the number of digenic combinations. Both digenic and trigenic interactions have been implicated in speciation (*40*, *41*), but the general extent to which digenic or complex negative genetic interactions drive speciation remains unknown. If digenic interactions do, in fact, play a major role in orchestrating speciation, then either the frequency and/or the strength of deleterious trigenic interactions must be relatively smaller than that of digenic interactions. Our systematic analysis shows that trigenic interactions are somewhat less likely to occur (by a factor of 3; ~3% versus ~1% for digenic versus trigenic, respectively) and generally weaker (~25% weaker) than digenic interactions. Nevertheless, modeling based on our findings suggests that trigenic interactions are substantially common and often strong enough to play a key role in the evolution of hybrid inviability (fig. S16 and table S4) (*16*). Because the connections associated with higher-order interactions may often overlap with those of simpler interactions and because those simpler interactions require fewer substitutions and will often manifest first, our findings may also suggest that the evolution of even more highly complex interactions may be limited, even though their absolute numbers increase exponentially, a possibility that is consistent with evolutionary theory (*38*).

Our trigenic interaction study is also relevant to synthetic biology efforts aimed at efficient synthesis (*42*, *43*) and design of minimal genomes (*44*, *45*). Digenic synthetic lethal interactions were recently noted as a major constraint in the design of the minimal genome for the bacteria *Mycoplasma mycoides*, for which a viable genome could be constructed only after resolving lethal interactions that arose between nonessential genes (*44*). For species in which systematic gene perturbation studies have been conducted, the proportion of essential genes is relatively small (e.g., ~20% in yeast, ~10% in human cells, which increases to 20% when only expressed genes are considered) (*13*, *46*–*48*). However, we expect that digenic and trigenic interactions will dictate much larger minimal genomes than the essential gene set, even for growth under simple laboratory conditions. With the complete digenic network (*7*), we estimate that the minimal yeast genome would encompass more than ~70% of genes after accounting for digenic interactions (table S5) (*16*). With the inclusion of constraints imposed by trigenic interactions, we expect that a minimal genome, without a substantial fitness defect, may nearly approach the complete set of genes encoded in the genome. Thus, genetic interactions may help to explain the large gap between the number of genes with strong individual fitness defects and the total genome size, and the prevalence of yeast negative trigenic interactions suggests that many genomes lack the potential for substantial compression while maintaining normal fitness.

It is important to consider other types of genetic interactions in addition to those associated with severe loss-of-function alleles due to entire open reading frame deletions of nonessential genes or temperature-sensitive alleles of essential

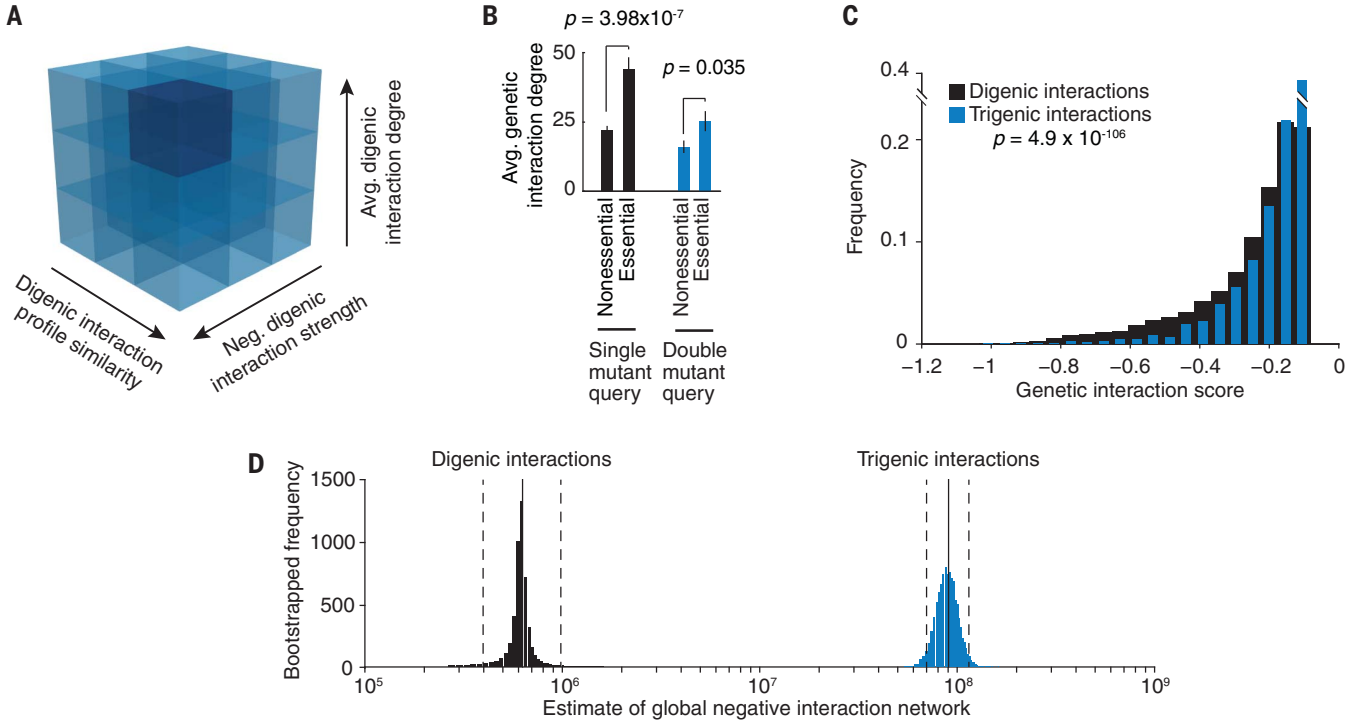


Fig. 7. Relation of digenic and trigenic interaction networks. (**A**) Trigenic interaction degree distribution correlated with three quantitative features of genes on the digenic interaction network: (i) Interaction profile similarity of the two genes in the double-mutant query gene pair (bin thresholds: −0.02, 0.03, 0.1, +∞), which generates three bins for average digenic interaction profile similarity (r): $-0.02 < r < 0.03$; $0.03 \le r < 0.1$; $0.1 \le r$. (ii) Negative digenic interaction strength associated with the double-mutant query gene pair (bin thresholds: 0, −0.08, −0.1, −∞), which generates three bins for digenic interaction score (ε): $\varepsilon < -0.1$; $-0.1 \le \varepsilon < -0.08$; $-0.08 \le \varepsilon < 0$. (iii) Average digenic interaction degree, which represents the average number of negative genetic interactions associated with each of the genes of the double-mutant query gene pair (bin thresholds: 10, 45, 70, +∞), which generates three bins for average digenic interaction degree: $10 \le$ degree < 45; $45 \le$ degree <70; $70 \le$ degree. The bin with the highest average negative trigenic interaction degree at the intermediate interaction score cutoff ($\tau < -0.08$) of 63.5 is shown in dark blue. (**B**) Essentiality determines trigenic interaction degree. Number of single mutants: 254 nonessential genes, 47 essential genes. Number of double mutants: 111 nonessential gene pairs, 40 essential or mixed essentiality gene pairs. Mean genetic interaction is represented; error bars indicate SEM; P values are based on a t test. Negative genetic interactions (ε or $\tau < -0.08$, $P < 0.05$) are depicted. (**C**) Cumulative distribution of negative digenic and trigenic interaction score magnitudes. Pairwise significance was assessed with a Wilcoxon rank sum test. (**D**) Estimates of the number of digenic and trigenic interactions at the intermediate score cutoff (ε or $\tau < -0.08$, $P < 0.05$). Bootstrapping was used to generate the estimate by sampling 10,000 times with replacement. Dashed lines indicate the 95% CIs; solid lines denote the estimated extent of the trigenic interaction landscape. This conservative estimate of the total number of trigenic interactions in the yeast genome covers ~26% of the interaction space. For the total genome-wide estimate, see fig. S15B and table S3.

genes. Our analysis revealed that double mutants with strong fitness defects often show rich trigenic interaction profiles (fig. S13C). Similarly, single-mutant fitness defects also correlate with digenic interaction degree (fig. S13A) (*7*). Presumably, the weaker fitness effects associated with the variation found in natural populations may require higher-order combinations, involving more than two genes, to influence trait heritability through genetic interactions (*49*). In the yeast model system, genetic interactions were found to play an important role in the heritability of a number of different quantitative traits, possibly with a greater contribution made by digenic interactions versus higher-order interactions (*4*, *5*, *50*). The genetic mechanism underlying conditional essentiality, in which a given yeast gene is nonessential in one genetic background but essential in another, often appears to be associated with a complex set of modifier loci (*49*), as do a number of other traits (*51*, *52*). Thus, both digenic and higher-order interactions are established components of the genetic architecture of yeast complex traits, and similar findings have been made in a number of other organisms (*53*). In part because model organism populations have allele distributions that differ from those in humans, the degree to which higher-order genetic interactions will contribute to the genetics of complex human disease remains to be seen (*50*, *54*). Nevertheless, the extensive landscape of trigenic interactions revealed here for yeast, as well as their capacity for generating functionally diverse phenotypes and driving speciation, suggests that higher-order genetic interactions may play a key role in the genotype-to-phenotype relationship.

Materials and methods summary

The supplementary materials contain a detailed description of materials and methods for the construction of yeast single-, double-, and triple-mutant strains as well as quantification of genetic interactions and any associated analyses. General methodological information and references to specific sections appear throughout the text.

REFERENCES AND NOTES

1. J. L. Hartman IV, B. Garvik, L. Hartwell, Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001). doi: 10.1126/science.291.5506.1001; pmid: 11232561
2. 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
3. O. Zuk, E. Hechter, S. R. Sunyaev, E. S. Lander, The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1193–1198 (2012). doi: 10.1073/pnas.1119675109; pmid: 22223662
4. J. S. Bloom, I. M. Ehrenreich, W. T. Loo, T. L. Lite, L. Kruglyak, Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–237 (2013). doi: 10.1038/nature11867; pmid: 23376951
5. J. S. Bloom *et al.*, Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nat. Commun.* **6**, 8712 (2015). doi: 10.1038/ncomms9712; pmid: 26537231

6. A. H. Tong *et al*., Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001). doi: 10.1126/science.1065810; pmid: 11743205
7. M. Costanzo *et al*., A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016). doi: 10.1126/science.aaf1420; pmid: 27708008
8. A. Baryshnikova *et al*., Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7**, 1017–1024 (2010). doi: 10.1038/nmeth.1534; pmid: 21076421
9. P. Novick, D. Botstein, Phenotypic analysis of temperature-sensitive yeast actin mutants. *Cell* **40**, 405–416 (1985). doi: 10.1016/0092-8674(85)90154-0; pmid: 3967297
10. A. Bender, J. R. Pringle, Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 1295–1305 (1991). doi: 10.1128/MCB.11.3.1295; pmid: 1996092
11. J. van Leeuwen *et al*., Exploring genetic suppression interactions on a global scale. *Science* **354**, aag0839 (2016). doi: 10.1126/science.aag0839; pmid: 27811238
12. E. A. Winzeler *et al*., Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999). doi: 10.1126/science.285.5429.901; pmid: 10436161
13. G. Giaever *et al*., Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002). doi: 10.1038/nature00935; pmid: 12140549
14. A. H. Tong *et al*., Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004). doi: 10.1126/science.1091317; pmid: 14764870
15. M. Costanzo *et al*., The genetic landscape of a cell. *Science* **327**, 425–431 (2010). doi: 10.1126/science.1180823; pmid: 20093466
16. Materials and methods are available as supplementary materials.
17. R. Deshpande, J. Nelson, S. W. Simpkins, M. Costanzo, J. S. Piotrowski, S. C. Li, C. Boone, C. L. Myers, Efficient strategies for screening large-scale genetic interaction networks. bioRxiv 159632 [Preprint]. 5 July 2017. doi: 10.1101/159632
18. J. E. Haber *et al*., Systematic triple-mutant analysis uncovers functional connectivity between pathways involved in chromosome regulation. *Cell Rep.* **3**, 2168–2178 (2013). doi: 10.1016/j.celrep.2013.05.007; pmid: 23746449
19. J. Zou *et al*., Regulation of cell polarity through phosphorylation of Bni4 by Pho85 G1 cyclin-dependent kinases in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **20**, 3239–3250 (2009). doi: 10.1091/mbc.E08-12-1255; pmid: 19458192
20. M. C. Jonikas *et al*., Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009). doi: 10.1126/science.1167983; pmid: 19325107
21. Y. Noda, T. Yamagishi, K. Yoda, Specific membrane recruitment of Uso1 protein, the essential endoplasmic reticulum-to-Golgi tethering factor in yeast vesicular transport. *J. Cell. Biochem.* **101**, 686–694 (2007). doi: 10.1002/jcb.21225; pmid: 17192843
22. A. K. Gillingham, A. H. Tong, C. Boone, S. Munro, The GTPase Arf1p and the ER to Golgi cargo receptor Erv14p cooperate to recruit the golgin Rud3p to the cis-Golgi. *J. Cell Biol.* **167**, 281–292 (2004). doi: 10.1083/jcb.200407088; pmid: 15504911
23. A. González, M. N. Hall, Nutrient sensing and TOR signaling in yeast and mammals. *EMBO J.* **36**, 397–408 (2017). doi: 10.15252/embj.201696010; pmid: 28096180
24. A. van der Zand, J. Gent, I. Braakman, H. F. Tabak, Biochemically distinct vesicles from the endoplasmic reticulum fuse to form peroxisomes. *Cell* **149**, 397–409 (2012). doi: 10.1016/j.cell.2012.01.054; pmid: 22500805
25. S. Kira *et al*., Dynamic relocation of the TORC1-Gtr1/2-Ego1/2/3 complex is regulated by Gtr1 and Gtr2. *Mol. Biol. Cell* **27**, 382–396 (2016). doi: 10.1091/mbc.E15-07-0470; pmid: 26609069
26. M. P. Péli-Gulli, A. Sardu, N. Panchaud, S. Raucci, C. De Virgilio, Amino acids stimulate TORC1 through Lst4-Lst7, a GTPase-activating protein complex for the Rag family GTPase Gtr2. *Cell Rep.* **13**, 1–7 (2015). doi: 10.1016/j.celrep.2015.08.059; pmid: 26387955
27. M. Usaj *et al*., TheCellMap.org: A web-accessible database for visualizing and mining the global yeast genetic interaction network. *G3* **7**, 1539–1549 (2017). doi: 10.1534/g3.117.040220; pmid: 28325812
28. Y. Sun, N. T. Leong, T. Wong, D. G. Drubin, A Pan1/End3/Sla1 complex links Arp2/3-mediated actin assembly to sites of clathrin-mediated endocytosis. *Mol. Biol. Cell* **26**, 3841–3856 (2015). doi: 10.1091/mbc.E15-04-0252; pmid: 26337384
29. I. V. Karpichev, G. M. Small, Global regulatory functions of Oaf1p and Pip2p (Oaf2p), transcription factors that regulate genes encoding peroxisomal proteins in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 6560–6570 (1998). doi: 10.1128/MCB.18.11.6560; pmid: 9774671
30. E. H. Hettema, R. Erdmann, I. van der Klei, M. Veenhuis, Evolving models for peroxisome biogenesis. *Curr. Opin. Cell Biol.* **29**, 25–30 (2014). doi: 10.1016/j.ceb.2014.02.002; pmid: 24681485
31. D. E. Bustard *et al*., During replication stress, non-SMC element 5 (NSE5) is required for Smc5/6 protein complex functionality at stalled forks. *J. Biol. Chem.* **287**, 11374–11383 (2012). doi: 10.1074/jbc.M111.336263; pmid: 22303010
32. R. J. Chi *et al*., Fission of SNX-BAR-coated endosomal retrograde transport carriers is promoted by the dynamin-related protein Vps1. *J. Cell Biol.* **204**, 793–806 (2014). doi: 10.1083/jcb.201309084; pmid: 24567361
33. J. R. Whyte, S. Munro, A yeast homolog of the mammalian mannose 6-phosphate receptors contributes to the sorting of vacuolar hydrolases. *Curr. Biol.* **11**, 1074–1078 (2001). doi: 10.1016/S0960-9822(01)00273-1; pmid: 11470415
34. A. Eugster, G. Frigerio, M. Dale, R. Duden, The α- and β′-COP WD40 domains mediate cargo-selective interactions with distinct di-lysine motifs. *Mol. Biol. Cell* **15**, 1011–1023 (2004). doi: 10.1091/mbc.E03-10-0724; pmid: 14699056
35. C. A. Argmann, S. M. Houten, J. Zhu, E. E. Schadt, A next generation multiscale view of inborn errors of metabolism. *Cell Metab.* **23**, 13–26 (2016). doi: 10.1016/j.cmet.2015.11.012; pmid: 26712461
36. S. Gavrilets, in *Fitness Landscapes and the Origin of Species*, S. A. Levin, H. S. Horn, Eds. (Monographs in Population Biology, Princeton Univ. Press, 2004), pp. 149–194.
37. B. R. Foley, C. G. Rose, D. E. Rundle, W. Leong, S. Edmands, Postzygotic isolation involves strong mitochondrial and sex-specific effects in *Tigriopus californicus*, a species lacking heteromorphic sex chromosomes. *Heredity* **111**, 391–401 (2013). doi: 10.1038/hdy.2013.61; pmid: 23860232
38. H. A. Orr, The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics* **139**, 1805–1813 (1995). pmid: 7789779
39. J. J. Welch, Accumulating Dobzhansky-Muller incompatibilities: Reconciling theory and data. *Evolution* **58**, 1145–1156 (2004). doi: 10.1111/j.0014-3820.2004.tb01695.x; pmid: 15266965
40. L. C. Moyle, T. Nakazato, Hybrid incompatibility "snowballs" between *Solanum* species. *Science* **329**, 1521–1523 (2010). doi: 10.1126/science.1193063; pmid: 20847271
41. S. Tang, D. C. Presgraves, Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* **323**, 779–782 (2009). doi: 10.1126/science.1169123; pmid: 19197064
42. S. M. Richardson *et al*., Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017). doi: 10.1126/science.aaf4557; pmid: 28280199
43. D. G. Gibson *et al*., One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20404–20409 (2008). doi: 10.1073/pnas.0811011106; pmid: 19073939
44. C. A. Hutchison 3rd *et al*., Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016). doi: 10.1126/science.aad6253; pmid: 27013737
45. D. G. Gibson *et al*., Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010). doi: 10.1126/science.1190719; pmid: 20488990
46. T. Wang *et al*., Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015). doi: 10.1126/science.aac7041; pmid: 26472758
47. T. Hart *et al*., High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015). doi: 10.1016/j.cell.2015.11.015; pmid: 26627737
48. V. A. Blomen *et al*., Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015). doi: 10.1126/science.aac7557; pmid: 26472760
49. R. D. Dowell *et al*., Genotype to phenotype: A complex problem. *Science* **328**, 469 (2010). doi: 10.1126/science.1189015; pmid: 20413493
50. S. K. Forsberg, J. S. Bloom, M. J. Sadhu, L. Kruglyak, Ö. Carlborg, Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat. Genet.* **49**, 497–503 (2017). doi: 10.1038/ng.3800; pmid: 28250458
51. M. B. Taylor, I. M. Ehrenreich, Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLOS Genet.* **10**, e1004324 (2014). doi: 10.1371/journal.pgen.1004324; pmid: 24784154
52. M. B. Taylor, I. M. Ehrenreich, Higher-order genetic interactions and their contribution to complex traits. *Trends Genet.* **31**, 34–40 (2015). doi: 10.1016/j.tig.2014.09.001; pmid: 25284288
53. P. C. Phillips, Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008). doi: 10.1038/nrg2452; pmid: 18852697
54. T. B. Sackton, D. L. Hartl, Genotypic context and epistasis in individuals and populations. *Cell* **166**, 279–287 (2016). doi: 10.1016/j.cell.2016.06.047; pmid: 27419868
55. A. Baryshnikova, Systematic functional annotation and visualization of biological networks. *Cell Syst.* **2**, 412–421 (2016). doi: 10.1016/j.cels.2016.04.014; pmid: 27237738
56. A. C. Gavin *et al*., Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006). doi: 10.1038/nature04532; pmid: 16429126
57. N. J. Krogan *et al*., Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006). doi: 10.1038/nature04670; pmid: 16554755
58. K. Tarassov *et al*., An in vivo map of the yeast protein interactome. *Science* **320**, 1465–1470 (2008). doi: 10.1126/science.1153878; pmid: 18467557
59. H. Yu *et al*., High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008). doi: 10.1126/science.1158684; pmid: 18719252
60. M. Babu *et al*., Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. *Nature* **489**, 585–589 (2012). doi: 10.1038/nature11354; pmid: 22940862
61. C. Huttenhower, M. Hibbs, C. Myers, O. G. Troyanskaya, A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics* **22**, 2890–2897 (2006). doi: 10.1093/bioinformatics/btl492; pmid: 17005538
62. Y. T. Chong *et al*., Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413–1424 (2015). doi: 10.1016/j.cell.2015.04.051; pmid: 26046442
63. P. Shannon *et al*., Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). doi: 10.1101/gr.1239303; pmid: 14597658

ACKNOWLEDGMENTS

We thank H. Friesen, J. Hou, and J. Moffat for discussions and M. P. Masinas and J. Nelson for logistical help. **Funding:** This work was primarily supported by the NIH (grant R01HG005853 to C.B., B.J.A., and C.L.M. and grants R01HG005084 and R01GM104975 to C.L.M.), the Canadian Institutes of Health Research (CIHR) (grants FDN-143264 and FDN-143265 to C.B. and B.J.A.), and the NSF (grant DBI\0953881 to C.L.M.). Computing resources and data storage services were partially provided by the Minnesota Supercomputing Institute and the University of Minnesota Office of Information Technology, respectively. Additional support was provided by the CIHR (grant MOP-79368 to G.W.B.), the NSF (DEB-1456462 to D.B.), the Swiss National Science Foundation (R.L.), the Canton of Geneva and the European Research Council Consolidator Grant program (R.L.), Natural Science and Engineering Research Council of Canada Postgraduate Scholarship-Doctoral PGS D2 (E.K.), a University of Toronto Open Fellowship (E.K.), and a University of Minnesota Doctoral Dissertation Fellowship (B.V). C.L.M, B.J.A., and C.B. are fellows of the Canadian Institute for Advanced Research. **Author contributions:** Conceptualization: E.K., B.V., B.J.A., C.B., and C.L.M.; Methodology and investigation: E.K., B.V., W.W., R.D., Y.C., A.B., M.M.U., J.v.L., E.N.K., C.P., A.J.D., M.P., J.Z.Y.W., J.H., M.R., K.X., H.H., B.-J.S.L., E.S., and H.Z.; Formal analysis: E.K., B.V., W.W., M.M.U., E.N.K., C.P., A.J.D., J.H., K.X., H.H., M.C., R.L., A.C., D.B., and G.W.B.; Resources: G.T.; Data curation: M.U.; Writing – original draft: E.K., B.V., B.J.A., C.B., and C.L.M.; Writing – review and editing: E.K., B.V., W.W., R.D., A.B., M.M.U., J.v.L., E.N.K., C.P., M.C., D.B., G.W.B., B.J.A., C.B., and C.L.M.; Supervision: B.J.A., C.B., and C.L.M.; Project administration: N.V.D. and S.S.; Funding acquisition: B.J.A., C.B., and C.L.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data files (data S1 to S7) associated with this study are described in detail and available in the supplementary materials and can be downloaded from http://boonelab.ccbr.utoronto.ca/supplement/kuzmin2018/supplement.html. Data files S1 to S7 have also been deposited in the DRYAD Digital Repository (doi: 10.5061/dryad.tt367)

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/eaao1729/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S5
References (*64*–*82*)
Data S1 to S7

25 June 2017; accepted 23 February 2018
10.1126/science.aao1729

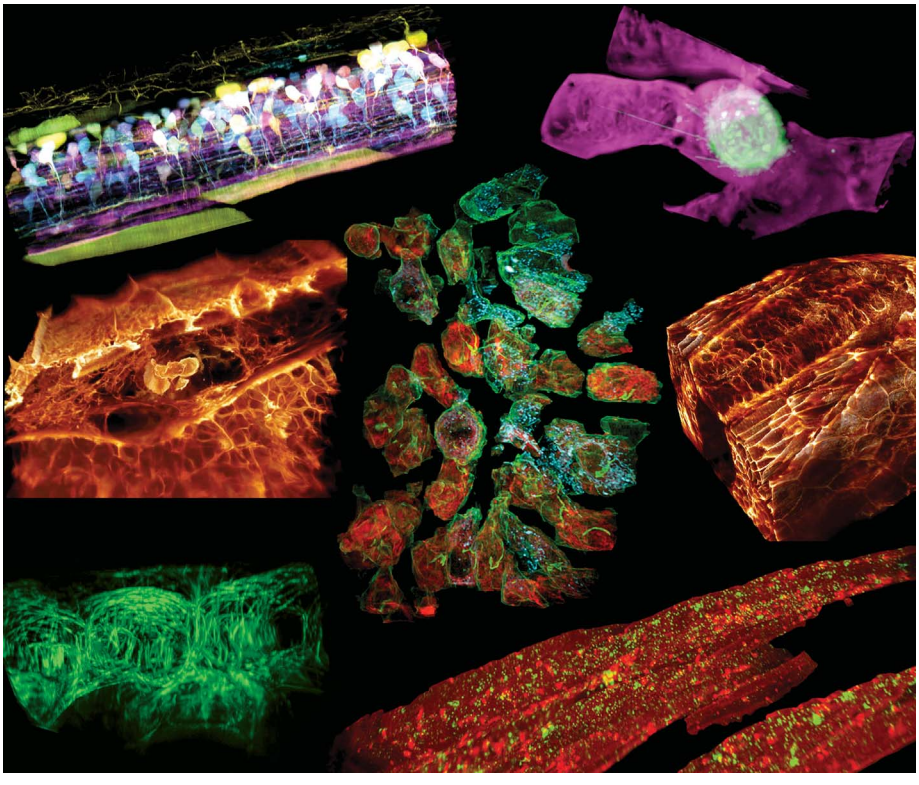
RESEARCH ARTICLE SUMMARY

ADVANCED IMAGING

Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms

Tsung-Li Liu,* Srigokul Upadhyayula,* Daniel E. Milkie, Ved Singh, Kai Wang, Ian A. Swinburne, Kishore R. Mosaliganti, Zach M. Collins, Tom W. Hiscock, Jamien Shea, Abraham Q. Kohrman, Taylor N. Medwig, Daphne Dambournet, Ryan Forster, Brian Cunniff, Yuan Ruan, Hanako Yashiro, Steffen Scholpp, Elliot M. Meyerowitz, Dirk Hockemeyer, David G. Drubin, Benjamin L. Martin, David Q. Matus, Minoru Koyama, Sean G. Megason, Tom Kirchhausen, Eric Betzig†

INTRODUCTION: Organisms live by means of the complex, dynamic, three-dimensional (3D) interplay between millions of components, from the molecular to the multicellular. Visualizing this complexity in its native form requires imaging at high resolution in space and time anywhere within the organism itself, because only there are all the environmental factors that regulate its physiology present. However, the optical heterogeneity of multicellular systems leads to aberrations that quickly compromise resolution, signal, and contrast with increasing imaging depth. Furthermore, even in the absence of aberrations, high resolution and fast imaging are usually accompanied by intense illumination, which can perturb delicate subcellular processes or even introduce permanent phototoxic effects.



High-resolution in vivo cell biology. AO-LLSM permits the study of 3D subcellular processes in their native multicellular environments at high spatiotemporal resolution, including (clockwise from upper left) growth of spinal cord axons; cancer cell metastasis; collective cellular motion; endocytosis; microtubule displacements; immune cell migration; and (center) organelle dynamics.

RATIONALE: We combined two imaging technologies to address these problems. The first, lattice light-sheet microscopy (LLSM), rapidly and repeatedly sweeps an ultrathin sheet of light through a volume of interest while acquiring a series of images, building a high-resolution 3D movie of the dynamics within. The confinement of the illumination to a thin plane insures that regions outside the volume remain unexposed, while the parallel collection of fluorescence from across the plane permits low, less perturbative intensities to be used. The second technology, adaptive optics (AO), measures sample-induced distortions to the image of a fluorescent "guide star" created within the volume—distortions that also affect the acquired light-sheet images—and compensates for these by changing the shape of a mirror to create an equal but opposite distortion.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaq1392

RESULTS: We applied AO-LLSM to study a variety of 3D subcellular processes in vivo over a broad range of length scales, from the nanoscale diffusion of clathrin-coated pits (CCPs) to axon-guided motility across 200 μm of the developing zebrafish spinal cord. Clear delineation of cell membranes allowed us to computationally isolate and individually study any desired cell within the crowded multicellular environment of the intact organism. By doing so, we could compare specific processes across different cell types, such as rates of CCP internalization in muscle fibers and brain cells, organelle remodeling during cell division in the developing brain and eye, and motility mechanisms used by immune cells and metastatic breast cancer cells. Although most examples were taken from zebrafish embryos, we also demonstrated AO-LLSM in a human stem cell–derived organoid, a *Caenorhabditis elegans* nematode, and *Arabidopsis thaliana* leaves.

CONCLUSION: AO-LLSM takes high-resolution live-cell imaging of subcellular processes from the confines of the coverslip to the more physiologically relevant 3D environment within whole transparent organisms. This creates new opportunities to study the phenotypic diversity of intracellular dynamics, extracellular communication, and collective cell behavior across different cell types, organisms, and developmental stages. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: betzige@janelia.hhmi.org.
Cite this article as T.-L. Liu *et al.*, *Science* **360**, eaaq1392 (2018). DOI: 10.1126/science.aaq1392

RESEARCH ARTICLE

ADVANCED IMAGING

Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms

Tsung-Li Liu,[1]* Srigokul Upadhyayula,[1,2,3,4]* Daniel E. Milkie,[1] Ved Singh,[1] Kai Wang,[1]† Ian A. Swinburne,[5] Kishore R. Mosaliganti,[5] Zach M. Collins,[5] Tom W. Hiscock,[5] Jamien Shea,[1] Abraham Q. Kohrman,[6] Taylor N. Medwig,[6] Daphne Dambournet,[7] Ryan Forster,[7] Brian Cunniff,[2,3]‡ Yuan Ruan,[8] Hanako Yashiro,[8] Steffen Scholpp,[9,10] Elliot M. Meyerowitz,[8] Dirk Hockemeyer,[7] David G. Drubin,[7] Benjamin L. Martin,[6] David Q. Matus,[6] Minoru Koyama,[1] Sean G. Megason,[5] Tom Kirchhausen,[1,2,3,4] Eric Betzig[1]§

True physiological imaging of subcellular dynamics requires studying cells within their parent organisms, where all the environmental cues that drive gene expression, and hence the phenotypes that we actually observe, are present. A complete understanding also requires volumetric imaging of the cell and its surroundings at high spatiotemporal resolution, without inducing undue stress on either. We combined lattice light-sheet microscopy with adaptive optics to achieve, across large multicellular volumes, noninvasive aberration-free imaging of subcellular processes, including endocytosis, organelle remodeling during mitosis, and the migration of axons, immune cells, and metastatic cancer cells in vivo. The technology reveals the phenotypic diversity within cells across different organisms and developmental stages and may offer insights into how cells harness their intrinsic variability to adapt to different physiological environments.

A common tenet, oft repeated in the field of bioimaging, is "seeing is believing." But when can we believe what we see? The question becomes particularly relevant when imaging subcellular dynamics by fluorescence microscopy. Traditional imaging tools such as confocal microscopy are often too slow to study fast three-dimensional (3D) processes across cellular volumes, create out-of-focus photoinduced damage (*1*, *2*) and fluorescence photobleaching, and subject the cell at the point of measurement to peak intensities far beyond those under which life evolved. In addition, much of what fluorescence microscopy has taught us about subcellular processes has come from observing isolated adherent cells on glass. True physiological imaging requires studying cells within the organism in which they evolved, where all the environmental cues that regulate cell physiology are present (*3*). Although intravital imaging achieves this goal (*4*, *5*) and has contributed pivotally to our understanding of cellular and developmental biology, the resolution needed to study minute subcellular processes in 3D detail is compromised by the optically challenging multicellular environment.

Two imaging tools have recently been developed to address these problems: Lattice light-sheet microscopy (LLSM) (*6*) provides a noninvasive alternative for volumetric imaging of whole living cells at high spatiotemporal resolution, often over hundreds of time points, and adaptive optics (AO) (*7*) corrects for sample-induced aberrations caused by the inhomogeneous refractive index of multicellular specimens and recovers resolution and signal-to-background ratios comparable to those attained for isolated cultured cells. The remaining challenge is to combine these technologies in a way that retains their benefits and thereby enables the in vivo study of cell biology at high resolution in conditions as close as possible to the native physiological state. Here we describe a technique based on an adaptive optical lattice light-sheet microscope designed for this purpose (AO-LLSM) and demonstrate its utility through high-speed, high-resolution, 3D in vivo imaging of a variety of dynamic subcellular processes.

Lattice light-sheet microscope with two-channel adaptive optics

Although several AO methods have been demonstrated in biological systems (*7*), including in the excitation (*8*) or detection (*9*) light paths of a light-sheet microscope, we chose an approach where the sample-induced aberrations affecting the image of a localized reference "guide star" created through two-photon excited fluorescence (TPEF) within the specimen are measured and then corrected with a phase modulation element (*10*). By scanning the guide star over the region to be imaged (*11*), an average correction is measured that is often more accurate than single-point correction—which is essential, because a poor AO correction is often worse than none at all. Scanning also greatly reduces the photon load demanded from any single point. Coupled with correction times as short as 70 ms (*11*), this AO method is compatible with the speed and noninvasiveness of LLSM.

In LLSM, light traverses different regions of the specimen for excitation and detection and therefore is subject to different aberrations. Hence, independent AO systems are needed for each. This led us to design a system (Fig. 1A, supplementary note 1, and fig. S1) where light (red) from a Ti:Sapphire ultrafast laser is ported to either the excitation or detection arm of a LLS microscope (left inset, Fig. 1A) by switching galvanometer 1. In the detection case, TPEF (green) generated within a specimen by scanning the guide star across the focal plane of the detection objective (DO) is descanned (*11*) and sent to a Shack-Hartmann wavefront sensor (DSH) via switching galvanometer 2 (SG2). We then apply the inverse of the measured aberration to a deformable mirror (DM) placed conjugate to both the DSH and the rear pupil plane of the DO (supplementary note 2). Because the signal (also green) generated by the LLS when in imaging mode travels the same path through the specimen as the guide star, and reflects from the same DM, the corrective pattern that we apply to the DM produces an AO-corrected image of the current excitation plane within the specimen on the camera when ported there by the SG2.

Similarly, for excitation correction, we send descanned TPEF generated and collected through the excitation objective (EO) to a second Shack-Hartmann sensor. However, because LLS excitation is confined to a thin annulus at the rear pupil of the DO (*6*), a DM placed conjugate to this pupil would be ineffective for AO correction over most of its surface. Instead, we apply wavefront correction at the same sample-conjugate spatial light modulator (SLM) that creates the light sheet itself, thereby enlisting thousands of independently corrective pixels. To do so, we

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. [2]Department of Cell Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA. [3]Program in Cellular and Molecular Medicine, Boston Children's Hospital, 200 Longwood Avenue, Boston, MA 02115, USA. [4]Department of Pediatrics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA. [5]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA. [6]Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY 11794-5215, USA. [7]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [8]Howard Hughes Medical Institute and Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [9]Living Systems Institute, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK. [10]Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, 76021 Karlsruhe, Germany.
*These authors contributed equally to this work. †Present address: Institute of Neuroscience, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai, China. ‡Present address: Department of Pathology, University of Vermont College of Medicine, Burlington, VT 05401, USA.
§Corresponding author. Email: betzige@janelia.hhmi.org

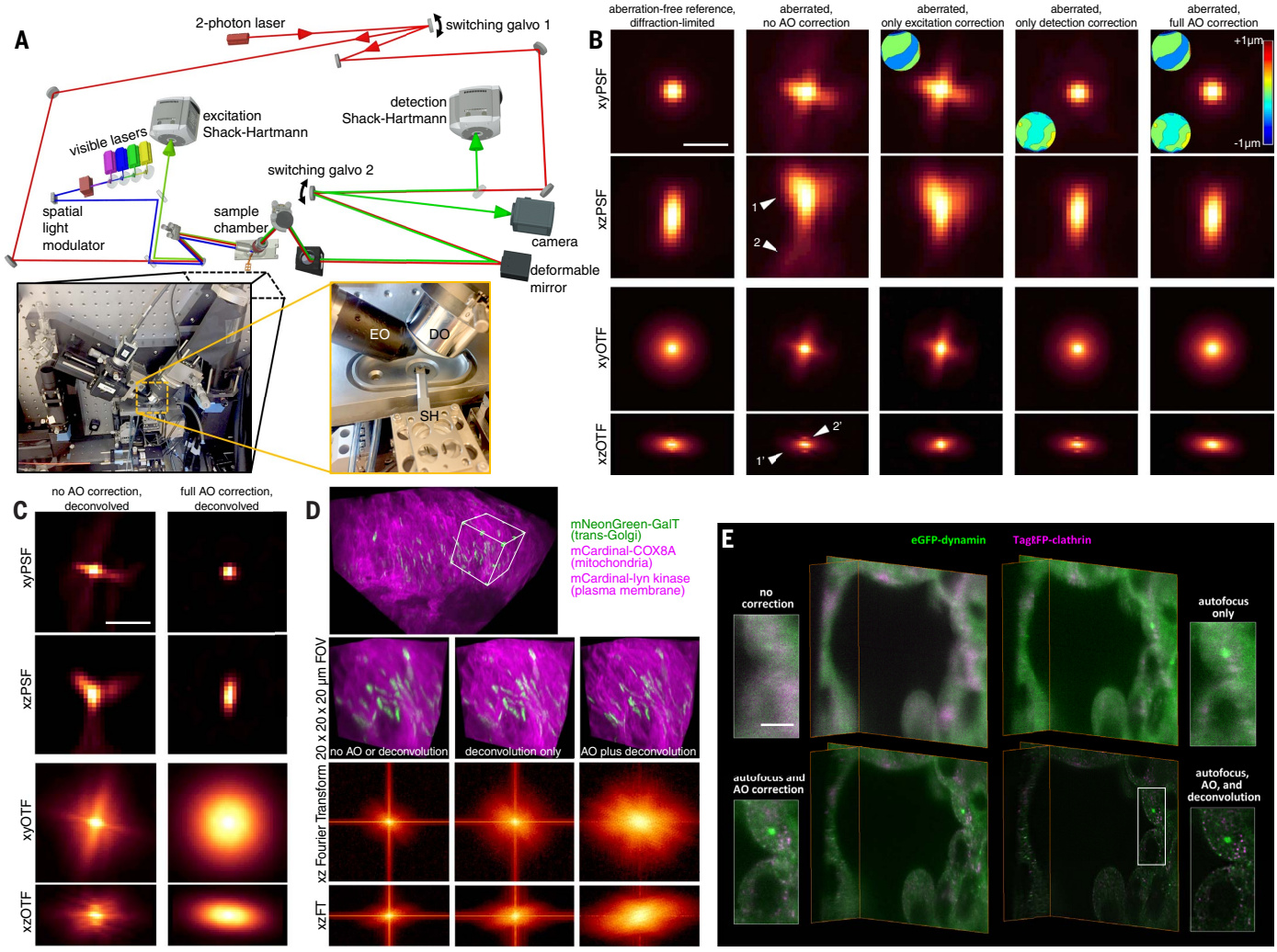
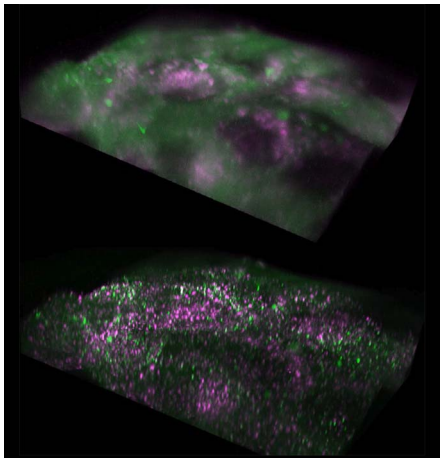


Fig. 1. Adaptive optical lattice light-sheet microscopy (AO-LLSM). (**A**) Simplified microscope schematic (fig. S1 shows a detailed version). EO, excitation objective; DO, detection objective; SH, sample holder. (**B**) *xy* and *xz* maximum intensity projections (MIPs) of the point spread function (PSF; top two rows) and corresponding optical transfer function (OTF; bottom two rows) of the microscope under five different degrees of AO correction (columns), as measured from a 200-nm fluorescent bead in an aberrating agarose gel. Insets show the corrective wavefronts applied. Arrowheads indicate lateral and axial aberrations. Scale bar, 1 μm. (**C**) MIPs and corresponding OTFs of the uncorrected (left column) and fully corrected (right column) bead images from (B), after deconvolution using the aberration-free reference PSF. Scale bar, 1 μm. (**D**) Cellular trans-Golgi, mitochondria, and plasma membranes in the spine of a live zebrafish embryo 24 hpf. Shown are unprocessed data without AO correction (left column), deconvolved data without AO correction (center), and deconvolved data after AO correction (top and right) (movie S2). MIP views (bottom two rows) of the Fourier transform (FT) of the data in all three cases indicate their respective degrees of information recovery. (**E**) Four different levels of correction, shown for orthoslices of a live human stem cell–derived organoid grown in Matrigel and gene-edited to express endogenous levels of clathrin and dynamin in coated endocytic pits (Movie 1). Scale bar, 5 μm.

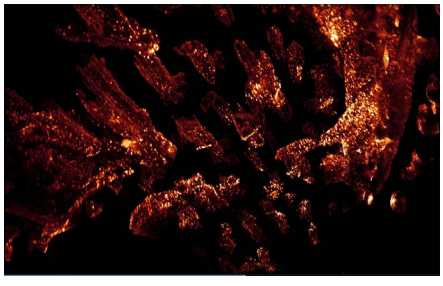
subtract the measured phase aberration from the phase of the Fourier transform (FT) of the ideal, aberration-free SLM pattern, then inverse-transform the result back to the sample-conjugate SLM plane (supplementary note 3 and fig. S2).

Lastly, optimal resolution requires the LLS to be coincident with the focal plane of the DO to less than ~0.46 μm over the entire field of view (FOV), whereas refractive index differences between the specimen and the surrounding media lead to tip-tilt alignment or axial displacement errors for the light sheet that might exceed this (*12*). Fortunately, we find that only displacement is a concern over FOVs typical of LLSM (supplementary note 4 and fig. S3) and that it can be robustly corrected by imaging, edge-on through the EO, the offset of the plane of TPEF that we generate when measuring the detection aberration relative to the plane of fluorescence generated by the LLS (supplementary note 5 and fig. S4).

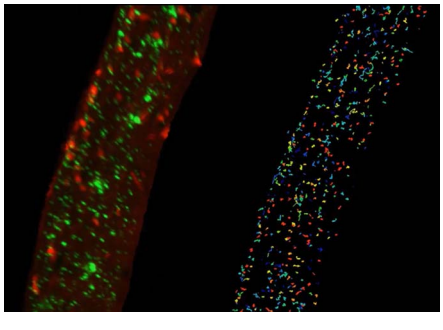
As an example (Fig. 1B), after correction for aberrations in the microscope itself, the point spread function (PSF) and corresponding optical transfer function (OTF), measured from a 200-nm bead, indicate nearly diffraction-limited performance (first column). However, when a similar bead is placed in 2% low-melt agarose, we observe substantial aberration (second column), both laterally (arrowheads 1 and 1′) and axially (arrowheads 2 and 2′). Correcting only the excitation aberration improves the axial resolution (third column) by returning the light sheet to its original width (fig. S2). Conversely, correcting only the detection aberration improves primarily the lateral performance (fourth column). However, when we combine the excitation and detection corrections, the image of the bead is returned to its diffraction-limited form (fifth column), with an eightfold recovery to its original peak intensity. Furthermore, the same



Movie 1. Endocytosis in a human stem cell–derived organoid. Gene-edited clathrin (magenta) and dynamin (green) before and after adaptive optical correction and deconvolution, showing comparative *xy* and *xz* orthoslices, volume renderings, and postcorrection tracking of the motion and lifetimes of individual CCPs and CCVs over 120 time points at 1.86-s intervals (Fig. 1E and fig. S5).



Movie 2. Clathrin-mediated endocytosis in vivo. Dynamics of CCPs and CCVs over 15 min at 10-s intervals in the dorsal tail region of a zebrafish embryo 80 hpf. Segmented cells reveal brighter clathrin puncta at the vascular endothelium than at muscle fibers (Fig. 2A).



Movie 3. Clathrin localization in muscle fibers. PM (red) and clathrin (green) in the tail of a zebrafish embryo 50 to 55 hpf, showing *xy* and *xz* orthoslices before and after AO correction and deconvolution, dynamics of individual CCPs and CCVs at and between t-tubules, large clathrin clusters and small clathrin puncta in volume-rendered and segmented cells, and tracked CCPs and CCVs in a segmented cell (Fig. 2, B to E; fig. S8; and movie S3).

correction proves valid over a 30-μm field of beads in agarose (movie S1).

One of the key advantages of complete AO correction is that it enables accurate deconvolution (Fig. 1C), giving the most truthful representation of the specimen possible within the limits of diffraction (second column). In contrast, applying deconvolution to an aberrated image gives a distorted result, because the OTF can then fall below the noise floor asymmetrically and at spatial frequencies well below the diffraction limit (first column), after which higher spatial frequencies cannot be recovered by deconvolution. These same trends can be seen in densely labeled specimens as well, such as across mitochondria, Golgi, and plasma membranes (PMs) of cells near the spinal midline in a living zebrafish embryo 24 hours postfertilization (hpf) (Fig. 1D, bottom three rows, and movie S2), where the greatest information content as seen in the FT of image volume is obtained by combined AO and deconvolution.

Next, we imaged organoids differentiated from human stem cells, gene-edited to express endogenous levels of red fluorescent protein (tagRFP)–clathrin and enhanced green fluorescent protein (EGFP)–dynamin in endocytic pits (clathrin-coated pits; CCPs). Such organoids permit the study of human cellular differentiation and tissue morphogenesis in vitro at subcellular resolution, with an experimental accessibility that is difficult to achieve in vivo. However, the extracellular matrix in which they are grown introduces considerable aberrations, and the fast dynamics and limited number of fluorophores in each CCP demand high-speed imaging with low photobleaching. The system is therefore well suited to the capabilities of AO-LLSM, with the CCPs doubling as distributed puncta of subdiffractive size to evaluate imaging performance.

Without AO or focus correction (upper left of Fig. 1E and Movie 1), no CCPs are visible, and the cell boundaries are poorly defined. Autofocus alone (upper right) reveals larger patches of clathrin and dynamin but must be combined with complete AO correction (excitation and detection, lower left) to identify individual CCPs. At that point, the imaging is diffraction-limited, so the data can be deconvolved using the system-corrected PSF to compensate for the spatial filtering properties of the microscope. Both dynamin and clathrin puncta then stand out clearly above the cytosolic background (lower right), allowing us to quantitatively measure their lifetimes and diffusion tracks (Movie 1). The increasing recovery of information as we progress through these four cases can also be seen quantitatively in their corresponding OTFs (fig. S5).

Clathrin dynamics in zebrafish

For transparent model organisms, we can apply AO-LLSM in vivo, where the complete physiological environment is preserved. The zebrafish has become the most widely used nonmammalian vertebrate model organism. We took advantage of its small size and transparency to visualize the formation of endocytic CCPs and clathrin-coated vesicles (CCVs) in the context of the developing organism. We first imaged a volume in the dorsal tail region of a fish larva 80 hpf stably expressing dsRed–clathrin light chain A (CLTA) (*13*) (Fig. 2A and Movie 2). We observed a high density of diffraction-limited clathrin spots that, after computational separation of all cells, were found to be mostly colocalized with the PM. These spots appeared and disappeared with the formation of new CCPs and their eventual uncoating. We determined that the areal density of CCPs was similar in muscle fibers (e.g., green arrowheads, Fig. 2A) and endothelial cells (magenta arrowheads) lining blood vessels, but the distribution of their intensities was broader in the latter (lower inset, Fig. 2A), which had a subpopulation of pits that were up to sixfold as bright as the median CCP intensity in the former. Because CCP size is proportional to intensity (*14*), these results indicate the presence of larger structures, possibly clusters of CCPs (*15*), in the vasculature endothelium.

To track CCPs for longer times, we turned to zebrafish embryos mRNA-injected to express the brighter and more photostable fluorescent fusion protein mNeonGreen-CLTA. These embryos also expressed mCardinal targeted to the PM, facilitating the computational separation of cells. Embryos imaged in the tail (Fig. 2B, fig. S6A, and Movie 3) and in the hindbrain (fig. S6B) displayed a variety of morphologies, trafficking behaviors (Fig. 2C), and lifetime distributions (Fig. 2D). Large, micron-scale intracellular spots of limited mobility (arrowheads, Fig. 2B) probably represent clathrin-rich vesicles clustered at the trans-Golgi network, whereas mobile diffraction-limited spots (Fig. 2C, arrowhead group 1) likely correspond to endosomal carriers similar to those seen in cultured mammalian cells (*16*). Diffraction-limited spots at the PM (Fig. 2C, arrowhead group 2) likely represent individual CCPs and CCVs. We also found CCPs at the t-tubules spanning muscle fibers (Fig. 2C, top left), in contrast to the diffuse clathrin signal observed using immunofluorescence in fixed rat muscle fibers (*17*). Most of these were pinned, but occasionally they would break free from a t-tubule and move rapidly along the fiber axis (Fig. 2C, arrowhead group 3, and movie S3), possibly by active transport along myofibrils or within the sarcoplasmic reticulum.

AO allowed us to detect more CCPs (fig. S7) and track all CCPs with higher precision (Fig. 2C, top right; fig. S8, A to C; and Movie 3). Comparing CCPs in muscle fibers and the brain, we found that, although their initiation frequencies and fluorescence intensities were similar, brain CCPs tended to internalize faster (Fig. 2E). Assuming that clathrin puncta in muscle and brain cells lasting >21 s corresponded to successful coated vesicles, each with an assumed membrane diameter of 60 nm, ~0.1% of the PM was internalized through clathrin-mediated endocytosis every minute (fig. S8D). This is similar to the values derived from measurements in cultured SUM-159 (*18*) or htertRPE-1 (*19*) mammalian cells at 37°C.

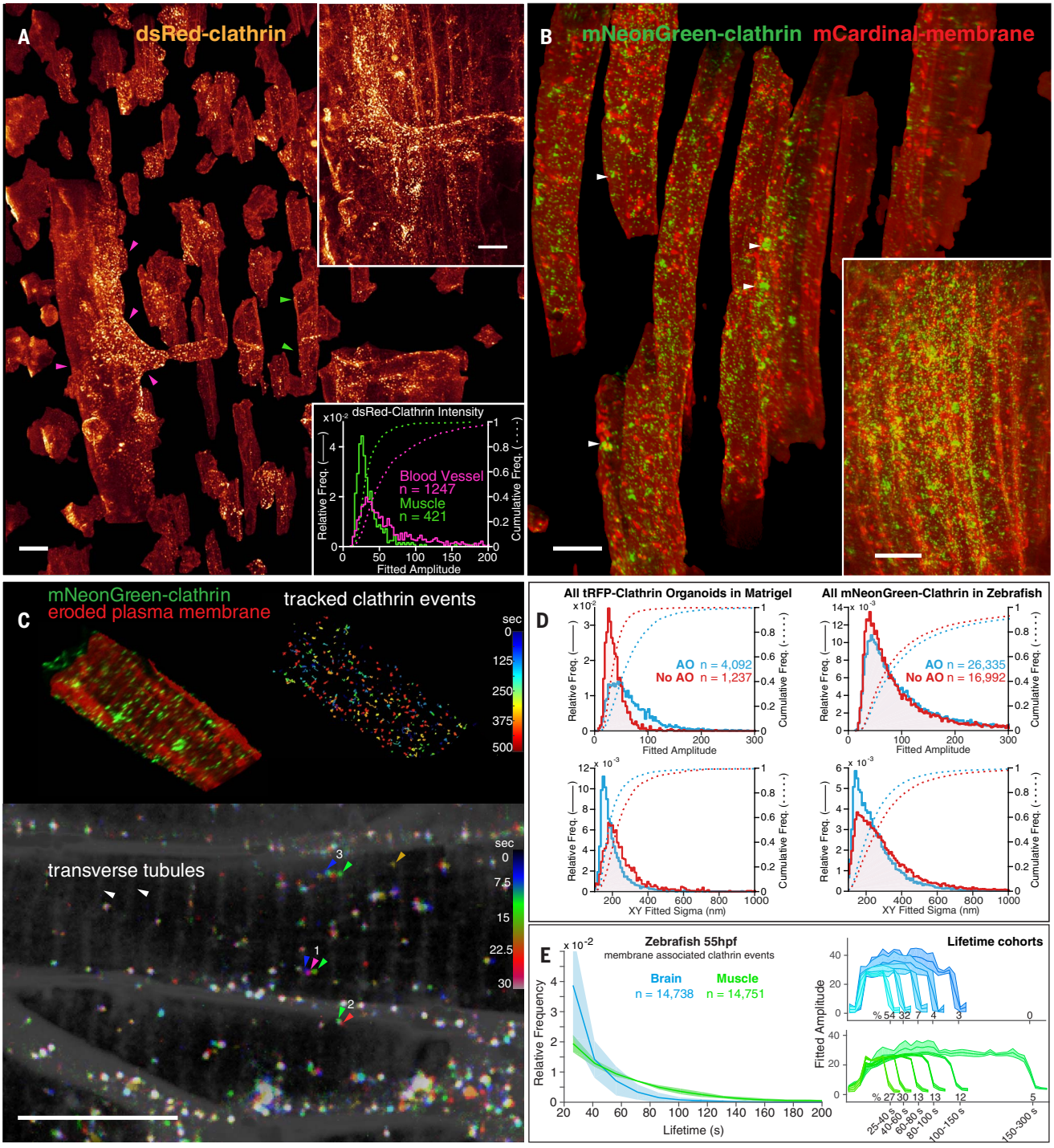


Fig. 2. Clathrin dynamics in zebrafish. (**A**) Computationally separated muscle fibers (e.g., green arrowheads) and vascular endothelial cells (e.g., magenta arrowheads), both expressing DsRed-CLTA to highlight CCPs and CCVs, from muscle tissue in a 75-μm by 99-μm by 41-μm region (upper inset) of the tail of a developing zebrafish larva 80 hpf (Movie 2). Brighter clathrin puncta were observed in the endothelial cells (lower inset). Scale bars, 10 μm. (**B**) Computationally separated muscle fibers from a region (lower inset) in the tail of a zebrafish embryo 50 hpf coexpressing an mCardinal-PM marker (red) and mNeonGreen-CLTA (green). Individual CCPs and CCVs and larger clathrin-rich vesicles (arrowheads) are visible (Movie 3). Scale bars, 10 μm. (**C**) Spatial distribution and dynamics of CCPs and CCVs tracked for 12 min at 7.5-s intervals in one muscle fiber from (B), showing CCPs localized at t-tubules (top left) and diffusion and lifetime characteristics for CCPs and CCVs across the cell (top right). A MIP through a 2-μm-thick slab at three consecutive time points (bottom) shows examples of a pinned CCV (arrowheads 2) and slowly diffusing (arrowheads 1) or rapidly shuttling CCVs (arrowheads 3) (movie S3). Scale bar, 10 μm. (**D**) Effect of AO on the measured quantity, intensity, and localization precision of CCPs and CCVs in the organoid in Fig. 1E and the zebrafish in (B). (**E**) Comparative distribution of CCP and CCV lifetimes (left) and intensity cohorts grouped by their lifetimes (right) in the brain and muscle of a developing zebrafish embryo 55 hpf.

In vivo imaging of organelle morphology and dynamics during embryogenesis

A major focus in cell biology is the study of the structure and function of organelles within the living cell. To study the dynamics of multiple organelles simultaneously throughout the cell cycle across a population of cells in vivo, we imaged brain progenitor cells with markers for the trans-Golgi, endoplasmic reticulum (ER), mitochondria, and PM for 200 image volumes at 44-s intervals (Fig. 3, A and B, and Movie 4). In interphase, we observed multiple trans-Golgi segments in most cells, often appearing as long filaments preferentially aligned along the axis of cell polarization (Fig. 3A) that fragmented during mitosis (Fig. 3B). The ER and mitochondria largely recapitulated the forms that they commonly take in cultured cells: The ER established a reticular network in interphase and sheetlike cisternae during mitosis (*20*), whereas mitochondria formed punctate structures near the surface and longer tubules in the subset of more deeply buried interphase cells that were well labeled. Analyzing one such cell, we found that all three organelles were distributed uniformly between the PM and the nucleus in interphase (fig. S9A), but mitochondria were preferentially located nearer the PM during mitosis (fig. S9A, 109 min).

The early synchrony of cell division is lost in zebrafish at the midblastula transition (3 hpf). Nevertheless, at 14 hpf, we observed instances of cascading cell division, where adjacent cells

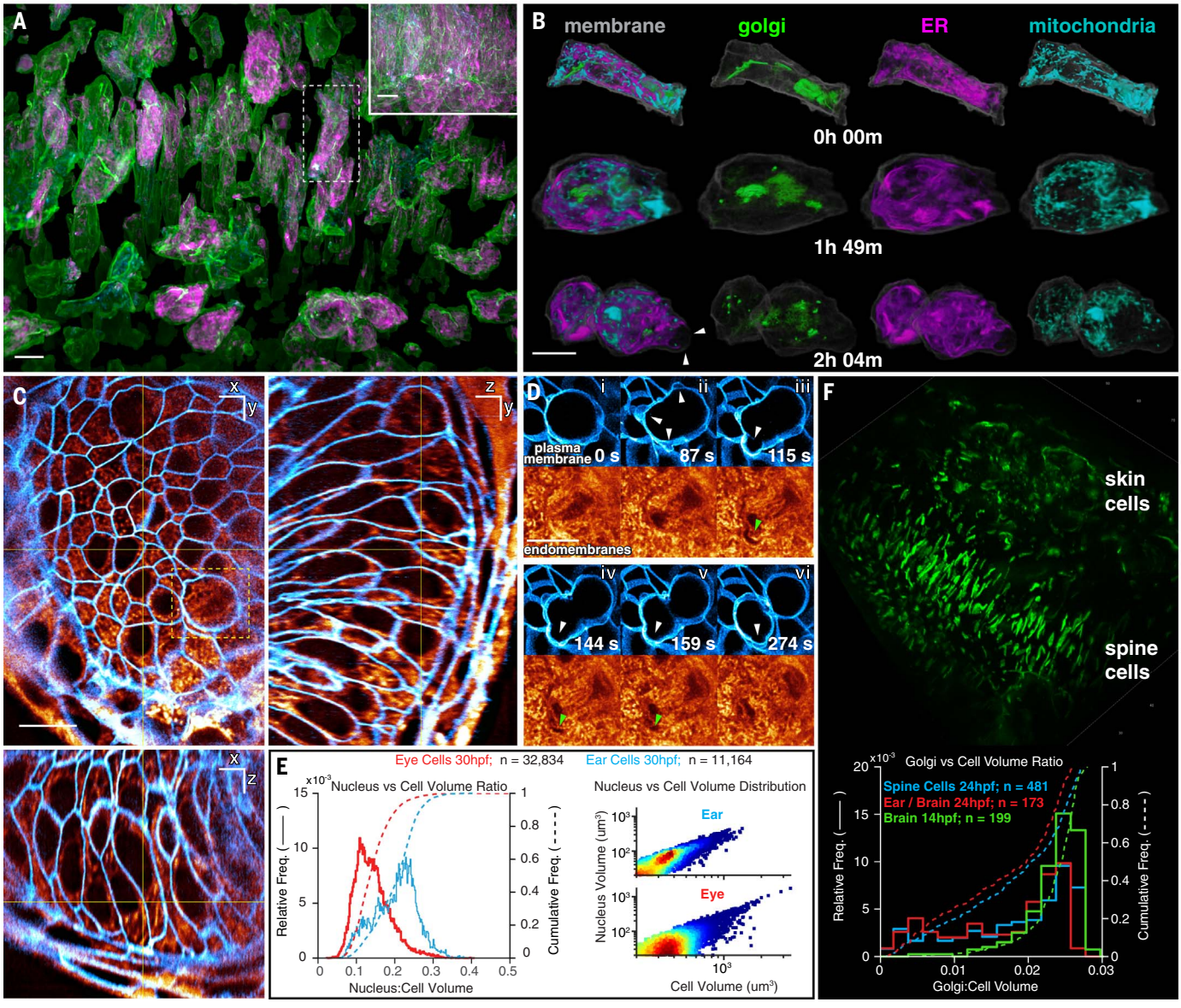


Fig. 3. Organelle morphologies and dynamics in zebrafish. (**A**) Computationally separated neural progenitor cells from a 70-μm by 35-μm by 35-μm region (inset) in the brain of a developing zebrafish embryo expressing GalT-mNeonGreen, tagRFP-Sec61β, and Citrine as markers of the trans-Golgi, ER, and PM, respectively, with additional labeling by MitoTracker Deep Red dye (Movie 4). Scale bars, 10 μm. (**B**) Changing morphologies of the organelles in the specific cell outlined in (A) at three time points through mitosis. Arrowheads indicate mitotic blebs. Scale bar, 10 μm. (**C**) MIP views from 1-μm-thick orthogonal slabs within the eye of a zebrafish embryo 30 hpf, showing PM (blue) and endomembranes (orange) (Movie 5). Scale bars, 10 μm. (**D**) Six time points from Movie 5, showing PM blebs (white arrowheads) during mitosis and the exclusion of endomembranes in early blebs (green arrowheads). (**E**) Correlation between nuclear volume and total cell volume in the eye and ear (Pearson's coefficient, 0.9 and 0.8, respectively). (**F**) Different morphologies of trans-Golgi (top) near the spine of a zebrafish embryo 24 hpf and distribution of trans-Golgi volume in different cell types and at different developmental stages (bottom).

underwent mitosis one after another (fig. S10 and Movie 4). Mitotic cells, as seen previously in cell cultures (*18*, *21*), decreased their surface area (fig. S9E) as they assumed a spheroidal shape and produced transient blebs before cytokinesis, but then recovered their initial area after division. Total cellular volume remained constant throughout mitosis (fig. S9E). We also identified instances of asymmetric cytokinesis (Fig. 3B and Movie 4), where the two daughter cells differed in surface area by ~50% (fig. S9F). When we quantified organelle intensity in one such instance, we also observed asymmetric fractional partitioning of the Golgi and mitochondria between the daughter cells during cytokinesis (fig. S9G).

Mitotic cells in the developing eye 30 hpf (white arrowheads in Fig. 3D and Movie 5) also produced transient blebs before cytokinesis. We discovered upon labeling with Bodipy–tetramethylrhodamine (TMR) that these blebs created voids (green arrowheads in Fig. 3D) that only slowly filled with endomembranes.

Lastly, we observed considerable variability in the size and morphology of specific subcellular features across different organs and developmental stages. Bodipy-TMR negative staining (e.g., Fig. 3, C and D) revealed that the nuclei of ear cells 30 hpf were nearly twice as large as those of eye cells when normalized by the total cellular volume (Fig. 3E, left) and that nuclear volume and cellular volume were positively correlated (Fig. 3E, right). Likewise, we found that the median trans-Golgi volume as a percentage of total cellular volume in brain progenitor cells 14 hpf was larger [2.46%; median absolute deviation (MAD) = 0.26%] than in ear, brain, or spine cells 24 hpf (2.0%; MAD = 0.66%) (Fig. 3F). Golgi took many forms, from the aforementioned narrow polarized filaments in the early brain (Fig. 3A) to shorter segments clustered near the midline in the spine and nuclear-wrapping filaments in skin cells (Fig. 3F).

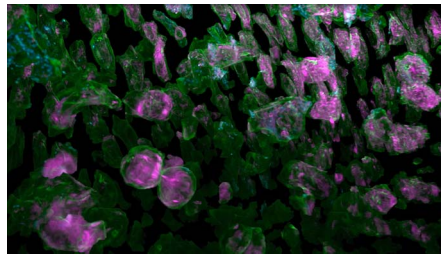
Tiled acquisition for aberrations varying in space and time

Because the refractive index profile can vary across a specimen and can also vary as the specimen develops, the AO corrections required can vary in both space and time. Unfortunately, it is difficult to estimate a priori the size or temporal stability of the isoplanatic patch (the FOV over which a single AO correction is valid). Empirically, we found in zebrafish embryos less than 72 hpf that a single excitation-detection correction pair obtained by scanning and descanning over the FOV is usually valid across 30 to 60 μm in each direction for at least 1 hour, provided that the light does not intersect the yolk. Fortuitously, these dimensions are comparable to those over which a LLS of submicron thickness does not deviate substantially in width. The examples shown above largely fall within these limits and hence, for them, a single AO correction pair at a single time point sufficed.

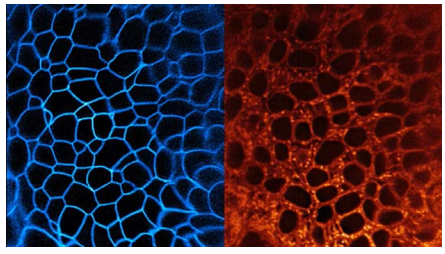
In other cases, however, we may wish to cover much larger FOVs. To do so, we must stitch together data from multiple image subvolume tiles, each with its own independent AO correction. To demonstrate the necessity of this, we imaged a 213-μm by 213-μm by 113-μm volume (Fig. 4A and movie S4) comprising 7 by 7 by 3 tiles in the tail region of a zebrafish embryo 96 hpf expressing membrane-EGFP, using three different protocols: no AO correction (Fig. 4B, left column), AO correction from the center tile applied to all tiles (middle column), and independent AO correction in each tile (right column). When viewed across 3-μm-thick slabs perpendicular (*xy*; Fig. 4B, top row) or parallel (*xz*; bottom row) to the detection axis, a small volume within the center tile (small orange boxes) showed substantial improvement by either center-tile or all-tiles AO correction, in both the *xy* lateral (upper-left orange boxes, top row) and *xz* axial (upper-left orange boxes, bottom row) planes. This is to be expected, because the site of AO correction coincided with the viewing area in these two cases. However, in a small volume at the edge of the fish (small blue boxes), only the data taken using individual AO corrections in each tile recovered optimal resolution in all directions (lower-right blue boxes, right column), because this volume was outside the isoplanatic patch over which the center-tile AO correction is valid. Applying the center correction across the larger stitched volume often results in greater wavefront errors (fig. S11) and poorer resolution (lower-right blue boxes, middle column) than applying no correction at all (lower-right blue boxes, left column), highlighting the importance of accurate and robust correction, if AO is to be applied.

Empirically, the largest aberrations that we observed in zebrafish embryos occur at regions of high curvature between the embryo and the surrounding media or at regions of rapid refractive index change, such as near the notochord (Fig. 4C and Movie 6). For example, when the LLS penetrates the embryo near-perpendicularly to its surface, the excitation aberration is initially small (Fig. 4C, red arrowhead, left column, top). However, after the light sheet passes through the notochord, it encounters substantial aberration, as seen in both the measured wavefront (green arrowhead, left column, top) and the uncorrected image (green arrowhead, middle column, bottom). On the detection side, aberrations increase with increasing depth in the embryo (left column, bottom, and middle column, top to bottom). In addition, substantial aberrations occur when the edge of the embryo is imaged tangentially (yellow arrowhead, left column, bottom), so that part of the detection light cone intersects the embryo and part does not.

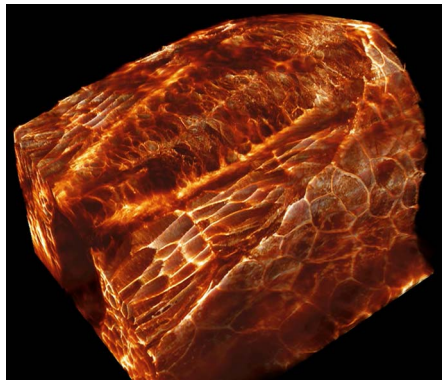
Provided that the specimen does not shift by more than a fraction of the isoplanatic patch size during imaging, a given set of tiled AO corrections can remain valid for hours (Movie 6). However, growth during development can cause an embryo to change its shape, position, or refractive index profile so that new corrections are occasionally needed. Fortunately, these changes often occur on a time scale that is slow compared with that needed to image even a large FOV by LLSM. In such cases, it is sufficient to update the



Movie 4. Subcellular imaging of organelle dynamics in the early zebrafish brain. Dynamics of PM (green or gray) and trans-Golgi (green), ER (magenta or red), and mitochondria (cyan) within neural progenitor cells over 200 time points at 44-s intervals from 14.0 hpf, showing complexity within the tissue, cross-sectional slab views through cells, sequential division of adjacent cells, segmentation and separation of all cells, and morphological changes to organelles during mitosis in one such cell (Fig. 3, A and B, and figs. S9 and S10).



Movie 5. Membrane dynamics in the zebrafish eye. PM (blue) and the endomembrane system (orange) 30 hpf viewed as *xy* orthoslices, cell divisions in a 1-μm-thick slab, and volume-rendered PM dynamics across the eye at 43.8-s intervals for 200 time points (Fig. 3, C to E).



Movie 6. Tiled AO correction for imaging large volumes. A 170-μm by 185-μm by 135-μm volume from the dorsal surface to the notochord in a PM-labeled zebrafish embryo, showing increasing aberration but continued full correction at increasing depth; corrective excitation and detection wavefronts in each of the tiled isoplanatic subvolumes of 5 by 4 by 7 tiles; and four views of PM dynamics within the complete volume from 30 to 39.5 hpf, imaged at 7.5 min intervals (Fig. 4C).

correction at only a subset of different tiles at each time point, as long as all subsets together encompass all tiles in the FOV before the previous round of corrections becomes inaccurate (Fig. 4D and movie S5). Usually, we chose subsets that broadly covered the FOV to monitor where the aberrations change the fastest.

One region that involves substantial specimen curvature, large spatial variation of the refractive index, and gradual aberration change is the eye of the developing zebrafish. Although in vivo optic cup development has been studied at subcellular 3D resolution by conventional (*22*, *23*) and multiview (*24*) light-sheet microscopy, tiled AO-LLSM permits a more detailed look at cellular morphology and organelle distributions during this process. We imaged across 4 by 4 by 3 tiles (Fig. 5A and Movie 7) spanning most of the eye of an embryo 24 to 27 hpf at 6-min intervals to study differences in the intracellular organization of various organelles (Fig. 5, B and C). Transgenic labeling of the PM allowed us to segment, isolate (Fig. 5D), and characterize each cell by type (Fig. 5E). Skin cells exhibited mitochondria clustered in the perinuclear region, like mitochondria seen in flat and thin cultured cells, to which these skin cells are morphologically similar. In contrast, mitochondria in retinal neuroepithelial (RNE) cells were generally longer, distributed across the length of the cell, and polarized along the same axis as the cell itself. The ER in RNE cells, although broadly distributed, was usually densest around the nucleus and

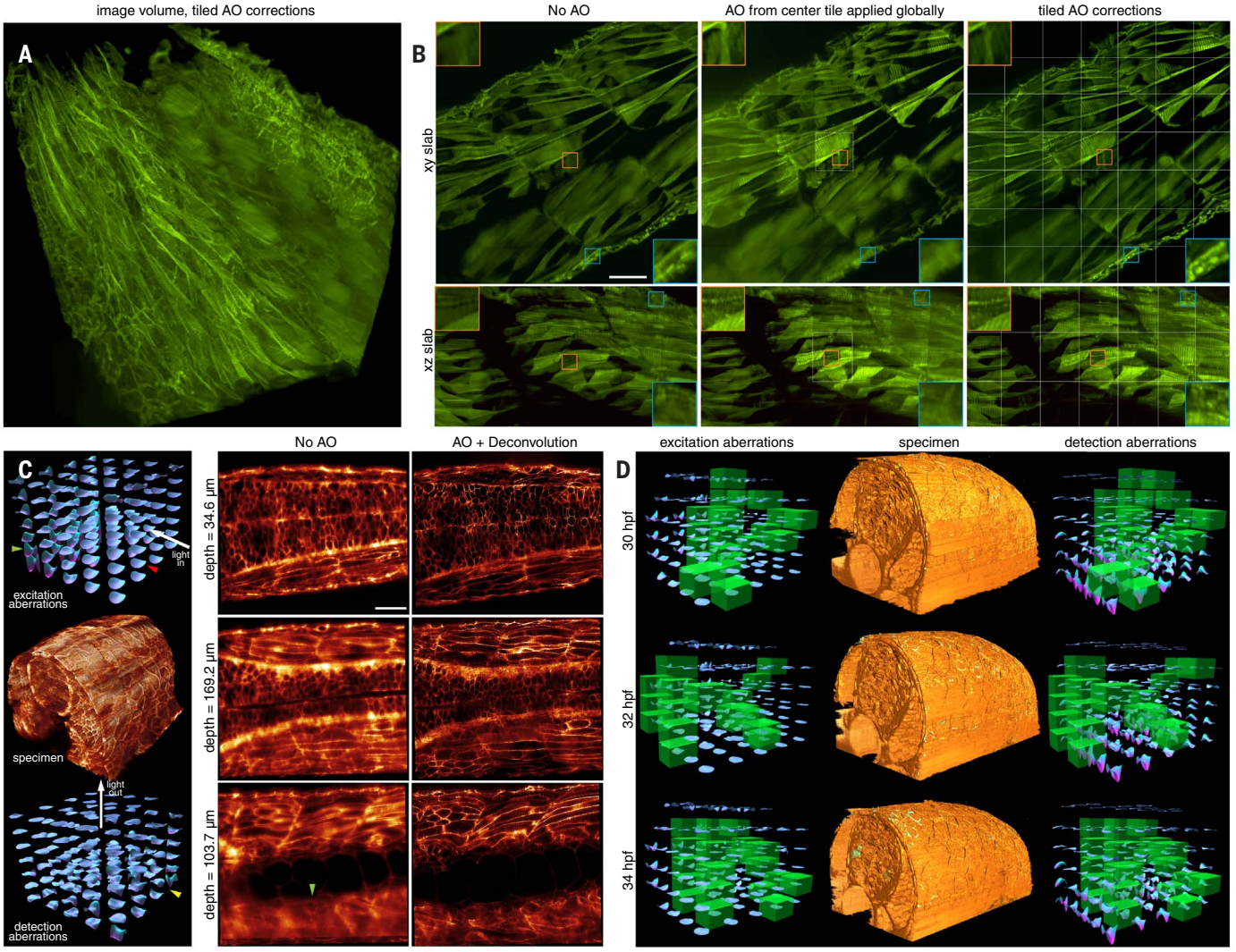


Fig. 4. AO-LLSM over large volumes. (**A**) Aberration-corrected volume rendering over 213 μm by 213 μm by 113 μm in the tail region of a live zebrafish embryo 96 hpf expressing PM-targeted EGFP, assembled from independently corrected subvolumes of 7 by 7 by 3 tiles (movie S4). (**B**) Increasing effectiveness of correction, as seen in orthogonal MIPs from 3-μm-thick slabs, under different scenarios: no AO (left column), AO correction from the center tile applied globally (middle column), and independent AO correction in each tile (right column) (fig. S11). Insets compare, at higher magnification, the quality of correction at the center tile (orange boxes) versus at the tiles at the periphery of the tail (blue boxes). Tile boundaries are shown in white. Scale bar, 30 μm. (**C**) A 5 by 4 by 7 set of measured excitation (left column, top) and detection (left column, bottom) aberrations which, after AO correction, yields diffraction-limited imaging over a 170-μm by 185-μm by 135-μm volume (left column, center) in the spine of a zebrafish embryo 30 hpf (Movie 6). Red and green arrowheads indicate excitation aberrations in specific tiles before and after passage through the notocord, respectively. The yellow arrowhead indicates a tile with a large detection aberration deep within the specimen. Orthoslices before (middle column) and after (right column) AO correction show increased aberration but continued recovery of high resolution at progressively greater depth. Scale bar, 30 μm. (**D**) Aberration-corrected volume renderings over 156 μm by 220 μm by 162 μm in the spine of a zebrafish embryo, at three points from a time series at 30-min intervals (movie S5), flanked by excitation and detection path aberrations at those points. Those tiles whose corrections were updated at a given time point are marked in green.

least dense near the polarized ends. In mitotic cells, however, we again observed (*20*) that the ER remodeled into sheetlike structures concentrated near the PM (Movie 7).

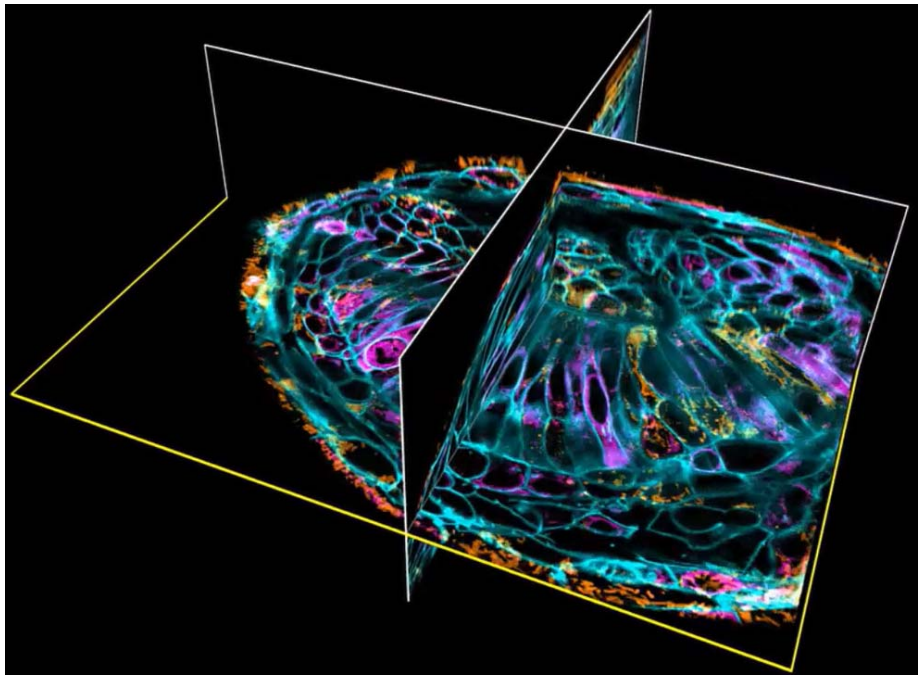
By imaging over time (Fig. 5F), we could follow the stages of RNE cell division (green arrowheads). As reported elsewhere (*25*), we found that, before mitosis, the nucleus retracts to the apical side of the retina (leftmost orange arrowheads), while the cell maintains a thin connection to the basal side (rightmost orange arrowheads). Despite its narrowness, mitochondria remain in this region. RNE cell divisions then occur at the apical surface (white arrowheads). This process has been found to be necessary to maintain the integrity of retinal tissue (*26*).

3D cell migration in vivo

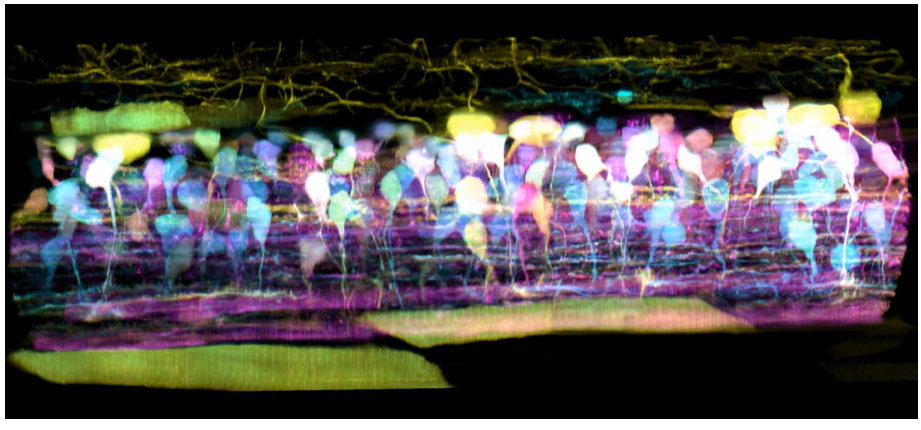
In vivo 3D migration of a cell in the densely crowded environment of living tissue involves forces, constraints, elasticity and adhesion heterogeneity, and chemical cues not found in the simple 2D environment on a cover glass. Furthermore, cell migration involves intricate and rapid remodeling of membranes, organelles, and the cytoskeleton that requires high spatiotemporal resolution to observe. It is therefore a problem well suited to AO-LLSM.

An example of this problem involves the wiring of neuronal circuits during development. To help them establish precise connections, axons are tipped with a highly complex and motile structure, the growth cone. This structure functions as both a sensor and a motor, driving the growth of neurites on the basis of environmental cues (*27*). Although its dynamics have been shown to be important for its proper function (*28*–*30*), its 3D dynamics in an intact animal have been difficult to study because of the lack of techniques for imaging the structure with sufficient resolution in vivo.

To address this, we used AO-LLSM to image growth cones in the spinal cord of a zebrafish embryo in which a subset of newly differentiated neurons expressed stochastic combinations of three different fluorophores via Autobow (*31*) (Fig. 6A and fig. S12), so that they could be spectrally distinguished from earlier differentiated neurons expressing only mCherry [e.g., those within the medial neuropil of the reticulospinal tract (magenta arrowheads, Fig. 6A)]. The Autobow-labeled neurons included Rohon-Beard sensory neurons in the dorsal spinal cord (e.g., yellow arrowheads, Fig. 6A) and interneurons with commissural axons. By imaging over more than two spinal segments at 10.4-min intervals from 58 to 70 hpf (Fig. 6B and Movie 8), we observed that the growth cones of axons migrating in the rostrocaudal direction primarily probed in the direction of their motion (Fig. 6C, top, and movie S6), whereas the growth cones of dorsoventrally aligned axons probed across a broader 2D fan (Fig. 6D, top). Transverse views (Fig. 6, C and D, bottom) revealed that most, if not all, growth cones of both types were located close to the surface of the spinal cord, with their filopodia preferentially aligned parallel to the surface,



Movie 7. Organelle dynamics across the zebrafish eye. PM (cyan), trans-Golgi (green), ER (magenta), and mitochondria (brown) across a 128-μm by 150-μm by 75-μm volume assembled from subvolumes of 4 by 4 by 3 tiles, showing orthoslices in a single tile, volume-rendered tiles before assembly into the combined volume, organelles in the combined volume, dynamics over 30 time points from 24.0 to 26.8 hpf in a 1-μm-thick slab through the combined volume, dynamics in perpendicular orthoslices, and organelle morphologies in different cell types in the computationally expanded volume (Fig. 5).



Movie 8. In vivo imaging of spinal cord neural circuit development. Autobow-labeled, newly differentiated neurons expressing stochastic combinations of three fluorophores in a zebrafish embryo, showing corrective excitation and detection wavefronts in subvolumes of 5 by 2 by 1 tiles, with scrolling updates at one tile (green box) per time point; AO-corrected orthoslices and volume-rendered views in each color channel 58 hpf; and axon pathfinding in each color channel from 58 to 70 hpf (Fig. 6, A to D; fig. S12; and movie S6).

even though the dorsoventrally projecting axons had to pass through the spinal cord to reach its surface. This is consistent with the previous observation that the neurites of late-born V2a ipsilateral projecting interneurons are located lateral to the preexisting ones, forming a layer-like organization based on the age of neurons (*32*), and extends this notion to other classes of spinal neurons. This also raises an interesting possibility that the shape of the growth cone is actively controlled in vivo to keep its exploration within a layer of its own age group.

Cell migration is also a key aspect of the innate immune system. Neutrophils, for example, migrate from the vasculature through the endothelium to reach and engulf infectious targets

(*33*). To study this process in vivo, we imaged endogenously produced immune cells moving through the perilymphatic space of the ear in a transgenic zebrafish larva ~80 hpf expressing the fluorescent protein Citrine in the PM of all cells (Fig. 6E and Movie 9). Acquiring 438 image volumes at 13-s intervals allowed us to accurately measure the 3D position and speed (fig. S13, A and B) of the cellular center of mass. Across five trials involving different embryos at 22°C, immune cells adopted a halting search pattern involving frequent changes in direction and speed (fig. S13, C to G), from nearly motionless to 10 μm/min. In contrast, neutrophilic mammalian HL-60 cells imaged in a collagen matrix at 37°C (*6*) exhibited peak speeds of ~25 μm/min (fig. S13H).

Immune cells next to the ear were remarkable for their rapidly changing and complex 3D morphologies (Fig. 6F). Their surface areas changed as much as 25% in 2 min (fig. S13I) as they remodeled themselves to exhibit a variety of protrusive features, including lamellar sheets, blebs, and short filopodia. Frequently, they also trailed a long filopodium behind them as they migrated to new regions (fig. S14). After injecting fluorescent Texas Red dextran into the heart, we also observed several immune cells containing granules of dextran up to several microns in size (light blue, Fig. 6F, and Movie 9, part 1), presumably

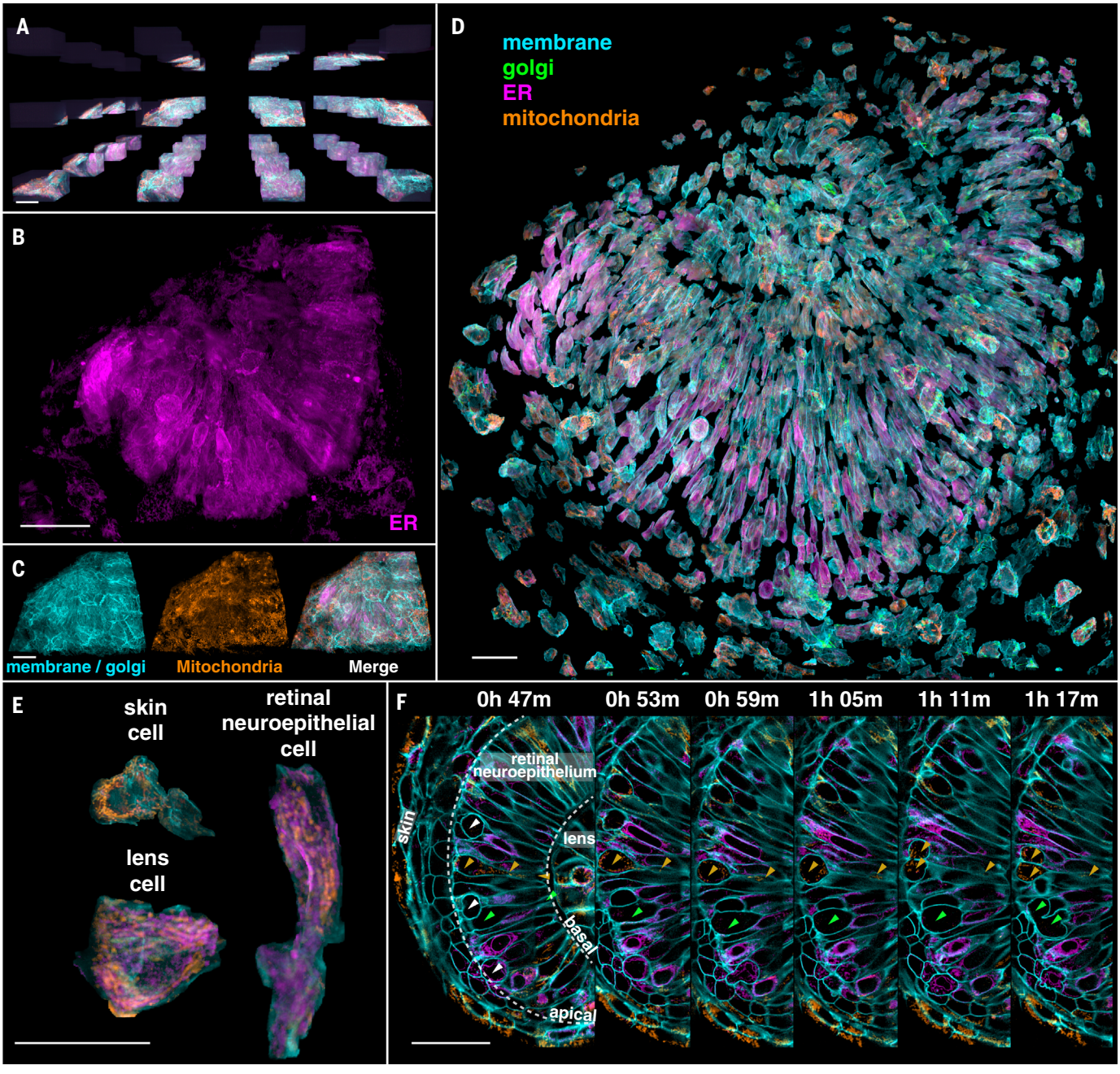


Fig. 5. Organelle diversity across the zebrafish eye. (**A**) Tiled array used to provide AO correction across the eye of a developing zebrafish embryo 24 hpf (Movie 7). Scale bar, 30 μm. (**B** and **C**) Distribution of three different types of organelles across the volume assembled from the tiles in (A). Scale bars, 30 μm. (**D**) Computationally separated cells across the eye, with the organelles colored as indicated. Scale bar, 30 μm. (**E**) Organelle morphologies in cells of three different types within the eye. Scale bar, 30 μm. (**F**) Orthoslices at six different time points highlighting cell divisions (white and green arrowheads, left panel) at the apical surface of the retinal neuroepithelium and mitochondria (orange arrowheads) present from the apical to the basal surface in one dividing cell. Scale bar, 30 μm.

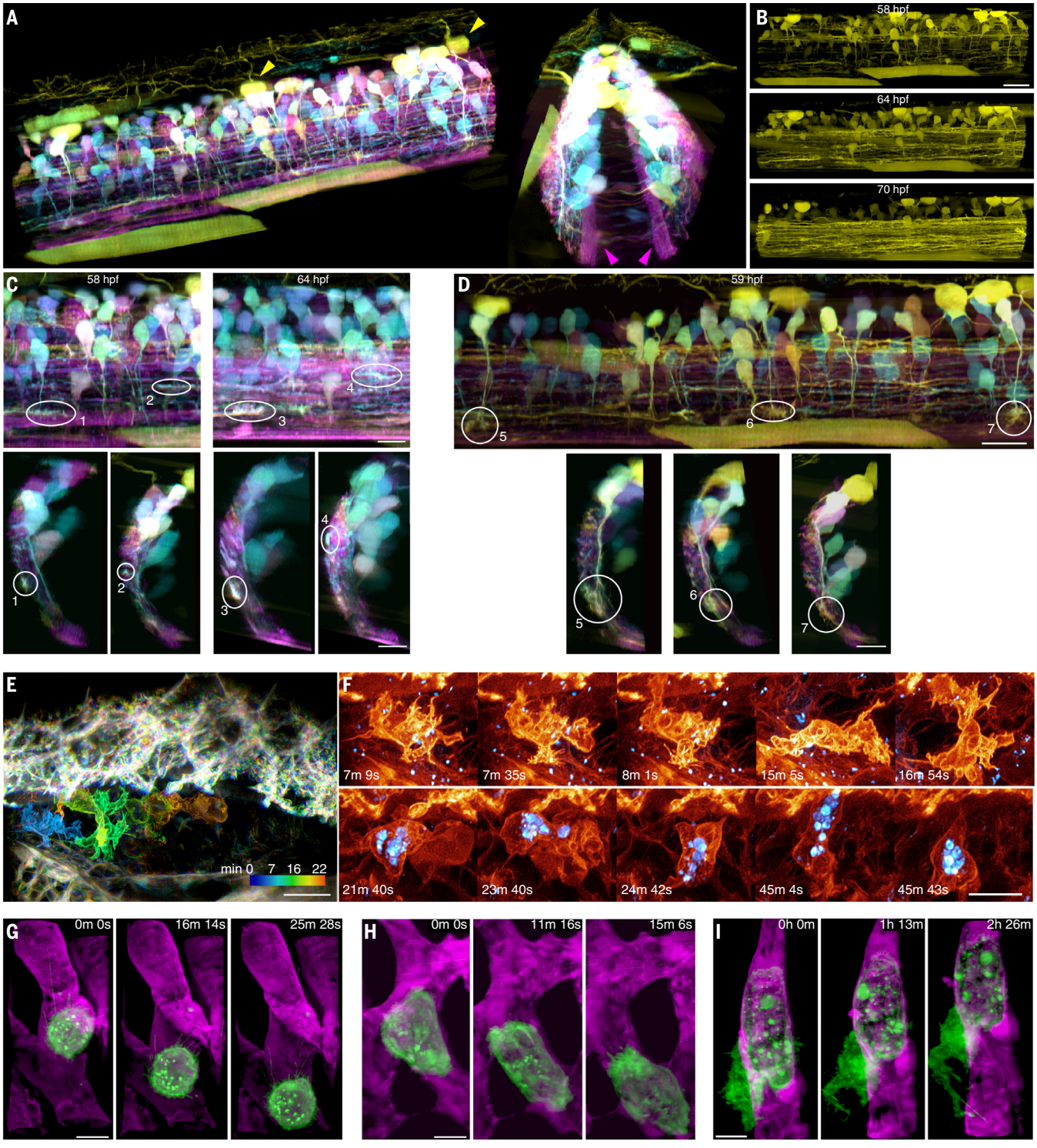


Fig. 6. 3D cell migration in vivo. (**A**) Two views of newly differentiated neurons highlighted by Autobow labeling in a 60-μm by 224-μm by 180-μm section of the spinal cord of a zebrafish embryo 58 hpf (Movie 8). Magenta and yellow arrowheads show neurons differentiated before and after Autobow expression, respectively. (**B**) Increase in the density of rostrocaudally projecting axons over time. Scale bar, 20 μm. (**C**) Sagittal (top) and transverse (bottom) views of the growth cones of four rostrocaudally projecting axons. Scale bars, 10 μm. (**D**) Sagittal (top) and transverse (bottom) views of the growth cones of three dorsoventrally projecting axons. Scale bars, 10 μm. (**E**) Time-coded color overlay of an immune cell migrating within the perilymphatic space next to the inner ear of a live transgenic zebrafish embryo 70 hpf expressing PM-targeted Citrine (Movie 9 and fig. S12). Texas Red dextran particles are shown in blue. Scale bar, 10 μm. (**F**) Changing morphologies of two different immune cells (top and bottom rows), one showing internalized dextran particles (blue) (fig. S13). Scale bar, 5 μm. (**G**) MDA-MB-231 human breast cancer cell (green) rolling in a blood vessel (magenta) in a zebrafish embryo 48 hpf. (**H**) Another MDA-MB-231 cell crawling through a blood vessel. (**I**) A partially extravasated MDA-MB-231 cell, showing an increasingly complex morphology over time (Movie 10 and fig. S1). Scale bars, 10 μm in (G) to (I).

ingested earlier by phagocytosis. However, these did not noticeably affect the motility of the cells or their ability to navigate through tight interstitial spaces.

Apart from immune cells, 3D AO-LLSM movies (e.g., Movie 9, part 2) of the developing ear region revealed a wealth of cellular morphologies and behaviors, including filopodial oscillations at the dorsal surface of skin cells, gradual inflation of the perilymphatic volume, rapid passage of cells through blood vessels, long and active filopodia on endothelial cells, and cellular motion in the hindbrain. With AO, the spatiotemporal resolution and noninvasiveness that we achieved was comparable to what we obtained when imaging cultured cells with our original LLSM approach (*6*), allowing us, for example, to follow the detailed morphological changes in a single endothelial cell lining the hindbrain over the entire course of its division (fig. S15 and Movie 9, part 2).

As a final example of 3D migration, during cancer metastasis, circulating tumor cells (CTCs) extravasate and seed new tumor formation at sites distant from the primary tumor (*34*). Extravasation has been studied extensively in vitro, but little is known about the process in vivo, owing to the highly dynamic nature of cells in circulation and the low density of CTCs in the vasculature. A long-standing hypothesis (*34*) based on in vitro studies is that CTCs co-opt a three-step process used by leukocytes to extravasate at sites of inflammation (*34*): Circulating leukocytes initially form tethers to weakly adhere to the vascular endothelium, causing them to roll slowly downstream (*35*); eventually, they attach and crawl along the endothelial wall; and finally, they penetrate the endothelium and migrate into the tissues beyond.

To determine whether CTCs follow this same pattern in vivo, we used a xenograft model (*36*), where we injected PM-labeled human breast cancer cells (MDA-MB-231) into the vasculature of 2-day-old zebrafish embryos that were transgenic for an endothelial reporter (*kdrl:gfp*). As hypothesized, we observed all three leukocyte migration behaviors in the cancer cells. First, we recorded MDA-MB-231 cells rolling through the blood vessels (Fig. 6G), trailing microvilli that adhered to the endothelium and stretched several microns before detaching as the cell continued to move downstream (Fig. 6G and Movie 10, part 1). Second, we visualized MDA-MB-231 cells crawling along the endothelium (Fig. 6H and Movie 10, part 2). Last, we observed an MDA-MB-231 cell actively engaged in transendothelial migration, with the portion of the cell outside the blood vessel projecting actin-rich extensions into the surrounding tissue (Fig. 6I and Movie 10, part 3) as the area of the cell increased by ~50% over 2 hours (fig. S16).

Discussion

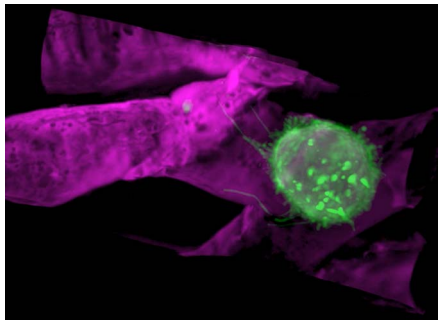
AO-LLSM enables minimally invasive high-speed 3D imaging of subcellular dynamics within optically challenging living specimens while maintaining diffraction-limited resolution, even over large FOVs. It corrects not only sample-induced aberrations, but also those introduced by mounting and immersion media (e.g., Fig. 1E) or imperfections in the optical path through the microscope. It therefore can provide practical 3D resolution exceeding that of nominally higher numerical aperture (NA) confocal or spinning disk microscopes, even in the comparatively benign optical environment encountered when imaging isolated adherent cells on cover slips.

This performance does not come without caveats, however. Because the fluorescence induced by the light sheet is captured with widefield optics, only weakly scattering specimens can be imaged. In addition, extremely sparse and/or weakly emitting fluorescent targets may require colabeling with a second, brighter color channel to provide a sufficient guide star signal for accurate wavefront measurement. Highly absorbing structures such as large blood vessels or melanin bodies within the detection light cone can block guide star light from reaching enough cells of the sensor for accurate wavefront measurement, although wavefront reconstruction algorithms might be made more robust against such missing information. Wavefront aberration can vary considerably across the specimen, and at present, this variation can only be determined empirically for each specimen type and developmental stage to determine how to subdivide the desired image volume into tiled isoplanatic subvolumes of relatively uniform aberration. Fortunately, such tile maps tend to be consistent between specimens of the same type and age, given similar mounting geometries. Lastly, specimens imaged after muscle development must be anesthetized and immobilized, or else a new correction must be measured and applied whenever sample motion exceeds the size of a given isoplanatic patch.

Another caveat is that all but one of the above described involved imaging subcellular dynamics within zebrafish embryos. Although we have shown that we can achieve substantial gains in imaging performance in both *Caenorhabditis elegans* larvae (fig. S17 and movie S7) and *Arabidopsis thaliana* leaves (fig. S18 and movie S8), *Danio rerio* represents an ideal model system in which to study cell and developmental biology in vivo, because it is a rapidly developing transparent vertebrate that is amenable to genetic manipulation. Furthermore, zebrafish embryos are small enough that most regions are optically accessible far into development, yet large enough to exhibit smoothly varying refractive index profiles that result in isoplanatic patch sizes that are comparable to imaging fields typical of LLSM. In contrast, *C. elegans* larvae and adults exhibit larger and more rapid spatial variations in refractive index, particularly near the gut, that can require a denser mesh of AO corrections, despite this organism's reputation as an optically tractable model.

Conventional light-sheet microscopy using weakly focused Gaussian beams is also susceptible to aberrations and would therefore also benefit from AO correction (*8*, *9*). However, conventional systems typically cover much wider FOVs and often operate at greater depth in larger organisms, such as in applications involving functional imaging of whole neural circuits (*37*) or in toto cellular tracking during development (*38*). They therefore usually image over regions much larger than a single isoplanatic patch, making it difficult to retain even cellular-level resolution at all locations and compromising the accuracy and resolution of approaches based on multiview fusion (*39*–*41*). A single AO correction would provide at best only partial correction, and a tiled AO approach, such as we use with LLSM, would negate the high-speed, large-field advantages of the conventional light-sheet microscopy. On the other hand, simultaneous full-field AO correction would likely require

Movie 9. Immune cell migration next to the zebrafish inner ear. Immune cells within the perilymphatic space of the inner ear of transgenic zebrafish embryos expressing PM-targeted Citrine 80 hpf, showing a MIP view before and after AO correction plus deconvolution of two immune cells (orange), one of which has ingested dextran particles (blue), for 438 time points at 13-s intervals; a volume-rendered view with a migrating immune cell and a dividing endothelial cell; and tracking of the position and velocity of an immune cell (Fig. 6, E and F, and figs. S13 to S15).



Movie 10. Cancer cell migration in a zebrafish xenograft model. MDA-MB-231 human breast cancer cells (green) exhibiting three different forms of motion within the vasculature (magenta) of different zebrafish embryos: rolling within a blood vessel while extending long, adhesive microvilli; crawling while conforming to the shape of a blood vessel; and partial extravasation from a blood vessel (Fig. 6, G to I, and fig. S16).

multiconjugate adaptive optics (42), substantially increasing cost and complexity.

Perhaps the greatest challenge of AO-LLSM involves mining the immense and complex data that it produces to extract as much biological insight as possible. In Fig. 6, for example, panels A to D represent a minute fraction of a 0.62-terabyte raw data set which first had to be deconvolved, creating a second copy, and then imported into 3D visualization software, generating a third. We deconvolve and store data in real time, but importation and visualization can take many hours, preventing meaningful real-time feedback on whether the biological structure and dynamic process of interest are being optimally recorded. If history is any guide, problems of petabyte-scale data storage and visualization at reasonable cost will yield to continued advancements in commercial hardware, but problems of image analysis and meaningful quantification of data at this scale may prove far less tractable. Although we have demonstrated quantification on a smaller scale through single-particle (Fig. 2, D and E, and fig. S8) and single-cell (fig. S13) tracking, segmentation (Figs. 2B, 3A, and 5D), and measurement of area and volume (Fig. 3, E and F, and figs. S9 and S16), the diversity of questions that can be asked when modern genetic and pharmacological tools are combined with high-resolution 5D in vivo data spanning hundreds of cells over many hours will demand bioinformatics expertise, machine learning, and custom algorithm development on an unprecedented level. Nevertheless, such efforts promise to offer insights into how cells harness their intrinsic variability to adapt to different physiological environments and have the potential to reveal the phenotypic diversity of organelle morphologies, intracellular dynamics, extracellular communication, and collective cell behavior across different cell types, organisms, and developmental stages.

Materials and methods

Lattice light-sheet subsystem

The lattice light-sheet excitation path of the AO-LLSM was designed as described previously (6). Noted here are the changes introduced in the AO-LLSM. The collinear laser beams from the combiner were first expanded using a pair of cylindrical lenses and aligned such that up to three different wavelengths illuminated three vertically separated thin stripes on spatial light modulator SLM, (Holoeye, PLUTO-Vis-014 1920 × 1080 pixels; fig. S1). As a grayscale phase modulation device, SLM was introduced to not only create the light sheet but to correct sample-induced aberrations as well. The diffraction orders reflected from SLM were then filtered using annular mask MSK (Photo Sciences) as before and conjugated to galvanometer scanning mirrors G3 and G4 (3 mm mirror, Cambridge Technology, 8315H) to scan the light sheet along the x and the z axis. During imaging, different offset voltages were applied to the z galvo to sequentially realign the light sheet from each laser to the same plane within the specimen. Sample plane conjugate resonant galvanometer RG (Electro-Optical Products Corp. 7 × 8 mm, SC-30) was also added prior to the excitation objective (Special Optics, 0.65 NA, 3.74 mm WD) to wobble the light sheet in the xy plane and thereby minimize stripe artifacts due to localized absorbing or scattering objects in the specimen. The fluorescence generated in the excitation plane was collected with detection objective DO (Nikon, CFI Apo LWD 25XW, 1.1 NA, 2 mm WD) and reflected off deformable mirror DM (ALPAO 97-15) conjugate to the rear focal plane of DO before being imaged at sCMOS camera CAM 1 (Hamamatsu Orca Flash 4.0 v2). Complete details of the optical design are given in supplementary note 1.

Adaptive optics subsystems

In principle, independent AO corrective systems are needed for excitation and detection in LLSM, since light traverses different regions of the specimen in each case and hence is subject to different aberrations. However, given that: (i) aberrations decrease quickly with decreasing numerical aperture (NA) (7); (ii) we use at most 0.6 NA for excitation, versus 1.1 NA for detection in LLSM; and (iii) only aberrations within a narrow annulus at the rear pupil of the excitation objective will affect a lattice light sheet, it is not obvious that AO correction of the light sheet itself is necessary. To check, we simulated the effect of aberrations consisting of random combinations of the 55 lowest-order Zernike modes up to a root mean square (RMS) amplitude of two wavelengths (λ). We found (fig. S19) that aberrations at this level could expand a 0.7-μm-thick lattice light sheet to as much as 20 μm, and displace it perpendicular to its plane by up to ±8 μm, indicating that correction of excitation as well as detection is essential.

Hence, during aberration measurement, light from Ti:Sapphire ultrafast pulsed laser 2PL (Coherent Cameleon Ultra II) was ported to either the excitation or detection arm by switching galvanometer SG1 (fig. S1). In either case, TPEF generated within the specimen by scanning the guide star focused by EO or DO was collected by the same objective, descanned (11) and sent to homebuilt Shack-Hartmann wavefront sensor ESH or DSH, each consisting of a square microlens array (Edmund Optics) focused onto an EMCCD camera (Andor iXon). Corrective wavefronts were then applied to SLM or DM as described in supplementary notes 3 or 2, respectively. Further hardware details are given in supplementary note 1.

Autofocus measurement was achieved by viewing, side-on through EO, both the light-sheet fluorescence and the plane of fluorescence generated by guide star TPE excitation through DO on camera CAM4, and correcting for any displacement between them as outlined in supplementary note 5.

Zebrafish immobilization, mounting, and imaging conditions

Zebrafish embryos were paralyzed with ~1 ng of α-bungarotoxin protein injected prior to imaging (43) or anesthetized using tricaine (0.16 mg/ml) for 15 min. 12-mm-diameter glass coverslips were precleaned as follows: ~20 coverslips were placed in a 50 ml Falcon tube containing 0.1M NaOH and the tube placed in a sonicator for 15 min, followed by at least 5 consecutives washes with Mili-Q water and then immobilized in the sample holder using superglue. An agarose holder containing narrow groves for mounting the embryos was created by solidifying a few drops of 0.5 to 2% (wt/wt) low-melting agarose between the coverslip and a mold containing ridges. For the immune cell experiments, larvae were placed in 3D-printed volcano-shaped mounts (https://www.shapeways.com/shops/megason-lab). A homemade hair-loop was used to position the embryo in the mold, which was then stabilized with a thin layer of agarose made by applying on top of the immobilized embryo ~10 to 20 μl 1% low-melting agarose at 37° to 40°C and then wicking the excess. After solidification, the sample holder was bolted onto a three-axis set of sample stages (Attocube, ECS3030 for x and y, ECS3050 for z) and submerged in a sample bath containing ~8 ml of 1x Danieau buffer. This assembly was then raised by a motorized actuator (Newport, LTA-HS Actuator, integrated with CONEX-CC Controller, CONEX-LTA-HS) until EO and DO were immersed in the media. The sample stages then positioned the desired FOV to the mutual focal point of the objectives. Detailed imaging conditions for each experiment discussed in the paper, including excitation power, imaging time, image, tile and voxel sizes, fluorophores, and proteins, are in table S1. Additional preparation conditions are discussed in supplementary note 6.

REFERENCES AND NOTES

1. J. Icha, M. Weber, J. C. Waters, C. Norden, Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* **39**, 1700003 (2017). doi: 10.1002/bies.201700003; pmid: 28749075
2. P. P. Laissue, R. A. Alghamdi, P. Tomancak, E. G. Reynaud, H. Shroff, Assessing phototoxicity in live fluorescence imaging. *Nat. Methods* **14**, 657–661 (2017). doi: 10.1038/nmeth.4344; pmid: 28661494
3. F. Pampaloni, E. G. Reynaud, E. H. Stelzer, The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.* **8**, 839–845 (2007). doi: 10.1038/nrm2236; pmid: 17684528
4. M. J. Pittet, R. Weissleder, Intravital imaging. *Cell* **147**, 983–991 (2011). doi: 10.1016/j.cell.2011.11.004; pmid: 22118457
5. R. Weigert, M. Sramkova, L. Parente, P. Amornphimoltham, A. Masedunskas, Intravital microscopy: A novel tool to study cell biology in living animals. *Histochem. Cell Biol.* **133**, 481–491 (2010). doi: 10.1007/s00418-010-0692-z; pmid: 20372919
6. B. C. Chen *et al.*, Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014). doi: 10.1126/science.1257998; pmid: 25342811
7. N. Ji, Adaptive optical fluorescence microscopy. *Nat. Methods* **14**, 374–380 (2017). doi: 10.1038/nmeth.4218; pmid: 28362438
8. D. Wilding, P. Pozzi, O. Soloviev, G. Vdovin, M. Verhaegen, Adaptive illumination based on direct wavefront sensing in a light-sheet fluorescence microscope. *Opt. Express* **24**, 24896–24906 (2016). doi: 10.1364/OE.24.024896; pmid: 27828430
9. C. Bourgenot, C. D. Saunter, J. M. Taylor, J. M. Girkin, G. D. Love, 3D adaptive optics in a light sheet microscope. *Opt. Express* **20**, 13252–13261 (2012). doi: 10.1364/OE.20.013252; pmid: 22714353

10. R. Aviles-Espinosa *et al*., Measurement and correction of in vivo sample aberrations employing a nonlinear guide-star in two-photon excited fluorescence microscopy. *Biomed. Opt. Express* **2**, 3135–3149 (2011). doi: 10.1364/BOE.2.003135; pmid: 22076274
11. K. Wang *et al*., Rapid adaptive optical recovery of optimal resolution over large volumes. *Nat. Methods* **11**, 625–628 (2014). doi: 10.1038/nmeth.2925; pmid: 24727653
12. L. A. Royer *et al*., Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nat. Biotechnol.* **34**, 1267–1278 (2016). doi: 10.1038/nbt.3708; pmid: 27798562
13. F. Arbizzani, M. Mayrhofer, M. Mione, Novel transgenic lines to fluorescently label clathrin and caveolin endosomes in live zebrafish. *Zebrafish* **12**, 202–203 (2015). doi: 10.1089/zeb.2015.1501; pmid: 25748563
14. M. Ehrlich *et al*., Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* **118**, 591–605 (2004). doi: 10.1016/j.cell.2004.08.017; pmid: 15339664
15. D. Li *et al*., Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015). doi: 10.1126/science.aab3500; pmid: 26315442
16. C. Kural *et al*., Dynamics of intracellular clathrin/AP1- and clathrin/AP3-containing carriers. *Cell Rep.* **2**, 1111–1119 (2012). doi: 10.1016/j.celrep.2012.09.025; pmid: 23103167
17. S. Vassilopoulos *et al*., Actin scaffolding by clathrin heavy chain is required for skeletal muscle sarcomere organization. *J. Cell Biol.* **205**, 377–393 (2014). doi: 10.1083/jcb.201309096; pmid: 24798732
18. F. Aguet *et al*., Membrane dynamics of dividing cells imaged by lattice light-sheet microscopy. *Mol. Biol. Cell* **27**, 3418–3435 (2016). doi: 10.1091/mbc.E16-03-0164; pmid: 27535432
19. Z. Kadlecova *et al*., Regulation of clathrin-mediated endocytosis by hierarchical allosteric activation of AP2. *J. Cell Biol.* **216**, 167–179 (2017). doi: 10.1083/jcb.201608071; pmid: 28003333
20. L. Lu, M. S. Ladinsky, T. Kirchhausen, Cisternal organization of the endoplasmic reticulum during mitosis. *Mol. Biol. Cell* **20**, 3471–3480 (2009). doi: 10.1091/mbc.E09-04-0327; pmid: 19494040
21. E. Boucrot, T. Kirchhausen, Endosomal recycling controls plasma membrane area during mitosis. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7939–7944 (2007). doi: 10.1073/pnas.0702511104; pmid: 17483462
22. J. Icha, C. Kunath, M. Rocha-Martins, C. Norden, Independent modes of ganglion cell translocation ensure correct lamination of the zebrafish retina. *J. Cell Biol.* **215**, 259–275 (2016). doi: 10.1083/jcb.201604095; pmid: 27810916
23. J. Sidhaye, C. Norden, Concerted action of neuroepithelial basal shrinkage and active epithelial migration ensures efficient optic cup morphogenesis. *eLife* **6**, e22689 (2017). doi: 10.7554/eLife.22689; pmid: 28372636
24. J. Icha *et al*., Using light sheet fluorescence microscopy to image zebrafish eye development. *J. Vis. Exp.* **110**, e53966 (2016).
25. F. C. Sauer, Mitosis in the neural tube. *J. Comp. Neurol.* **62**, 377–405 (1935). doi: 10.1002/cne.900620207
26. P. J. Strzyz *et al*., Interkinetic nuclear migration is centrosome independent and ensures apical cell division to maintain tissue integrity. *Dev. Cell* **32**, 203–219 (2015). doi: 10.1016/j.devcel.2014.12.001; pmid: 25600237
27. L. A. Lowery, D. Van Vactor, The trip of the tip: Understanding the growth cone machinery. *Nat. Rev. Mol. Cell Biol.* **10**, 332–343 (2009). doi: 10.1038/nrm2679; pmid: 19373241
28. D. Bentley, A. Toroian-Raymond, Disoriented pathfinding by pioneer neurone growth cones deprived of filopodia by cytochalasin treatment. *Nature* **323**, 712–715 (1986). doi: 10.1038/323712a0; pmid: 3773996
29. C. B. Chien, D. E. Rosenthal, W. A. Harris, C. E. Holt, Navigational errors made by growth cones without filopodia in the embryonic *Xenopus* brain. *Neuron* **11**, 237–251 (1993). doi: 10.1016/0896-6273(93)90181-P; pmid: 8352941
30. J. Q. Zheng, J. J. Wan, M. M. Poo, Essential role of filopodia in chemotropic turning of nerve growth cone induced by a glutamate gradient. *J. Neurosci.* **16**, 1140–1149 (1996). pmid: 8558243
31. D. Cai, K. B. Cohen, T. Luo, J. W. Lichtman, J. R. Sanes, Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547 (2013). doi: 10.1038/nmeth.2450
32. A. Kinkhabwala *et al*., A structural and functional ground plan for neurons in the hindbrain of zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1164–1169 (2011). doi: 10.1073/pnas.1012185108; pmid: 21199947
33. S. de Oliveira, E. E. Rosowski, A. Huttenlocher, Neutrophil migration in infection and wound repair: Going forward in reverse. *Nat. Rev. Immunol.* **16**, 378–391 (2016). doi: 10.1038/nri.2016.49; pmid: 27231052
34. N. Reymond, B. B. d'Água, A. J. Ridley, Crossing the endothelial barrier during metastasis. *Nat. Rev. Cancer* **13**, 858–870 (2013). doi: 10.1038/nrc3628; pmid: 24263189
35. A. Marki, K. Buscher, Z. Mikulski, A. Pries, K. Ley, Rolling neutrophils form tethers and slings under physiologic conditions in vivo. *J. Leukoc. Biol.* **103**, 67–70 (2017).
36. K. Stoletov *et al*., Visualizing extravasation dynamics of metastatic tumor cells. *J. Cell Sci.* **123**, 2332–2341 (2010). doi: 10.1242/jcs.069443; pmid: 20530574
37. M. B. Ahrens, M. B. Orger, D. N. Robson, J. M. Li, P. J. Keller, Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods* **10**, 413–420 (2013). doi: 10.1038/nmeth.2434; pmid: 23524393
38. R. Tomer, K. Khairy, F. Amat, P. J. Keller, Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nat. Methods* **9**, 755–763 (2012). doi: 10.1038/nmeth.2062; pmid: 22660741
39. R. K. Chhetri *et al*., Whole-animal functional and developmental imaging with isotropic spatial resolution. *Nat. Methods* **12**, 1171–1178 (2015). doi: 10.1038/nmeth.3632; pmid: 26501515
40. A. Kumar *et al*., Dual-view plane illumination microscopy for rapid and spatially isotropic imaging. *Nat. Protoc.* **9**, 2555–2573 (2014). doi: 10.1038/nprot.2014.172; pmid: 25299154
41. J. Swoger, P. Verveer, K. Greger, J. Huisken, E. H. Stelzer, Multi-view image fusion improves resolution in three-dimensional microscopy. *Opt. Express* **15**, 8029–8042 (2007). doi: 10.1364/OE.15.008029; pmid: 19547131
42. R. D. Simmonds, M. J. Booth, Modelling of multi-conjugate adaptive optics for spatially variant aberrations in microscopy. *J. Opt.* **15**, 094010 (2013). doi: 10.1088/2040-8978/15/9/094010
43. I. A. Swinburne, K. R. Mosaliganti, A. A. Green, S. G. Megason, Improved long-term imaging of embryos with genetically encoded α-bungarotoxin. *PLOS ONE* **10**, e0134005 (2015). doi: 10.1371/journal.pone.0134005; pmid: 26244658

ACKNOWLEDGMENTS

We thank the Shared Resource teams at the Janelia Research Campus for their skill and dedication in specimen handling and preparation, K. Schaefer for the preparation of cancer cells for the xenograph experiments, and the Instrument Design and Fabrication team for their manufacturing expertise. We also gratefully acknowledge the support of the Janelia Visitor Program. S.U. thanks H. Elliott, D. Richmond, and R. Gao for discussions and acknowledges the MATLAB code repository received from the Computational Image Analysis Workshop supported by NIH grant GM103792. **Funding:** T.-L.L., D.E.M., V.S., J.S., M.K., E.M.M., and E.B. are funded by the Howard Hughes Medical Institute (HHMI). T.K. and S.U. are funded by grants from Biogen, Ionis Pharmaceuticals, and NIH grant R01GM075252 (to T.K.). S.U. is a Fellow at the Image and Data Analysis core at Harvard Medical School. S.S. is funded by a Living Systems Institute start-up grant, University of Exeter. K.R.M., I.A.S., Z.M.C., T.W.H., and S.G.M. were supported by NIH grant R01DC015478. D.Q.M. is funded by the NIH (5R00CA154870-05 and 1R01GM121597-01). D.Q.M. and B.L.M. are funded by the Carol M. Baldwin Foundation and are Damon Runyon-Rachleff Innovators supported (in part) by the Damon Runyon Cancer Research Foundation (DRR-47-17). B.L.M. is also funded by the NSF (IOS1452928). D.H. is a Pew-Stewart Scholar for Cancer Research supported by the Pew Charitable Trusts and NIH grant R01CA196884. D.D. was supported by a Human Frontier Science Program Fellowship. D.G.D. was supported by NIH grant R35GM118149. **Author contributions:** E.B. supervised the project and wrote the manuscript with input from all coauthors. T.-L.L. built the microscope with input from E.B., D.E.M., and K.W. and performed all microscope characterization experiments. D.E.M. created the instrument control software. T.-L.L., V.S., and S.U. acquired all biological data with coauthors. D.D., D.G.D., R.F., and D.H. provided organoids and led related experiments. T.K., S.U., B.C., and S.S. provided transgenic zebrafish and AP2 cells and led related clathrin experiments. Z.M.C., T.W.H., and S.G.M. created the mCardinal-PM transgenic zebrafish. T.K., S.G.M., S.U., and I.A.S. provided zebrafish and led related organelle dynamics and in vivo immune cell experiments. J.S. and M.K. created the Autobow zebrafish, and M.K. led related experiments. A.Q.K., T.N.M., and D.Q.M. provided *C. elegans* specimens and led related experiments. B.L.M. provided MDA-MB-231 cells and vasculature-labeled zebrafish and led related experiments with D.Q.M. Y.R., H.Y., and E.M.M. provided *Arabidopsis* specimens and led related experiments. S.U., K.R.M., T.-L.L., and V.S. processed all image data. S.U. performed quantitative analysis of all image data. S.U., T.-L.L., and E.B. produced all figures and movies. **Competing interests:** Portions of the technology described herein are covered by U.S. Patent 7,894,136 issued to E.B., assigned to Lattice Light of Ashburn, VA, and licensed to Carl Zeiss Microscopy; U.S. Patents 8,711,211 and 9,477,074 issued to E.B., assigned to HHMI, and licensed to Carl Zeiss Microscopy; U.S. Patent application 13/844,405 filed by E.B. and K.W. and assigned to HHMI; and U.S. Patent 9,500,846 issued to E.B. and K.W. and assigned to HHMI. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. Documentation for construction of a lattice light-sheet microscope can be obtained by execution of a research license agreement with HHMI.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/eaaq1392/suppl/DC1
Supplementary Text
Figs. S1 to S20
Table S1
References (*44*–*75*)
Movies S1 to S8

5 October 2017; accepted 19 February 2018
10.1126/science.aaq1392

RESEARCH ARTICLE

QUANTUM OPTICS

Multidimensional quantum entanglement with large-scale integrated optics

Jianwei Wang,[1,2]*† Stefano Paesani,[1]* Yunhong Ding,[3,4]*† Raffaele Santagati,[1] Paul Skrzypczyk,[5] Alexia Salavrakos,[6] Jordi Tura,[7] Remigiusz Augusiak,[8] Laura Mančinska,[9] Davide Bacco,[3,4] Damien Bonneau,[1] Joshua W. Silverstone,[1] Qihuang Gong,[2] Antonio Acín,[6,10] Karsten Rottwitt,[3,4] Leif K. Oxenløwe,[3,4] Jeremy L. O'Brien,[1] Anthony Laing,[1]† Mark G. Thompson[1]†

The ability to control multidimensional quantum systems is central to the development of advanced quantum technologies. We demonstrate a multidimensional integrated quantum photonic platform able to generate, control, and analyze high-dimensional entanglement. A programmable bipartite entangled system is realized with dimensions up to 15 × 15 on a large-scale silicon photonics quantum circuit. The device integrates more than 550 photonic components on a single chip, including 16 identical photon-pair sources. We verify the high precision, generality, and controllability of our multidimensional technology, and further exploit these abilities to demonstrate previously unexplored quantum applications, such as quantum randomness expansion and self-testing on multidimensional states. Our work provides an experimental platform for the development of multidimensional quantum technologies.

As a generalization of two-level quantum systems (qubits), multidimensional quantum systems (qudits) exhibit distinct quantum properties and can offer improvements in particular applications. For example, qudit systems allow higher capacity and noise robustness in quantum communications (*1*–*3*), can be used to strengthen the violations of generalized Bell and Einstein-Podolsky-Rosen (EPR) steering inequalities (*4*–*6*), provide richer resources for quantum simulation (*7*, *8*), and offer higher efficiency and flexibility in quantum computing (*9*, *10*). Moreover, encoding and processing qudits can represent a more viable route to larger Hilbert spaces. These advantages motivate the development of multidimensional quantum technologies in a variety of systems, such as photons (*11*, *12*), superconductors (*8*, *13*), and atomic systems (*14*, *15*). Unlike superconducting and atomic qudits, which cannot be encoded and manipulated without complex interaction engineering and control sequences, photons represent a promising platform able to naturally encode and process qudits in various degrees of freedom, such as orbital angular momentum (OAM) (*11*, *12*), temporal modes (*3*, *16*), and frequency (*17*, *18*). Previous work on qudits includes realizations of complex entanglement (*19*), entanglement in ultrahigh dimensions (*20*), and practical applications in quantum communication (*1*–*3*) and computing (*7*–*9*). However, these approaches incur limitations in terms of controllability, precision, and universality, which represent bottlenecks for further developments of multidimensional technologies. For example, the arbitrary generation of high-dimensional entanglement is a key experimental challenge, typically relying on complex bulk-optical networks and postselection schemes (*12*, *16*–*18*). In general, these approaches lack the ability to perform arbitrary multidimensional unitary operations with high fidelity (*16*, *19*), an important factor in quantum information tasks. Integrated microring resonators able to emit multidimensional OAM (*21*) and frequency (*18*) states have been reported, but these present limited fidelity and difficulties for on-chip state control and analysis, thus not fully exploiting the high precision, scalability, and programmability of integrated optics.

We report a multidimensional integrated quantum photonic device that is able to generate, manipulate, and measure multidimensional entanglement fully on-chip with unprecedented precision, controllability, and universality. Path-encoded qudits are obtained in which each photon exists over d spatial modes simultaneously, and entanglement is produced by a coherent and controllable excitation of an array of d identical photon-pair sources. Our device allows the generation of multidimensional entangled states with an arbitrary degree of entanglement. Universal operations on path-encoded qudits are possible in linear optics for any dimension (*22*, *23*), and our device performs arbitrary multidimensional projective measurements with high fidelity. The capabilities achieved allow us to demonstrate high-quality multidimensional quantum correlations, verified by generalized Bell and EPR steering violations, and to implement unexplored multidimensional quantum information protocols.

Large-scale integrated quantum photonic circuit

Entangled path-encoded qubits can be generated by coherently pumping two photon-pair sources produced by spontaneous parametric down-conversion (*24*) or by spontaneous four-wave mixing (SFWM) (*25*). The approach can be generalized to qudits via the generation of photons entangled over d spatial modes by coherently pumping d sources (*24*). However, scaling this approach to high numbers of dimensions has been challenging because of the requirement for a stable and scalable technology able to coherently embed large arrays of identical photon sources and to precisely control qudit states in large optical interferometers.

Silicon quantum photonics offer intrinsic stability (*26*), high precision (*27*), and dense integration (*28*) and can therefore provide a natural solution. We devised a large-scale silicon quantum photonic circuit to implement the scheme (Fig. 1A). A total of 16 SFWM sources are coherently pumped, generating a photon pair in a superposition across the array. Because both photons must originate from the same source, the bipartite state created is $\sum_{k=0}^{d-1} c_k |1\rangle_{i,k} |1\rangle_{s,k}$, where $|1\rangle_{i,k}$ and $|1\rangle_{s,k}$ respectively indicate the Fock state of the idler and signal photon in the kth spatial mode, and c_k represents the complex amplitude in each mode (with $\sum |c_k|^2 = 1$). The mapping between the Fock state of each photon and the logical state is as follows: We say that the qudit state is $|k\rangle$ ($k = 0, \ldots, d-1$) if the associated photon is in its kth optical mode. This yields a multidimensional entangled state of the form

$$|\psi\rangle_d = \sum_{k=0}^{d-1} c_k |k\rangle_i |k\rangle_s \qquad (1)$$

where the coefficients c_k can be arbitrarily chosen by controlling the pump distribution over the d sources and the relative phase on each mode. This is achieved with a network consisting of Mach-Zehnder interferometers (MZIs) at

[1]Quantum Engineering Technology Labs, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1FD, UK. [2]State Key Laboratory for Mesoscopic Physics, School of Physics, Collaborative Innovation Center of Quantum Matter, Peking University, Beijing 100871, China. [3]Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. [4]Center for Silicon Photonics for Optical Communication, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. [5]H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK. [6]ICFO–Institut de Ciencies Fotoniques, Barcelona Institute of Science and Technology, 08860 Castelldefels, Spain. [7]Max-Planck-Institut für Quantenoptik, 85748 Garching, Germany. [8]Center for Theoretical Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland. [9]QMATH, Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen Ø, Denmark. [10]ICREA–Institucio Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain.
*These authors contributed equally to this work.
†Corresponding author. Email: jianwei.wang@bristol.ac.uk (J.W.); yudin@fotonik.dtu.dk (Y.D.); anthony.laing@bristol.ac.uk (A.L.); mark.thompson@bristol.ac.uk (M.G.T.)

the input and phase shifters on each mode. In particular, maximally entangled states $|\psi_d^+\rangle = \sum_{k=0}^{d-1}|k\rangle_i|k\rangle_s/\sqrt{d}$ can be obtained with a uniform excitation of the sources. The two nondegenerate photons are deterministically separated using asymmetric MZI filters and routed by a network of waveguide crossings, grouping the signal photon into the top modes and the idler photon into the bottom modes (Fig. 1A). We can then locally manipulate and measure the state of each qudit. Linear optical circuits enable the implementation of any local unitary transformation $\hat{U}_d$ in dimension d (*22*, *23*). A triangular network of MZIs and phase shifters is used, which allows us to perform arbitrary local projective measurements, and two detectors are used to measure the outcomes. In this scheme, the measurement outcomes on a specific basis are collected sequentially by rotating the qudit reference frame and using one detector per photon. The collection of the d^2 outcomes thus requires d^2 detections in total. For more general implementations, the simultaneous collection of all the outcomes can be achieved via universal qudit operations (*23*) and the detection of each photon on all the output modes with $2d$ detectors (Fig. 1A, inset).

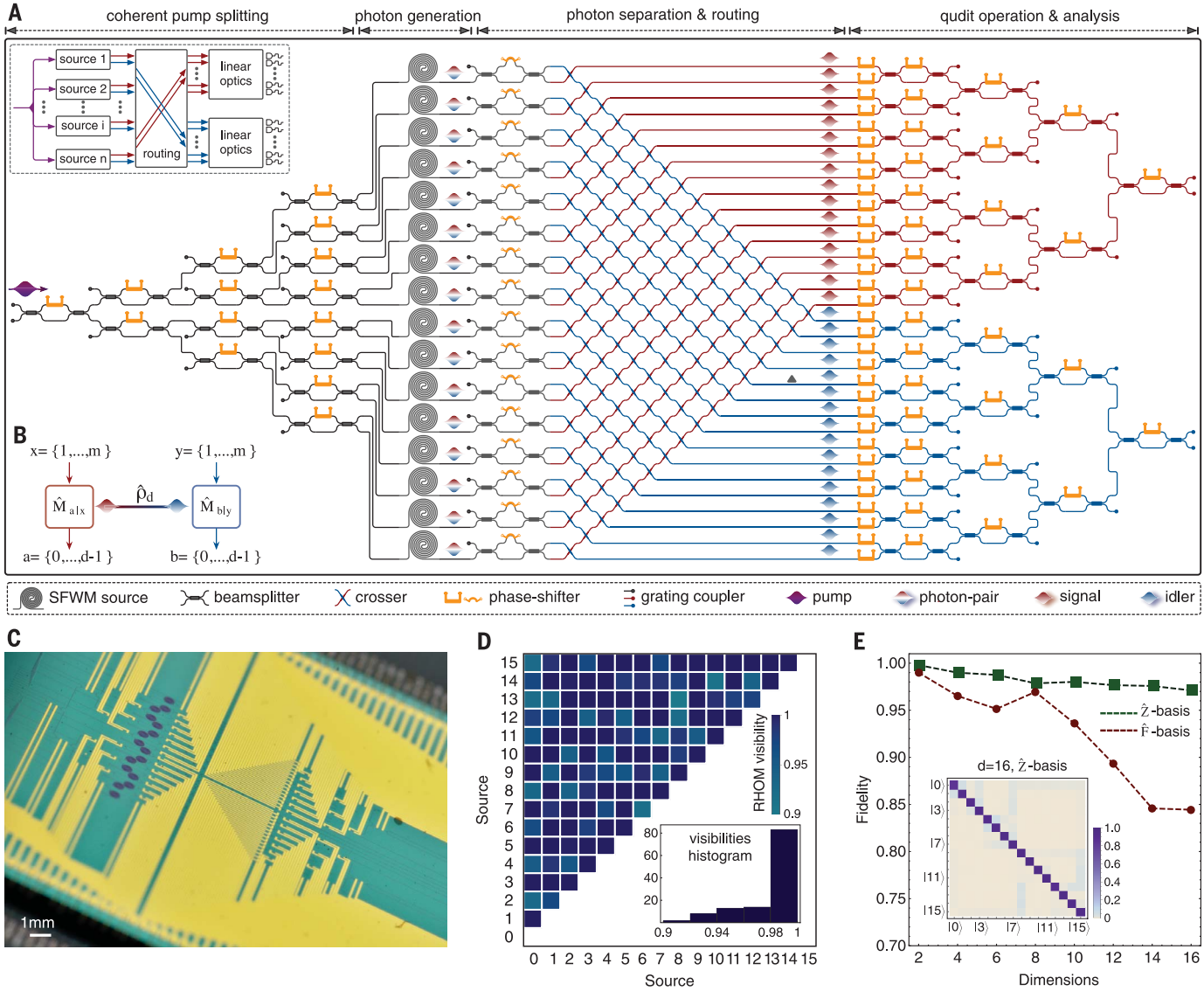


Fig. 1. Diagram and characterization of the multidimensional silicon quantum photonic circuit. (**A**) Circuit diagram. The device monolithically integrates 16 SFWM photon-pair sources, 93 thermo-optical phase shifters, 122 multimode interferometer (MMI) beamsplitters, 256 waveguide crossers, and 64 optical grating couplers. A photon pair is generated by SFWM in superposition across 16 optical modes, producing a tunable multidimensional bipartite entangled state. The two photons, signal and idler, are separated by an array of asymmetric MZI filters and routed by a network of crossers, allowing the local manipulation of the state by linear optical circuits. Using triangular networks of MZIs, we perform arbitrary local projective measurements. The photons are coupled off the chip into fibers by means of grating couplers and are detected by two superconducting nanowire detectors. The inset represents a general schematic for universal generation and manipulation of bipartite multidimensional entangled states. (**B**) Framework for correlation measurements on a shared d-dimensional state $\hat{\rho}_d$. $\hat{M}_{a|x}$ and $\hat{M}_{b|y}$ represent the operators associated with local measurements x on Alice and y on Bob, with outcomes a and b, respectively. (**C**) Photograph of the device. Silicon waveguides and 16 SFWM sources can be observed as black spirals. Gold wires allow the electronic access of each phase shifter. (**D**) Visibilities for the two-photon RHOM experiments to test sources' indistinguishability. The inset shows the histogram of all 120 measured visibilities, with a mean value of 0.984 ± 0.025. (**E**) Statistical fidelity for d-dimensional projectors in both the computational $\hat{Z}$-basis and the Fourier $\hat{F}$-basis. The inset shows the measured distribution for the 16-dimensional projector in the $\hat{Z}$-basis.

See (*29*) for more details of the device and experimental setup.

The 16 photon-pair sources are designed to be identical. Two-photon reversed Hong-Ou-Mandel (RHOM) interference is used to verify their performance, where the fringe visibility gives an estimate of the sources' indistinguishability (*26*). RHOM interference is tested between all the possible pairs of the 16 sources, performing $\binom{16}{2} = 120$ quantum interference experiments and evaluating the corresponding visibilities. The pair of sources used for each interference experiment is selected each time by reconfiguring the interferometric network. The observed rate of photon-pair coincidences was approximately 2 kHz in typical measurement conditions, from which accidentals were subtracted. For the measured visibilities (Fig. 1D), in all cases we obtained a visibility greater than 0.90, and more than 80% of cases presented a visibility of >0.98. These results show a state-of-the-art degree of source indistinguishability in all 120 RHOM experiments, leading to the generation of high-quality entangled qudit states.

Each of the MZIs and phase shifters can be rapidly reconfigured (kHz rate) with high precision (*26*, *28*). The quality of the qudit projectors is characterized by the classical statistical fidelity, which quantifies the output distribution obtained preparing and measuring a qudit on a fixed basis. We measured the fidelity of projectors in dimension d = 2 to 16 in the computational basis $\hat{Z} = |k\rangle\langle k|$ as well as in the Fourier-transform basis $\hat{F} = |\ell\rangle\langle\ell|$, where $|\ell\rangle = \sum_{k=0}^{d-1}\exp(2\pi ik\ell/d)|k\rangle/\sqrt{d}$ and k, ℓ = 0, ..., d – 1 (Fig. 1E). For d = 8, we observed fidelities of 98% in the $\hat{Z}$-basis and 97% in the $\hat{F}$-basis; for d = 16, fidelities were 97% in the $\hat{Z}$-basis and 85% in the $\hat{F}$-basis (*29*). The residual imperfections are mainly due to thermal cross-talk between phase shifters (higher in the $\hat{F}$-basis), which can be mitigated using optimized designs for the heaters (*28*) or ad hoc characterization techniques (*23*).

Because of a fabrication imperfection in the routing circuit, one of the modes (triangle label in Fig. 1A) for the idler photon presents an additional 10-dB loss. For simplicity, we exclude this lossy mode in the rest of our experiments and study multidimensional entanglement for dimensions up to 15.

Figure 1B represents the experiment in the standard framework for bipartite correlation. The correlations between two parties, Alice (A) and Bob (B)—here identified by the signal and idler photon, respectively—are quantified by joint probabilities $p(ab|xy) = \mathrm{Tr}[\hat{\rho}_d(\hat{M}_{a|x} \otimes \hat{M}_{b|y})]$, where $\hat{\rho}_d$ is the shared d-dimensional state; $x, y \in \{1, ..., m\}$ represent the m measurement settings chosen by Alice and Bob; and $a, b \in \{0, ..., d-1\}$ denote the possible outcomes with associated measurement operators $\hat{M}_{a|x}$ and $\hat{M}_{b|y}$. The joint probabilities for each measurement are calculated by normalizing the coincidence counts over all the d^2 outcomes in a given basis.

Quantum state tomographies

Quantum state tomography (QST) allows us to estimate the full state of a quantum system, providing an important diagnostic tool. In general, performing a complete tomography is an expensive task both in terms of the number of measurements and the computational time to reconstruct the density matrix from the data. For these reasons, complete QST on entangled qudit states has been achieved only up to eight-dimensional systems (*30*). To perform the tomographic reconstructions of larger entangled states, we use quantum compressed sensing techniques. Inspired by advanced classical methods for data analysis, these techniques reduce the experimental cost for state reconstruction (*31*), are general for density matrices of arbitrary dimension (*32*), and have been experimentally demonstrated to characterize complex quantum systems (*32*, *33*). Compressed sensing QST was implemented to reconstruct bipartite entangled states with local dimension up to d = 12. Fidelities with ideal states $|\psi_d^+\rangle$ are reported in Fig. 2A. For dimensions d = 4, 8, and 12, we plot the reconstructed density matrices, with fidelities of 96%, 87%, and 81%, respectively (Fig. 2, B to D). These results show an improvement of the quality for multidimensional entanglement (*18*, *30*). See (*29*) for more details.

Certification of system dimensionality

The dimension of a quantum system quantifies its ability to store information and represents a key resource for quantum applications. Device-independent (DI) dimension witnesses enable us to set a lower bound for the dimension of a quantum system solely from the observed statistics [i.e., correlation probabilities $p(ab|xy)$], making no prior assumptions on the experimental apparatus [see, e.g., (*34*, *35*)]. Here, we adopt the approach of (*35*) to verify the local dimension of entangled states in a DI way in the context where

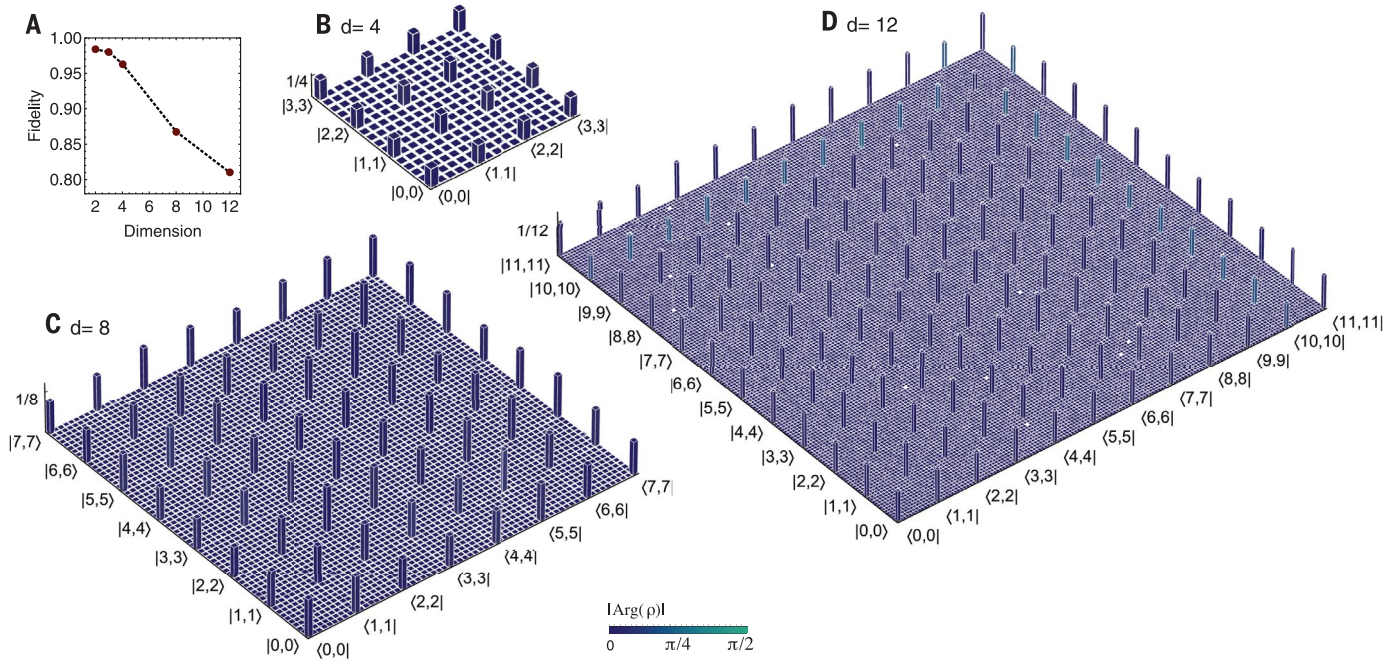


Fig. 2. Experimental quantum state tomographies. (**A**) Measured quantum fidelities $\langle\psi_d^+|\hat{\rho}_d|\psi_d^+\rangle$, where $\hat{\rho}_d$ represents the reconstructed states and $|\psi_d^+\rangle$ refers to the ideal d-dimensional maximally entangled state. (**B** to **D**) Reconstructed density matrices for the entangled states in dimension 4 (B), 8 (C), and 12 (D) using compressed sensing techniques. Column heights represent the absolute values $|\rho|$; colors represent the phases $|\mathrm{Arg}(\rho)|$. The phase information for matrix elements with module $|\rho_{ij}| < 0.01$ is approximately randomly distributed but is not displayed for clarity.

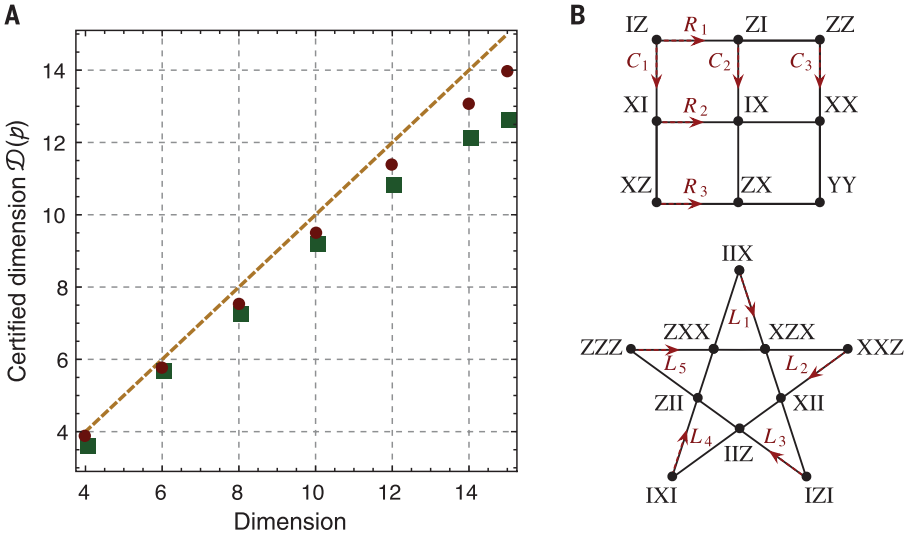


Fig. 3. Verification of system dimensionality. (**A**) Experimental results. Data points refer to the measured lower bounds on the local dimension of the generated entangled states; green and red points represent data for measurement scenarios I and II, respectively. The yellow line refers to ideal values. Errors are smaller than the markers and are neglected here for clarity. (**B**) Correlation measurements associated with optimal strategies for Magic Square and Magic Pentagram games. *X*, *Y*, and *Z* are Pauli operators and *I* is the identity. Red lines C_i, R_i, and L_i are associated with different measurement settings. A single Magic Square game, a single Magic Pentagram game, and two copies of the Magic Square game are used in an attempt to certify the dimension for states with local dimension $d = 4$, $d = 6$ or 8, and $d = 10$, 12, 14, or 15, respectively. The correct dimension is certified for up to $d = 10$ when using measurement scenario I and up to $d = 14$ when using scenario II.

shared randomness is not a free resource. The lower bound on the system dimension is given by $\lceil \mathcal{D}(p) \rceil$, where $\mathcal{D}(p)$ is a nonlinear function of the correlations, and $\lceil \varepsilon \rceil$ indicates the least integer greater than or equal to ε. Whereas with d we indicate the local dimension of the qudit encoded in d modes, $\lceil \mathcal{D}(p) \rceil$ represents the certified dimension of the quantum system—that is, the minimum dimension required to describe the observed correlations. We adopt two different DI measurement scenarios, with experimental results shown in Fig. 3A. In scenario I, we calculate the bound from the measured (partial) correlations for the Magic Square and Magic Pentagram games (*36*) (Fig. 3B). For example, to certify 8 × 8 entangled states (locally equivalent to a three-qubit system), we perform a $\hat{Z}$-basis measurement on Alice's system, while on Bob's system we use the $\hat{Z}$-basis and the one that simultaneously diagonalizes the commuting operators *ZZZ*, *ZXX*, *XZX*, and *XXZ* (Fig. 3B, lines L_3 and L_5, respectively). In the absence of noise, we would achieve $\mathcal{D} = 8$. Using the measured correlations, we obtain $\mathcal{D}(p_8^{\mathrm{I}}) \geq 7.22 \pm 0.05$, which yields the optimal lower bound $\lceil \mathcal{D}(p_8^{\mathrm{I}}) \rceil = 8$. In scenario II, we compute $\mathcal{D}(p_d^{\mathrm{II}})$ for correlations p_d^{II} obtained by performing $\hat{Z}$-basis measurements on both sides of the maximally entangled state of local dimension d. We expect less experimental noise in this scenario (see Fig. 1D). As shown in Fig. 3A, the experimentally observed correlations p_d^{II} yield $\lceil \mathcal{D}(p_d^{\mathrm{II}}) \rceil = d$ for all $d \leq 14$, certifying the correct dimensions. See (*29*) for further details.

Multidimensional Bell correlations and state self-testing

Bell inequalities enable experimental studies of quantum nonlocality, which indicates the presence of correlations incompatible with local hidden variables (LHV) theories. Nonlocality can be demonstrated by the violation of Bell inequalities of the form $S_d \leq C_d$, where S_d is a linear function of the joint probabilities and C_d is the classical bound for LHV models. Although our implementation does not represent a rigorous and loophole-free test of nonlocality, Bell inequalities are here used as an experimental tool to benchmark the quality of the multidimensional entanglement and to investigate possible future applications. We study two types of generalized Bell-type inequalities for d-dimensional bipartite systems: the SATWAP (Salavrakos-Augusiak-Tura-Wittek-Acin-Pironio) inequalities, recently introduced in (*6*), and the standard CGLMP (Collins-Gisin-Linden-Massar-Popescu) inequalities (*5*). In contrast to CGLMP inequalities, SATWAP inequalities are explicitly tailored to obtain a maximal violation for maximally entangled qudit states. Here, we test the two-input version of the SATWAP inequalities by measuring the joint probabilities to obtain the quantity

$$\tilde{I}_d = \sum_{i=1}^{2} \sum_{l=1}^{d-1} \langle A_i^l \bar{B}_i^l \rangle \qquad (2)$$

Table 1. Experimental values for multidimensional CGLMP Bell correlations. Measured CGLMP values are given with experimental errors. Values in parentheses refer to the LHV classical bound; those in braces refer to theoretical bounds for d-dimensional maximally entangled states. Errors are given by photon Poissonian noise.

Dimension	CGLMP S_d
2	(2) 2.810 ± 0.014 {2.828}
3	(2) 2.845 ± 0.012 {2.873}
4	(2) 2.867 ± 0.014 {2.896}
5	(2) 2.763 ± 0.014 {2.910}
6	(2) 2.629 ± 0.010 {2.920}
7	(2) 2.532 ± 0.013 {2.927}
8	(2) 2.650 ± 0.012 {2.932}

where the $2(d-1)$ values $\langle A_i^l \bar{B}_i^l \rangle$ represent generalized Bell correlators, whose explicit form is given in (*29*). The Bell inequality here is given by $\tilde{I}_d \leq C_d$, where the bound for classical LHV models is $C_d = [3\cot(\pi/4d) - \cot(3\pi/4d)]/2 - 2$. The maximum value of $\tilde{I}$ obtainable with quantum states (known as the Tsirelson bound, Q_d) is known analytically for arbitrary dimensions and is given by $\tilde{I}_d \leq Q_d = 2d - 2$. This maximal violation is achieved with maximally entangled states (*6*).

In Fig. 4A we show the experimental values of the generalized correlators $\mathrm{Re}[\langle A_i^l \bar{B}_i^l \rangle]$. The correlation measurements are performed in the Fourier bases provided in (*29*). Figure 4B shows the obtained values of $\tilde{I}_d$ for dimensions 2 to 8, together with the analytical quantum and classical bounds. In all cases the classical bound is violated. In particular, in dimensions 2 to 4, a strong violation is observed, closely approaching Q_d.

We report in Table 1 the experimental values for the CGLMP inequalities. Again, strong violations of LHV models are observed. As an example, for $d = 4$ we observe $S_4 = 2.867 \pm 0.014$, which violates the classical bound (i.e., $C_d = 2$ for CGLMP inequalities) by 61.9σ and is higher than the maximal value achievable by two-dimensional quantum systems ($S_2 = 2\sqrt{2}$) by 2.8σ, indicating stronger quantumness for higher dimensions.

The near-optimal Bell violations enable the self-testing of multidimensional entangled states. The task of self-testing represents the DI characterization of quantum devices by a classical user, based solely on the observed Bell correlations (*37*, *38*), and thus does not require making any assumption about the devices being tested—a desirable attribute for practical quantum applications. If the maximal violation of a Bell inequality can only be achieved by a unique quantum state and set of measurements (up to local isometries), a near-optimal violation enables characterization of the experimental device. In (*6*) it was shown that the SATWAP inequality can be used to self-test the maximally entangled state of two qutrits $|\psi_3^+\rangle$; in particular, using a numerical approach

from (*39*), a lower bound on the state fidelity can be obtained from the measured value of $\tilde{I}_3$. In (*29*) we generalize it also for arbitrary qutrit states of the form $|00\rangle + \gamma|11\rangle + |22\rangle$ (up to normalization). In Fig. 4C, we report the experimental self-tested lower bounds on the fidelities for different values of $\gamma = 1$, 0.9, and $(\sqrt{11} - \sqrt{3})/2 \approx 0.792$. This is made possible by exploiting the capability of the device to generate multidimensional states with tunable entanglement. In particular, $\gamma = 1$ indicates $|\psi_3^+\rangle$, and $\gamma = (\sqrt{11} - \sqrt{3})/2$ represents the state that maximally violates the CGLMP inequality (*39*). We experimentally achieve self-tested lower bounds on the fidelities of 79.9%, 83.2%, and 68.0%, respectively, for the three states. Note that the certification of high fidelities in a self-testing context is achievable only in the presence of near-ideal experimental correlations. The measured self-tested fidelities are comparable with the reported values obtained from full tomographies in other experimental approaches (*18*, *30*). Although our device also provides high violations for dimensions higher than 3, it remains unknown whether the approach based on SATWAP inequalities can be generalized to self-test states in arbitrary dimension (*6*).

Multidimensional randomness expansion

Randomness is a key resource in many practical applications. Generating certified randomness, however, is a notoriously difficult problem. Quantum theory, being fundamentally nondeterministic, provides a natural solution. The probabilistic nature of measurements forms the basis of quantum random number generators (*40*). Remarkably, quantum theory can go one step further and allows a stronger form of certified randomness: Measurement statistics that exhibit nonlocal correlations are necessarily uncertain and contain randomness. This remains true even if some or all of the experimental apparatuses used to generate the nonlocal correlations are uncharacterized or untrusted (*41*). That is, the measurement statistics are random not only for the user of the devices but also for any other party who may have additional knowledge about the devices, such as a potential eavesdropper.

The two scenarios considered here are (i) randomness certified by Bell inequality violations, where two untrusted measuring apparatuses are used to generate randomness, providing a fully DI certification (*41*), and (ii) randomness certified by EPR steering inequality violations, where one trusted and one untrusted measuring apparatus are used, with randomness generated from the untrusted device, providing a one-sided DI (1SDI) certification (*42*).

The protocol in either case consists of performing n runs of a Bell or steering test, using a small seed of randomness to choose the measurement settings in each run. The violation of the corresponding Bell or steering inequality is then estimated from the raw data. The randomness of the string **s** of measurement outcomes of the untrusted apparatuses is then lower-bounded as a function of the observed violation. Because an initial seed of randomness is necessary, this process achieves randomness expansion as new private randomness is generated in the process. The randomness of the string **s** is quantified in terms of min-entropy $H_{\min} = -\log_2 P_g$, where P_g is the predictability of **s** (i.e., the probability that it can be correctly guessed).

To study DI randomness expansion, we use violations of the above SATWAP Bell inequalities

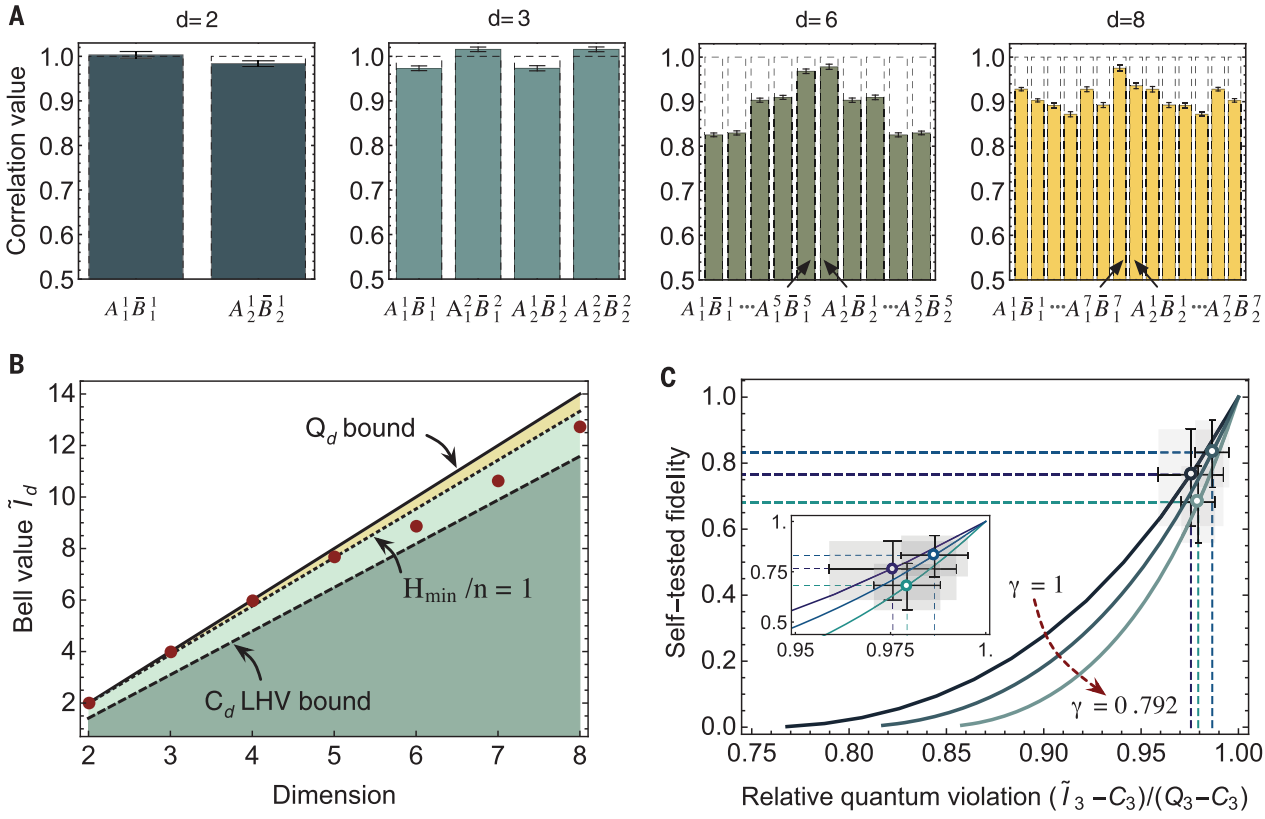


Fig. 4. Bell violation and self-testing on multidimensional entangled states. (**A**) Measured values of the $2(d-1)$ correlators $\mathrm{Re}[\langle A_i^l \bar{B}_i^l \rangle]$. Dashed boxes indicate theoretical values. (**B**) Violation of the generalized SATWAP Bell-type inequalities for d-dimensional states. Red points are experimentally measured $\tilde{I}_d$ values. Bell inequalities of the form $\tilde{I}_d \leq C_d$ are here violated, where C_d is the classical LHV bound (dashed line). The Tsirelson bound Q_d (solid line) represents the maximal violation for quantum systems. The dotted line represents the threshold above which more than 1 random bit can be extracted per output symbol from Bell correlations. (**C**) DI self-testing of entangled qutrit states $(|00\rangle + \gamma|11\rangle + |22\rangle)/\sqrt{2+\gamma^2}$ for $\gamma = 1$, 0.9, or 0.792. Self-tested lower bounds on the fidelities to ideal states are plotted as a function of the relative violation for more clarity. The significant uncertainty of the fidelity value is due to the general limited robustness of self-testing protocols. All errors are estimated from photon Poissonian statistics; those in (B) are smaller than markers.

(*6*), whereas in the 1SDI case we use violations of the EPR steering inequalities tailored for maximal randomness expansion recently introduced in (*29*, *42*):

$$\beta_d = \sum_{\substack{a=b \\ x=y}} p(a|x)\mathrm{Tr}[\hat{M}_{b|y}\hat{\rho}_{a|x}] \leq 1 + 1/\sqrt{d} \quad (3)$$

where $p(a|x)$ are the probabilities of Alice's uncharacterized measurements; $\hat{M}_{k|0} = |k\rangle\langle k|$ and $\hat{M}_{\ell|1} = |-\ell\rangle\langle -\ell|$ are the characterized measurements of Bob, with $|k\rangle$ corresponding to the $\hat{Z}$-basis and $|\ell\rangle$ to the $\hat{F}$-basis, defined above; and $\hat{\rho}_{a|x}$ indicates the reduced state for Bob when the measurement x is performed on Alice and outcome a is obtained. Quantum states can violate this inequality and maximally achieve $\beta_d = 2$. Figure 5A reports the experimentally measured values of β_d up to dimension 15, which display violations of the local bound in all dimensions. We note that this provides a 1SDI certification of the presence of bipartite multidimensional entanglement between Alice and Bob in all cases (*4*).

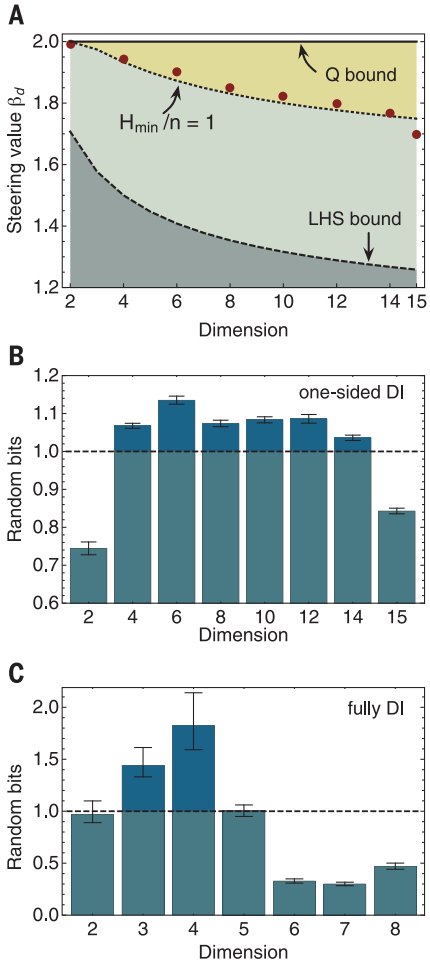


Fig. 5. Certification of multidimensional randomness expansion. (**A**) Multidimensional EPR steering is certified by violating the inequality $\beta_d \leq 1 + 1/\sqrt{d}$ (dashed line). Red points are experimentally measured steering values β_d. The dotted line denotes the threshold above which more than 1 random bit is generated per round from steering correlations. (**B**) Randomness per round certified in a one-sided DI scenario by *d*-dimensional steering correlations. (**C**) Randomness per round certified in a fully DI scenario by *d*-dimensional Bell correlations. Above the dashed line in (B) and (C), more than 1 private random bit is generated per round. Error bars are given by Poissonian statistics; those in (A) are smaller than markers.

In the DI setting, the measuring apparatuses of Alice and Bob are both untrusted, and the string of outcomes $\mathbf{s} = (\mathbf{a}, \mathbf{b})$, where $\mathbf{a}$ and $\mathbf{b}$ are the lists of data collected by Alice and Bob, respectively. In the 1SDI setting, the measuring apparatus of Alice is untrusted, and the string of outcomes is $\mathbf{s} = \mathbf{a}$. [See (*29*) for details of how randomness is certified, as well as the maximal theoretical amount of randomness that can be certified in each case.] A particularly demanding task is the efficient generation of randomness so as to generate more than one bit of randomness per experimental run (i.e., to achieve $H_{\min} > n$). For qubits, this is only possible using nonprojective measurements (with more than two outcomes) (*43*) or with sequences of measurements (*44*). In contrast, multidimensional entangled states provide a natural route based on projective measurements.

In Figs. 4B and 5A, the minimum values of $\tilde{I}_d$ and β_d above which more than one bit of randomness per run is certified in the DI or 1SDI setting are shown as a function of dimension (yellow regions). Figure 5C shows the randomness associated with the Bell violations shown in Fig. 4A. Efficiency $H_{\min}/n > 1$ is achieved for $d = 3$ and 4. The largest amount of randomness per run is obtained for $d = 4$, where $H_{\min}/n = 1.82 \pm 0.35$ random bits. The experimentally measured values of β_d are shown in Fig. 5A, and the associated randomness is reported in Fig. 5B. Here, efficiency $H_{\min}/n > 1$ is preserved for the range $4 \leq d \leq 14$, indicating, as expected, stronger robustness in the 1SDI case.

Conclusion

We have shown how silicon photonics quantum technologies have reached the maturity level that enables fully on-chip generation, manipulation, and analysis of multidimensional quantum systems. The achieved complexity of our integrated device represents a major step forward for large-scale quantum photonic technologies. Note that in the experimental tests performed here, the detection and locality loopholes were not closed, which was not an immediate goal of this work. However, the results demonstrate the unprecedented capabilities of multidimensional integrated quantum photonics, which will enable a wide range of practical applications. For example, high-rate, device-independent randomness generators can be realized by harnessing the abilities of efficient randomness expansion shown here, in combination with high-speed on-chip state manipulation. As a single-chip proof-of-principle demonstration of quantum key distribution, we report in (*29*) that Alice and Bob can share high-rate secure keys enabled by the high-fidelity control and analysis of the entangled qudits. Together with developed techniques for phase-coherent chip-to-chip qudit transmission—for example, via a multicore optical fiber (*45*), encoding in other degrees of freedom in free space (*46*), or exploiting reference frame–independent schemes (*47*)—our integrated platform can allow the development of high-dimensional quantum communications. All these possible applications can benefit from the monolithic integration of high-performance sources, universal operations, and detectors (*26*). Moreover, the scalability of silicon quantum photonics can further increase system dimensionality and allow the coherent control of multiple photons entangled over a large number of modes. Our results pave the way for the development of advanced multidimensional quantum technologies.

REFERENCES AND NOTES

1. N. J. Cerf, M. Bourennane, A. Karlsson, N. Gisin, *Phys. Rev. Lett.* **88**, 127902 (2002).
2. F. Bouchard, R. Fickler, R. W. Boyd, E. Karimi, *Sci. Adv.* **3**, e1601915 (2017).
3. N. T. Islam, C. C. W. Lim, C. Cahall, J. Kim, D. J. Gauthier, *Sci. Adv.* **3**, e1701491 (2017).
4. H. M. Wiseman, S. J. Jones, A. C. Doherty, *Phys. Rev. Lett.* **98**, 140402 (2007).
5. D. Collins, N. Gisin, N. Linden, S. Massar, S. Popescu, *Phys. Rev. Lett.* **88**, 040404 (2002).
6. A. Salavrakos *et al.*, *Phys. Rev. Lett.* **119**, 040402 (2017).
7. R. Kaltenbaek, J. Lavoie, B. Zeng, S. D. Bartlett, K. J. Resch, *Nat. Phys.* **6**, 850–854 (2010).
8. M. Neeley *et al.*, *Science* **325**, 722–725 (2009).
9. B. P. Lanyon *et al.*, *Nat. Phys.* **5**, 134–140 (2009).
10. A. Bocharov, M. Roetteler, K. M. Svore, *Phys. Rev. A* **96**, 012306 (2017).
11. A. Mair, A. Vaziri, G. Weihs, A. Zeilinger, *Nature* **412**, 313–316 (2001).
12. A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, E. Andersson, *Nat. Phys.* **7**, 677–680 (2011).
13. E. Svetitsky *et al.*, *Nat. Commun.* **5**, 5617 (2014).
14. C. Senko *et al.*, *Phys. Rev. X* **5**, 021026 (2015).
15. S. Choi, N. Y. Yao, M. D. Lukin, *Phys. Rev. Lett.* **119**, 183603 (2017).
16. A. Martin *et al.*, *Phys. Rev. Lett.* **118**, 110501 (2017).
17. Z. Xie *et al.*, *Nat. Photonics* **9**, 536–542 (2015).
18. M. Kues *et al.*, *Nature* **546**, 622–626 (2017).
19. M. Malik *et al.*, *Nat. Photonics* **10**, 248–252 (2016).
20. R. Fickler *et al.*, *Science* **338**, 640–643 (2012).
21. X. Cai *et al.*, *Science* **338**, 363–366 (2012).
22. M. Reck, A. Zeilinger, H. J. Bernstein, P. Bertani, *Phys. Rev. Lett.* **73**, 58–61 (1994).
23. J. Carolan *et al.*, *Science* **349**, 711–716 (2015).
24. C. Schaeff *et al.*, *Opt. Express* **20**, 16145 (2012).
25. R. Santagati *et al.*, *Sci. Adv.* **4**, eaap9646 (2018).
26. D. Bonneau, J. W. Silverstone, M. G. Thompson, *Silicon Quantum Photonics* (Springer, 2016), pp. 41–82.
27. J. Wang *et al.*, *Nat. Phys.* **13**, 551–555 (2017).
28. N. C. Harris *et al.*, *Nat. Photonics* **11**, 447–452 (2017).
29. See supplementary materials.
30. M. Agnew, J. Leach, M. McLaren, F. S. Roux, R. W. Boyd, *Phys. Rev. A* **84**, 062101 (2011).
31. D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, J. Eisert, *Phys. Rev. Lett.* **105**, 150401 (2010).
32. E. Bolduc, G. Gariepy, J. Leach, *Nat. Commun.* **7**, 10439 (2016).
33. C. A. Riofrío *et al.*, *Nat. Commun.* **8**, 15305 (2017).
34. M. Navascués, T. Vértesi, *Phys. Rev. Lett.* **115**, 020501 (2015).
35. J. Sikora, A. Varvitsiotis, Z. Wei, *Phys. Rev. Lett.* **117**, 060401 (2016).
36. N. D. Mermin, *Phys. Rev. Lett.* **65**, 3373–3376 (1990).
37. D. Mayers, A. Yao, in *39th Annual Symposium on Foundations of Computer Science (FOCS)* (1998), p. 503.
38. A. Coladangelo, K. T. Goh, V. Scarani, *Nat. Commun.* **8**, 15485 (2017).

39. T. H. Yang, T. Vértesi, J.-D. Bancal, V. Scarani, M. Navascués, *Phys. Rev. Lett.* **113**, 040401 (2014).
40. M. Herrero-Collantes, J. C. Garcia-Escartin, *Rev. Mod. Phys.* **89**, 015004 (2017).
41. S. Pironio *et al.*, *Nature* **464**, 1021–1024 (2010).
42. P. Skrzypczyk, D. Cavalcanti, arXiv:1803.05199 (2018).
43. A. Acín, S. Pironio, T. Vértesi, P. Wittek, *Phys. Rev. A* **93**, 040102 (2016).
44. F. J. Curchod *et al.*, *Phys. Rev. A* **95**, 020102 (2017).
45. Y. Ding *et al.*, *npj Quantum Inform.* **3**, 25 (2017).
46. R. Fickler *et al.*, *Nat. Commun.* **5**, 4502 (2014).
47. A. Laing, V. Scarani, J. G. Rarity, J. L. O'Brien, *Phys. Rev. A* **82**, 012304 (2010).

ACKNOWLEDGMENTS

We thank A. C. Dada, P. J. Shadbolt, J. Carolan, C. Sparrow, W. McCutcheon, A. A. Gentile, D. A. B. Miller, and Q. He for useful discussions, and W. A. Murray, M. Loutit, and R. Collins for experimental assistance. **Funding:** Supported by the Engineering and Physical Sciences Research Council (EPSRC), European Research Council (ERC), and European Commission, including PICQUE, BBOI, QuChip, QITBOX, UK Quantum Technology Hub for Quantum Communications Technologies (EP/M013472/1) and the Center of Excellence, Denmark SPOC (ref DNRF123), Bristol NSQI; the Spanish MINECO (QIBEQI FIS2016-80773-P and Severo Ochoa SEV-2015-0522), the Fundacio Cellex, the Generalitat de Catalunya (SGR875 and CERCA Program), and the AXA Chair in Quantum Information Science; and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreements 705109, 748549, and 609405. Also supported by the Chinese National Young 1000 Talents Plan (J.W.); the Royal Society through a University Research Fellowship (UHQT) (P.K.); the Villum Fonden via the QMATH Centre of Excellence (no. 10059) and EPSRC grant EP/L021005/1 (L.M.); National Key R&D Program of China grant 2013CB328704 (Q.G.); a Royal Academy of Engineering Chair in Emerging Technologies (J.L.O.); EPSRC fellowship EP/N003470/1 (A.L.); and ERC starter grant ERC-2014-STG 640079 and EPSRC Early Career Fellowship EP/K033085/1 (M.G.T.). **Author contributions:** J.W. designed the experiment; Y.D. designed and fabricated the device. J.W., S.P., Y.D., R.S., and J.W.S. built the setup and carried out the experiment. S.P., P.S., A.S., J.T., R.A., L.M., D.Ba., and D.Bo. performed the theoretical analysis; A.L. contributed the tomography scheme; and Q.G., A.A., K.R., L.K.O., J.L.O., A.L., and M.G.T. managed the project. All authors discussed the results and contributed to the manuscript. **Competing interests:** All authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/285/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S9
References (*48*–*64*)

8 December 2017; accepted 27 February 2018
Published online 8 March 2018
10.1126/science.aar7053

REPORT

NANOPHOTONICS

Probing the ultimate plasmon confinement limits with a van der Waals heterostructure

David Alcaraz Iranzo,[1]* Sébastien Nanot,[1,2]* Eduardo J. C. Dias,[3] Itai Epstein,[1] Cheng Peng,[4] Dmitri K. Efetov,[1,4] Mark B. Lundeberg,[1] Romain Parret,[1] Johann Osmond,[1] Jin-Yong Hong,[4] Jing Kong,[4] Dirk R. Englund,[4] Nuno M. R. Peres,[3] Frank H. L. Koppens[1,5]†

The ability to confine light into tiny spatial dimensions is important for applications such as microscopy, sensing, and nanoscale lasers. Although plasmons offer an appealing avenue to confine light, Landau damping in metals imposes a trade-off between optical field confinement and losses. We show that a graphene-insulator-metal heterostructure can overcome that trade-off, and demonstrate plasmon confinement down to the ultimate limit of the length scale of one atom. This is achieved through far-field excitation of plasmon modes squeezed into an atomically thin hexagonal boron nitride dielectric spacer between graphene and metal rods. A theoretical model that takes into account the nonlocal optical response of both graphene and metal is used to describe the results. These ultraconfined plasmonic modes, addressed with far-field light excitation, enable a route to new regimes of ultrastrong light-matter interactions.

Van der Waals heterostructrures are constructed by vertically stacking atomically thin materials, selected from a rich palette of thousands of materials such as graphene (semimetal), hexagonal boron nitride (h-BN, dielectric), and transition metal dichalcogenides (semiconductors) (*1*). These are key enablers for tailoring electronic, optical, and optoelectronic properties (*2*). The most common heterostructure for two-dimensional (2D) electronics is graphene encapsulated by h-BN, and recently, this system has also emerged as a platform for polaritonics (*3*, *4*), with the capability to strongly confine plasmon polaritons with a relatively long plasmon lifetime exceeding 500 fs at room temperature (*5*). Heterostructures of graphene, h-BN, and metals have revealed so-called propagating acoustic plasmons (*6*–*8*), in which metal screening confines the light in the space between the metal and the graphene, and it slows down the plasmon to a velocity almost as low as $c/300$ (with c the speed of light) (*9*). What is then the ultimate limit on the confinement of propagating or resonant plasmons? For bulk metal–based plasmonic systems such as tapers (*10*), grooves (*11*), metal-insulator-semiconductor (*12*), and metal-insulator-metal (*13*, *14*) waveguides, the confinement of propagating surface plasmon polaritons is limited by Landau damping (*15*). For example, the quality factor of plasmonic Fabry-Pérot modes dropped below one for a confinement below 15 nm (*14*). Further enhancement of optical fields to the nanometer scale is possible in hotspots (*16*–*18*), although these are broadband in character.

We present a graphene-insulator-metal platform that allows us to realize and probe the ultimate physical limits of (out-of-plane) confinement of propagating plasmons down to the ultimate physical boundary of a one-atom-thick layer (here, $\lambda_0/26{,}000$), and without sacrificing damping. We take advantage of the fact that plasmons in two-dimensional materials are fundamentally different from plasmons in bulk metals because the restoring force by the long-range Coulomb interactions, which are essential for the plasmon velocity and confinement, can be controlled by tailoring the external environment. For that reason, the out-of-plane confinement and wavelength compression can be increased strongly without suffering from Landau damping. We use far-field light to couple to these strongly confined plasmons and find a vertical mode length down to 0.3 nm, whereas higher-order Fabry-Pérot resonances reveal that the propagating character of the plasmons is preserved.

The basic device geometry consists of graphene as the plasmonic material, encapsulated by atomically thin dielectric materials (h-BN or Al_2O_3) and covered by a metallic rod array (Fig. 1A) [details on the fabrication processes are provided in (*19*)]. The advantages of the periodic metal-insulator-graphene system are twofold. First, the presence of the metal results in efficient screening of the graphene plasmons, squeezing the so-called screened graphene plasmons (SGPs) into the graphene-metal gap without reducing their lifetime. Although screening and coupling of radiation to plasmons in classical 2D electron gases (2DEGs) was previously achieved by using grating-gate field-effect-transistors in the terahertz range (*7*), a relatively thick barrier layer (~100 nm) prevented the confinement of plasmons to the subnanometer limit, as we report here. Second, metal rods facilitate efficient coupling between far-field light and the strongly confined plasmons, where the width of the rods defines the resonant conditions for the plasmon modes. This approach does not require patterning of the graphene into nanoribbons or nanodiscs such as for previous infrared graphene plasmonic studies with far-field light (*20*, *21*).

The effect of the coupling of far-field light into our devices is represented by the field profiles obtained with finite-difference time-domain (FDTD) simulations (Fig. 1, B and C). The plasmons are launched at the metal edges, and most of the electric field is confined between metal and graphene, with virtually no leakage into the metal. This confinement arises because the metal acts as a nearly perfect conductor that prevents most of the field penetration. Because image charges are induced in the metal, the plasmon mode is analogous to the antisymmetric plasmon mode of two nearby graphene sheets with twice the spacer thickness (*8*), known as acoustic plasmons, carrying larger momentum than for conventional plasmon resonances in graphene ribbons [(*19*), section 5.7]. Within the dielectric gap, the plasmons maintain their propagating character and reflect at the edges of the rod, forming what appears to be a standing wave pattern similar to two coupled Fabry-Pérot resonators (*22*).

Bringing the metal closer will increase the plasmon screening, which slows down the plasmons, as previously observed with scattering-type scanning near-field optical microscopy (*8*, *9*), and also enhances the vertical confinement, as shown for two different dielectric spacer materials and graphene conductivity models (local and nonlocal) (Fig. 1D). The most extreme case is that of a monolayer h-BN spacer, in which in theory the vertical plasmon confinement is below 1 nm. The calculated width of the resonance [(*19*), section 6] does not increase when reducing the spacer thickness s. Therefore, this platform allows us to access the ultimate confinement limits of propagating plasmons in two spatial directions: out-of-plane confinement defined by s and in-plane confinement governed by λ_0/λ_p.

The far-field approach presented above allows probing of this plasmon confinement by use of Fourier transform infrared (FTIR) transmission measurements. Gate-dependent, spectral extinction

[1]Institut de Ciències Fotòniques (ICFO)–The Institute of Photonic Sciences, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain. [2]Laboratoire Charles Coulomb (L2C), Université de Montpellier, CNRS, 34095 Montpellier Cedex, France. [3]Centro de Física and Departamento de Física and QuantaLab, Universidade do Minho, P-4710-057 Braga, Portugal. [4]Department of Electrical Engineering and Computer Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
*These authors contributed equally to this work.
†Corresponding author. Email: frank.koppens@icfo.eu

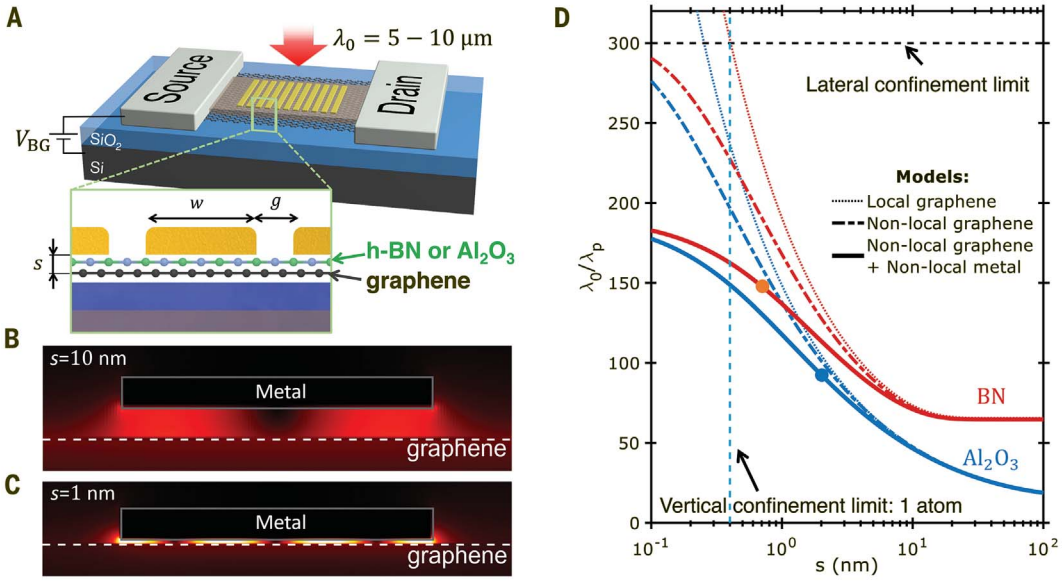


Fig. 1. Device design for probing ultimate plasmon confinement limits. (**A**) Graphene is encapsulated in a dielectric (few-nanometer-thick Al_2O_3 or monolayer h-BN) and covered by an array of gold rods. A gate voltage V_{BG} is applied between Si and graphene in order to control the Fermi energy of the graphene E_F. (Bottom left inset) Schematic cross section of the device. (**B** and **C**) Simulated plasmonic field magnitude profiles for metal-graphene separation *s* of 10 and 1 nm. (**D**) Simulated plasmon wavelength λ_p as a function of metal-graphene spacer *s* for the two materials used in the experiments (λ_0 = 8 μm and E_F = 0.54 eV). The vertical dashed line refers to the fundamental limit: a monolayer h-BN spacer. Colored circles correspond to the two sets of devices discussed in the main text. The dotted lines represent the model where the metal was considered as a perfect conductor in combination with the local graphene conductivity model. The dash-dotted lines represent the nonlocal graphene conductivity model (obtained from the random-phase-approximation), but still metal as a perfect conductor. The solid lines represent the model where nonlocal optical response for both metal and graphene is considered.

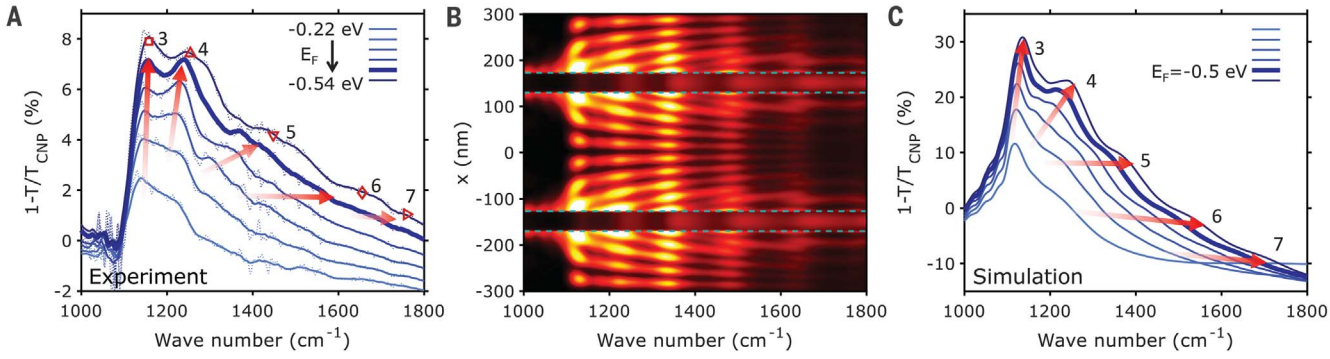


Fig. 2. Resonant excitation of 2-nm confined Fabry-Pérot SGP modes. (**A**) Gate-dependent FTIR extinction spectra referenced to the charge neutrality point (CNP) for metal rod width *w* = 256 nm, gap *g* = 44 nm, and 2-nm Al_2O_3 spacer between graphene and the metal. For increasing E_F, the resonances increase in intensity and blue shift, and high-order modes become visible. (**B**) Simulation of the electric field intensity at the graphene and along *x*, which is along the short axis of the rods, for the same device geometry as in (A), with E_F = 0.5 eV. The model considers the local response of the gold and the local response of graphene. The color scale is linear. (**C**) Simulated extinction for the device geometry as in (A), and same model as for (B).

$(1 - T/T_{CNP})$ curves were measured (Fig. 2A) for a device with continuous graphene (unpatterned), covered with 2 nm Al_2O_3 spacer (*s* = 2 nm) and metallic rods of 256 nm (*w* = 256 nm). These curves are obtained from the transmission curves *T* normalized by the transmission of undoped graphene T_{CNP} [at the charge neutrality point (CNP)] for several gate voltages (hence, Fermi energies E_F) [(*19*), section 2.2] at the same position, with light polarized perpendicular to the rod's long axis. The spectra in Fig. 2A exhibit multiple resonances, which become more pronounced and continuously blue shift with increasing Fermi energies, thus confirming their plasmonic nature. The appearance of multiple peaks demonstrates that the incoming light can resonantly couple to several higher-order plasmonic modes, in contrast to patterned graphene ribbons, as reported previously (*21*). We found up to five visible resonances that can be controlled by the spacer thickness and the metallic array geometry.

We corroborated the phenomenon by examining the simulated electric field intensity profiles at the graphene surface for the same geometry as that of the measured device (Fig. 2B) and the corresponding simulated extinction spectra (Fig. 2C). These FTDT simulations [(*19*), section 4] show a good agreement with the experiment (for *s* = 2 nm) in terms of peak shape and position by using the local optical response of graphene as well as a local metal permittivity model. The resonances can be related to Fabry-Pérot behavior of the propagating SGPs below the metal, and the resonance orders *m* can be approximated by half the number of nodes in the field profile (Fig. 2B). This observation confirms that resonances up to the seventh order are contributing to optical extinction, whereas for patterned graphene (*21*, *23*), only very weak second-order resonances have been experimentally reported. We attribute the efficient launching of these higher-order resonances to the strong dipole modes at the metal edges [(*19*), section 9], which provide the required momentum to scatter light into graphene plasmons.

To probe the physical limits of SGP confinement, we studied devices with a metal-graphene

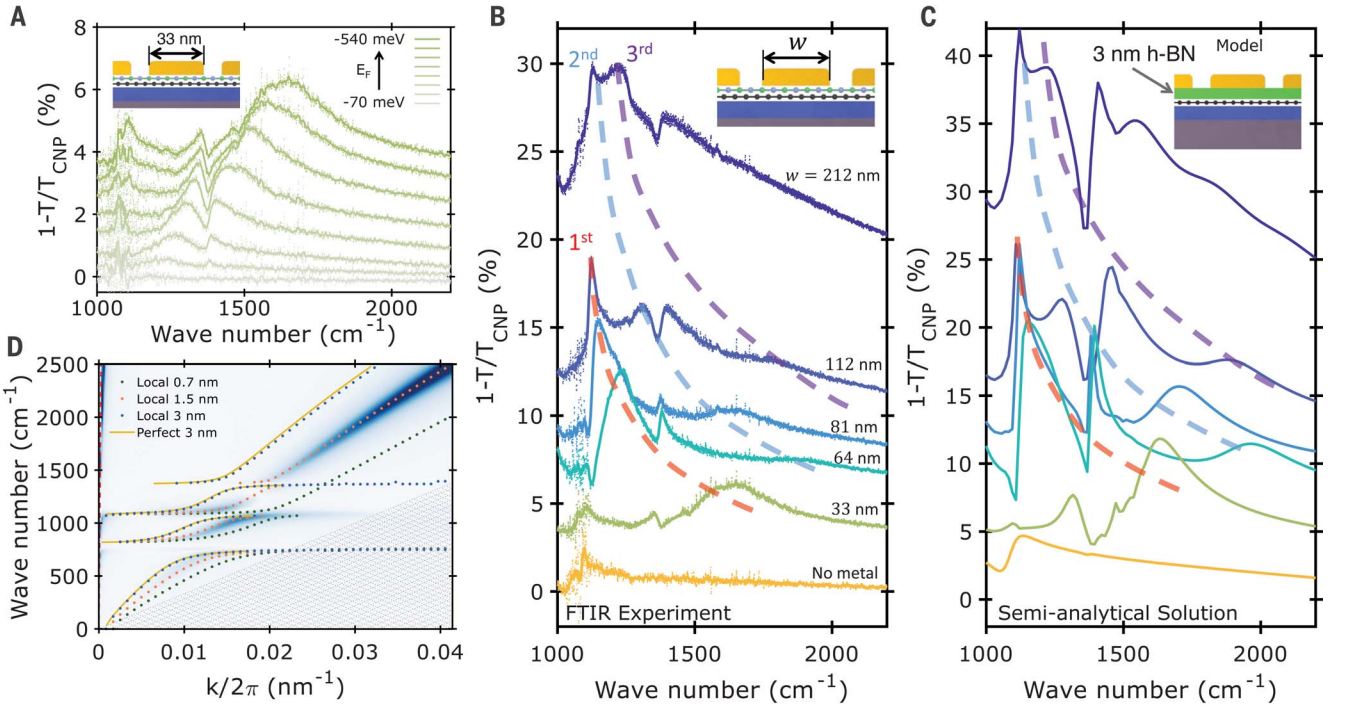


Fig. 3. Single mode and Fabry-Pérot SGP modes confined to a monolayer h-BN. (**A**) Extinction spectra for E_F ranging from 70 to 540 meV, and fixed $w \approx 33$ nm, gap $g \approx 37$ nm. (**B**) Extinction spectra for w ranging from 33 to 212 nm, and fixed $g = 38 \pm 4$ nm, $E_F = 540$ meV. Dashed lines are guides to the eye showing the evolution of each resonance with w. (Inset) Monolayer device schematic for data in (A) and (B). (**C**) Simulated extinction spectra where the nonlocal metal effects are accounted for by modeling a perfectly conducting metal but an effective thicker 3-nm h-BN spacer. (Inset) Model schematic. (**D**) Plasmon dispersion relation for a (continuous) SiO_2/graphene/h-BN/metal/air heterostructure. Dotted curves correspond to local metal response [with nonzero loss; (*19*), section 3] and are plotted for h-BN thickness of 0.7 nm (green), 1.5 nm (orange), and 3 nm (blue). The solid yellow curve corresponds to h-BN thickness of 3 nm (yellow) and modeling the metal as perfectly conducting. The blue color gradient represents the loss function of the heterostructure for 0.7-nm-thick h-BN with nonlocal metal (titanium) and nonlocal graphene response. This illustrates that accounting for nonlocality comes down to adding an extra spacer thickness of ~2 nm to a model that considers only the local metal response.

spacer with only one monolayer of chemical vapor deposition–grown h-BN of thickness ~0.7 nm (*24*, *25*). The extinction spectra were obtained for various E_F of a device with $w = 33$ nm and $g = 37$ nm (Fig. 3A). Even in this case, the SGP resonances were still visible, with high extinction values. A single plasmon peak was clearly visible and blue shifted for increasing E_F while hybridizing and anticrossing with SiO_2 and h-BN phonons. This shift with Fermi energy confirms that the extinction resonance is due to graphene plasmons.

The propagating character of the SGP modes can be assessed by further increasing w, which allows us to probe higher-order Fabry-Pérot resonances. We will show that this measurement is equivalent to changing the plasmonic cavity width w, which defines the resonant conditions. We studied the extinction of several devices with increasing w and similar gap around $g = 40$ nm and doping of $E_F = 0.54$ eV (Fig. 3B). Additional doping dependences of these devices can be found in fig. S5. The first-order resonance (Fig. 3B, red dashed line) displays hybridization with the optical phonons of its dielectric environment (as observed for plasmons in graphene nanoribbons) (*26*). We observed that when increasing w, the first-order resonance shifts to lower energies and hybridizes not only with the h-BN phonons but also with the SiO_2 phonons. More resonances appear with increasing w. For example, for $w = 64$ nm, a second-order resonance appears at 1900 cm^{-1} (Fig. 3B, blue dashed line), which also shifts for larger w and starts hybridizing when increasing w. For the largest w, the third-order (Fig. 3B, purple dashed line) and even fourth-order resonances appear for our measurement range.

The same trends are seen in the calculated spectra (Fig. 3C), obtained from a semi-analytical approach, which consists of a Fourier decomposition of the fields (for transverse magnetic modes) in each dielectric region of the system [(*19*), section 5]. In addition, for these very tightly confined optical fields, one must also take into account that the nonlocal optical response of both the metal and graphene can have appreciable effects. First of all, the in-plane momentum of the graphene plasmon is strongly enhanced by the presence of the metal and approaches $\omega/v_{F,graphene}$, where the momentum dependence of the graphene optical conductivity (nonlocal corrections to the conductivity) increases (*9*). These graphene nonlocal corrections are modeled within the framework of the random-phase-approximation (RPA) [(*19*), section 5.6]. Second, because of additional strong vertical field confinement on the order of 1 nm, the components of out-of-plane wave vectors approach $\omega/v_{F,metal}$, which is the Fermi momentum of the charge carriers in the metal (*16*). This trend can lead to additional nonlocal effects in the metal, resulting in field penetration into the metal. It is therefore relevant to quantify these effects in order to determine fundamental limits of the vertical field confinement of the propagating plasmon.

Before we discuss a more rigorous treatment of the metal nonlocal effects, we provide a qualitative picture (*27*), in which the nonlocal metal permittivity (NMP) was modeled as a dielectric shell surrounding a local metal permittivity (LMP) bulk material. Applying this model to our system, we used a perfect conductor model (with zero damping) for the bulk metal, and the thickness s of a uniform dielectric spacer was used as a fitting parameter. Simulations for this case, with an effective dielectric thickness of 3 nm (Fig. 3C), show good agreement with the experiment (Fig. 3B). Even though this result conflicts with the estimated spacer thickness (0.7 nm), the nonlocal effect can be understood by recognizing that the

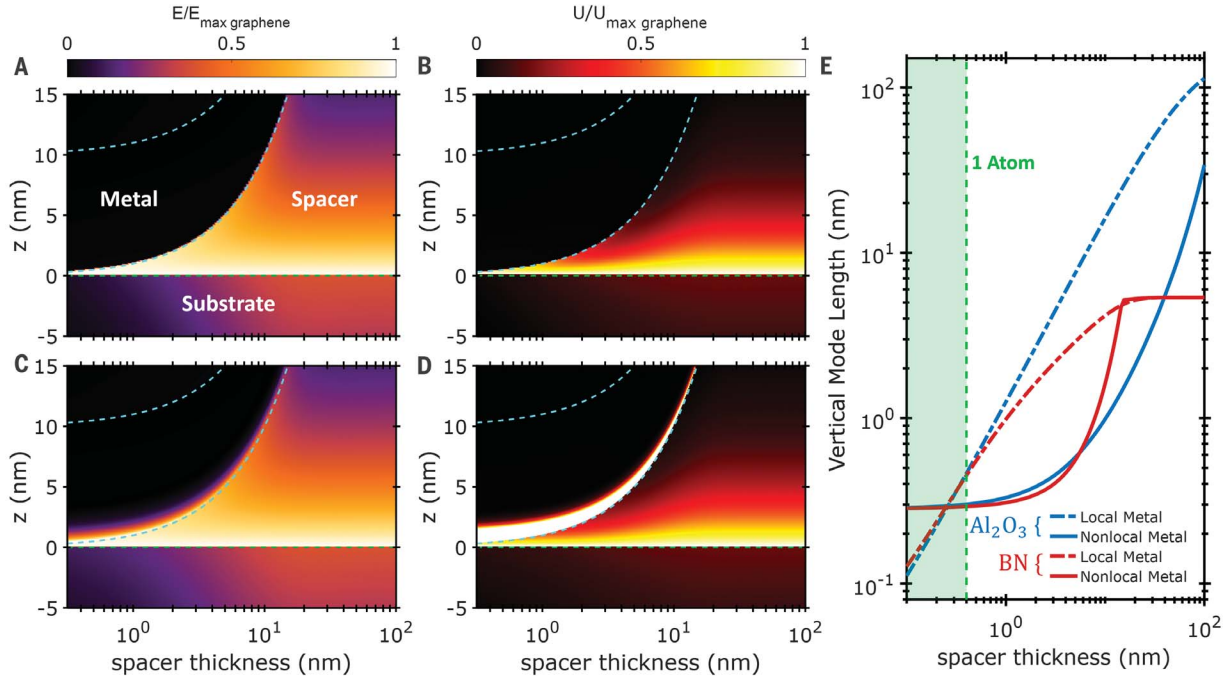


Fig. 4. Energy density and field confinement. (**A** and **C**) Electric field magnitude distribution of the plasmons associated to a continuous heterostructure of air/Ti/h-BN/graphene/SiO_2 as a function of h-BN thickness for (A) LMP model and (C) NMP model. The top and bottom metal limits are depicted by blue dashed lines, and graphene is located at $z = 0$. Normalization by the maximum electric field strictly above graphene shows the confinement and screening effects. (**B** and **D**) Same as (A) and (C), respectively, but for energy density. (**E**) Vertical field confinement for both types of dielectrics as a function of the spacer thickness for local (dash-dotted lines) and nonlocal (solid lines) metal permittivity.

electromagnetic field does penetrate more into the metal for smaller s (Fig. 4 and fig. S13), increasing the effective s.

The excited plasmon modes associated with the extinction peaks in Figs. 2 and 3 can be seen as a combination of SGPs under the metal and unscreened graphene plasmons in the gap region, with their relative contribution depending on the geometry. Nevertheless, the calculated field profiles in Fig. 1, B and C, and the scaling of the plasmon resonance peak energy with w clearly show that the electric field (associated with the plasmon modes) is mostly confined between the metal and the graphene. This confinement is also consistent with the observations that the plasmon resonances shift most substantially when varying w compared with a change of the gap g [(*19*), section 3] and that the system can be seen as a plasmonic crystal (analogous to photonic crystals) (*22*).

To further explore that a thicker dielectric in the LMP model results in nonlocal effects in the metal, the dispersion relation of the SGP modes is shown in Fig. 3D. The dispersion relation for a fully NMP model is offset to higher energies as compared with a LMP model, accounting for the proper h-BN thickness (0.7 nm) (Fig. 3D, green dotted line). Accounting for a thicker dielectric spacer used to fit our data ($s = 1.5$ nm), the dispersion curves shift down and overlap with the NMP model. Qualitatively, this is in agreement with our findings. An additional shift is expected for a periodic structure instead of a continuous one, as calculated in Fig. 3D, because of a small coupling between the modes below the rods.

We can now evaluate the ultimate limit on the plasmonic vertical field and mode volume confinement. We have calculated the electric field intensity and energy density distribution for different spacer thicknesses (Fig. 4) and considered both LMP and NMP models. We used the definitions in section 5 of (*19*), taking into account dispersion effects. Inspecting first the electric field magnitude obtained from the LMP model and normalized by the maximum of the graphene plasmon (Fig. 4, A and B), we found that most of it is confined between the graphene and the metal and that the penetration of the field inside the metal is negligible. On the other hand, considering metal nonlocal effects (NMP model) offers a more complete picture of the physics of the electrons that accumulate at the metal surface in order to screen the electromagnetic field (*18*). When the out-of-plane wave vector is increased for thinner spacers, the electrons start suffering from Pauli and electrostatic repulsion. This effect results in a saturation of the electron density and leads to field penetration into the metal as the metal screening capability is reduced (*28*). The penetration of the field into the metal becomes considerable for s below 3 nm, although the field remains maximum in the spacer region.

This nonlocal field penetration limits the vertical mode length, which is defined by the ratio of the energy density integrated over the out-of-plane coordinate z to the maximum of the field intensity in the region of the spacer (*29*, *30*)

$$L = \frac{\int u_E(z)dz}{\max u_E(z)} \quad (1)$$

The energy density distribution (Fig. 4D) peaks at the metal surface for a NMP model as a consequence of charge accumulation and the combination of high field and permittivity values, which is in contrast to the LMP (Fig. 4C). Below a certain spacer thickness, the energy density at the metal surface becomes larger than the energy density in the dielectric region near graphene. At this point, the transition of the maximum energy density from graphene to the metal dominates the vertical mode length (Fig. 4E). The out-of-plane confinement is calculated from Eq. 1 and the energy density distribution that is shown in Fig. 4E. Whereas for LMP the vertical mode length can be made arbitrarily small, the nonlocal metal behavior limits the vertical mode length to ~0.3 nm. This corresponds to a plasmonic mode confinement down to the atomic scale, as confirmed with our experiments.

These results raise the question: Why is the plasmonic mode for $s < 1$ nm, which substantially penetrates the metal (because of nonlocal response), not overdamped through Landau damping? For metals, the strongest confinement normal to the surface is limited by direct excitation of electron-hole pairs (Landau damping), which is accounted for by the nonlocal

response function (*15*). A quantitative analysis of the metallic nonlocal corrections to the dynamic response and damping through the Feibelman parameters (*18*, *31*) revealed that these effects are much stronger for ω approaching ω_p. For $\omega << \omega_p$, the imaginary part of the Feibelman parameters approaches zero as the phase space for electon-hole excitations decreases with ω. However, the real part of the Feibleman parameters does remain finite for $\omega << \omega_p$, which means that nonlocal effects persist for smaller ω, but damping contributions vanish. Thus, for our experimental conditions (with $\omega << \omega_p$), the metal nonlocal effects are relevant because the electrons in the metal cannot perfectly screen the field, but the additional damping from electronic excitations in the metal is weak.

In terms of ultimate mode volume limit, a reduction of vertical confinement reduces also the lateral propagating plasmon wavelength (Fig. 1D). Although our experiment discloses the fundamental limits of the vertical mode length of $\lambda_0/26{,}000$, it also allows us to estimate the complete mode volume confinement V_p with respect to the volume associated with the extent of the free photons $V_0 = \lambda_0^3$. If instead of an array (as in this work), a single resonant structure (for example, a metal disc on graphene, with a monolayer h-BN spacer) was built with our approach, a vertical mode length down to 0.3 nm and plasmon wavelenth $\lambda_p = \lambda_0/170$ (Fig. 1D) would be attainable, and potentially even smaller λ_p is possible with twisted bilayer graphene (*32*). Using these values (for λ_0 = 8 μm), which include the full nonlocal response for both graphene and metal, this would correspond to a mode volume of $V_p = 664\ nm^3$ and mode volume ratio $V_0/V_p \approx 10^{-9}$, which is at least two orders of magnitude smaller than reported so far with graphene nanoribbons (*20*) and sufficiently small to explore new regimes of light-matter interactions, such as the ultrastrong coupling regime.

Our results show that 2D-material heterostructures can be considered as a powerful toolbox for nanophotonics with vertical subnanometer precision. In general, patterned metals in the vicinity of van der Waals heterostructures do strongly influence the Coulomb interactions, collective excitations, and even the electronic band structure, which proves itself as an exciting platform for designing novel quantum and topological phenomena. The metallic structure can also be used as a nearby efficient gate (*7*) and provides a route to applications such as molecular sensing with even higher resolution, to enhance nonlinear effects, or to design photodetectors with plasmon-enhanced sensitivity and tunability.

REFERENCES AND NOTES

1. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
2. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, *Science* **353**, aac9439 (2016).
3. T. Low *et al*., *Nat. Mater.* **16**, 182–194 (2017).
4. D. N. Basov, M. M. Fogler, F. J. García de Abajo, *Science* **354**, aag1992 (2016).
5. A. Woessner *et al*., *Nat. Mater.* **14**, 421–425 (2015).
6. A. Principi, R. Asgari, M. Polini, *Solid State Commun.* **151**, 1627–1630 (2011).
7. V. V. Popov, *J. Infrared Millim. Terahertz Waves* **32**, 1178–1191 (2011).
8. P. Alonso-González *et al*., *Nat. Nanotechnol.* **12**, 31–35 (2017).
9. M. B. Lundeberg *et al*., *Science* **357**, 187–191 (2017).
10. M. I. Stockman, *Phys. Rev. Lett.* **93**, 137404 (2004).
11. S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, T. W. Ebbesen, *Phys. Rev. Lett.* **95**, 046802 (2005).
12. R. F. Oulton, V. J. Sorger, D. A. Genov, D. F. P. Pile, X. Zhang, *Nat. Photonics* **2**, 496–500 (2008).
13. J. A. Dionne, L. A. Sweatlock, H. A. Atwater, A. Polman, *Phys. Rev. B* **73**, 035407 (2006).
14. H. T. Miyazaki, Y. Kurokawa, *Phys. Rev. Lett.* **96**, 097401 (2006).
15. J. B. Khurgin, G. Sun, in *Quantum Plasmonics*, S. Bozhevolnyi, L. Martin-Moreno, F. Garcia-Vidal, Eds. (Springer, 2016), ch. 13, pp. 303–322.
16. C. Ciracì *et al*., *Science* **337**, 1072–1074 (2012).
17. F. Benz *et al*., *Science* **354**, 726–729 (2016).
18. T. V. Teperik, P. Nordlander, J. Aizpurua, A. G. Borisov, *Phys. Rev. Lett.* **110**, 263901 (2013).
19. Materials and methods are available as supplementary materials.
20. V. W. Brar, M. S. Jang, M. Sherrott, J. J. Lopez, H. A. Atwater, *Nano Lett.* **13**, 2541–2547 (2013).
21. H. Yan *et al*., *Nat. Photonics* **7**, 394–399 (2013).
22. I.-T. Lin, C. Fan, J.-M. Liu, *IEEE J. Sel. Top. Quantum Electron.* **23**, 144–147 (2017).
23. Z. Fang *et al*., *ACS Nano* **7**, 2388–2395 (2013).
24. K. K. Kim *et al*., *Nano Lett.* **12**, 161–166 (2012).
25. J. H. Park *et al*., *ACS Nano* **8**, 8520–8528 (2014).
26. V. W. Brar *et al*., *Nano Lett.* **14**, 3876–3880 (2014).
27. Y. Luo, A. I. Fernandez-Dominguez, A. Wiener, S. A. Maier, J. B. Pendry, *Phys. Rev. Lett.* **111**, 1 (2013).
28. C. David, F. J. García de Abajo, *ACS Nano* **8**, 9558–9566 (2014).
29. R. Ruppin, *Phys. Lett. A* **299**, 309–312 (2002).
30. S. A. Maier, *Opt. Express* **14**, 1957–1964 (2006).
31. T. Christensen, W. Yan, A. P. Jauho, M. Soljačić, N. A. Mortensen, *Phys. Rev. Lett.* **118**, 157402 (2017).
32. T. Stauber, P. San-Jose, L. Brey, *New J. Phys.* **15**, 113050 (2013).

ACKNOWLEDGMENTS

The authors thank G. Konstantatos and V. Pruneri for the intensive use of their respective FTIRs and are grateful for very insightful discussions with M. Polini, T. Christensen, A. Mortenson, and J. Aizpurua on nonlocal effects and with A. Woessner on simulation and modeling of graphene acoustic plasmons modes. **Funding:** We acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (grant SEV-2015-0522), and support by Fundacio Cellex Barcelona, the Mineco grants Ramon y Cajal (grant RYC-2012-12281), Plan Nacional (grants FIS2013-47161-P and FIS2014-59639-JIN), and the government of Catalonia through the SGR grant (2014-SGR-1535). Furthermore, the research leading to these results has received funding from the European Union H2020 Programme under grant agreement 604391 Graphene Flagship, the European Research Council starting grant (307806, CarbonLight), and project GRASP (FP7-ICT-2013-613024-GRASP). D.A.I. acknowledges the FPI grant BES-2014-068504. N.M.R.P. and E.J.C.D acknowledge support from the Portuguese Foundation for Science and Technology (FCT) in the framework of the Strategic Financing UID/FIS/04650/2013. This work was supported in part by the Center for Excitonics, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under award DE-SC0001088, and the Army Research Office (grant 16112776). J.-Y.H. and J.K. acknowledge support from the U.S. Air Force Office of Scientific Research Foldable and Adaptive 2D Electronics Multidisciplinary University Research Initiative, grant FA9550-15-1-0514. **Author contributions:** F.H.L.K, D.A.I., and S.N. conceived the idea; E.J.C.D and N.M.R.P. developed the analytical model; S.N., C.P., J.O., D.K.E., and D.A.I. fabricated the devices; S.N., R.P., and D.A.I. performed measurements; D.A.I., M.B.L., I.E., and S.N. performed data analysis; J.-Y.H. and J.K. provided h-BN; D.A.I., S.N., E.J.C.D., N.M.R.P, I.E., D.K.E., and F.H.L.K wrote the manuscript; and D.R.E. and F.H.L.K. supervised the project. **Competing interests:** None of the authors have 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.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/291/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S17
References (*33*–*47*)

25 December 2017; accepted 26 February 2018
10.1126/science.aar8438

MATERIALS SCIENCE

Capillarity-induced folds fuel extreme shape changes in thin wicked membranes

Paul Grandgeorge,[1] Natacha Krins,[2] Aurélie Hourlier-Fargette,[1,3] Christel Laberty-Robert,[2] Sébastien Neukirch,[1] Arnaud Antkowiak[1,4]*

Soft deformable materials are needed for applications such as stretchable electronics, smart textiles, or soft biomedical devices. However, the design of a durable, cost-effective, or biologically compatible version of such a material remains challenging. Living animal cells routinely cope with extreme deformations by unfolding preformed membrane reservoirs available in the form of microvilli or membrane folds. We synthetically mimicked this behavior by creating nanofibrous liquid-infused tissues that spontaneously form similar reservoirs through capillarity-induced folding. By understanding the physics of membrane buckling within the liquid film, we developed proof-of-concept conformable chemical surface treatments and stretchable basic electronic circuits.

Geometry is behind the curious mechanical behavior of the capture silk spun by ecribellate spiders: Whether stretched or compressed, this fiber remains straight while seemingly adjusting its length, as if telescopic. In reality, the pulling force caused by surface tension allows coiling, spooling, and packing of excess fiber within the glue droplets decorating the thread. These fiber reservoirs can be recruited on demand and give the thread an apparent extreme stretchability of +10,000% (*1*). Another example are cells that display a particular ability to cope with stretch. Macrophages extend their surface area by a factor of 5 to engulf large microbes or cellular debris (*2*), patroling T lymphocytes stretch by 40% to squeeze into the microvasculature (*3*), hundreds of micrometer-sized neuronal projections extrude from 10-μm-wide neurons (*4*, *5*), and the osmotic swelling of fibroblasts leads to a 70% increase in area (*6*). Such an extreme deformability is all the more spectacular given that the lytic stretching level at which the plasma membrane ruptures is about 4% (*4*, *7*). Why do these cells not burst under stress? Cells store excess membrane in the form of folds and microvilli (*8*, *9*) that can be recruited and deployed on demand. These local geometrical ruffles do not alter the global shape of the cell, because cellular tension is preserved with the pulling action of the underlying cortical actin layer (*10*). The considerable deformations that biological materials undergo could inspire a new generation of synthetic stretchable materials, which are in demand for emerging technologies such as stretchable electronics (*11*), flexible batteries (*12*), smart textiles (*13*), biomedical devices, tissue engineering, and soft robotics (*14*, *15*).

Our technique for making synthetic fabrics with high stretchability takes advantage of spontaneously formed membrane folds and ruffles. Figure 1 illustrates the key steps in designing such an extensible tissue. We first manufacture, using an electrospinning technique, a light and free-standing nonwoven fabric made of poly(vinylidene fluoride-co-hexafluoropropylene) (PVDF-HFP; Fig. 1) (*16*). Without further treatment, this membrane would show early signs of damage above a few percent of extension and would rupture at 30% area extension. We therefore infuse the fibrous membrane with a wetting liquid, so as to mimic the pulling action of the cortical actin layer with surface tension. The membrane becomes tight while seemingly adjusting its surface, storing any excess membrane into a venation network (visible in Fig. 1 and movie S1), thus remaining globally flat. These veins are made of membrane ruffles and furrows that can be unfolded as required, fueling any imposed shape change to the membrane. By buffering excess membrane and mediating stretching, these veins play the same role as the membrane reservoirs in living cells, and we therefore refer to these structures as reservoirs. The unfolding process is reversible; the membrane reservoirs, after being smoothed out upon extension, spontaneously reform upon compression.

To elucidate the mechanics of membrane reservoir formation, we investigated the inner conformation of the membrane within the liquid film. Figure 2 presents microscopic views revealing that the flat membrane portions are actually lightly wrinkled with a wavelength λ, whereas

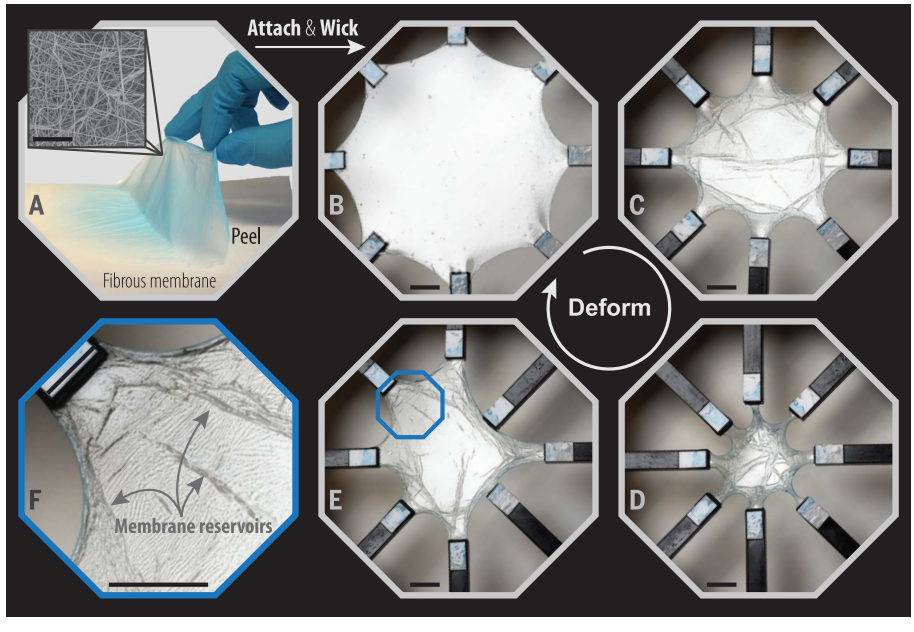


Fig. 1. Design of the ultrastretchable wicked membrane. (**A**) A thin fibrous membrane (electrospun PVDF-HFP membrane dyed blue) and the corresponding scanning electron microscope micrograph. The membrane has a typical thickness of a few micrometers; the fibers composing it have diameters around 300 nm. Scale bar, 50 μm. (**B**) The membrane is attached to eight translational supports and wicked with 3 cSt silicone oil. (**C** to **E**) The eight supports are brought together, but the wicked membrane remains globally flat by locally storing excess membrane in apparent veins. (**F**) A closer view of these veins, or membrane reservoirs, which buffer the imposed deformations on the wicked membrane. Scale bars in (B) to (F), 1 cm.

[1]Sorbonne Université, CNRS, Institut Jean le Rond ∂'Alembert, F-75005 Paris, France. [2]Sorbonne Université, CNRS, Laboratoire de Chimie de la Matière Condensée de Paris, F-75005 Paris, France. [3]PSL Research University, CNRS, École Normale Supérieure, Département de Physique, F-75005 Paris, France. [4]Saint-Gobain, CNRS, Surface du Verre et Interfaces, F-93303 Aubervilliers, France.
***Corresponding author. Email: arnaud.antkowiak@upmc.fr**

the veins are made up of a stack of folds and both the lightly wrinkled and stacked regions are sheathed between slightly rippled liquid interfaces (movie S2). We first focus on the emergence of the lightly wrinkled pattern (Fig. 2B, panel b). Wrinkling is a trademark of thin elastic sheets, and it develops spontaneously in a variety of contexts: pinched skin, shriveling fruits (*17*), brain sulci (*18*), hanging curtains (*19*), or, more generally, thin sheets under tension (*20*). This elastic instability occurs whenever a compressed slender structure is bound to a substrate resisting deformation. The emerging wavelength λ of this particular form of buckling therefore results from a trade-off between the deformation of the membrane and that of the substrate in order to minimize global energy. Here, the substrate role is played by the liquid film interfaces, which can be seen as soft capillary walls confining the membrane. Experiments revealed that the wavelength λ is neither particularly sensitive to the interfacial tension γ, nor to the fibrous membrane thickness t_0, but scales linearly with the liquid film thickness h, which is measured by means of colorimetry (Fig. 2D and figs. S1 and S2) (*16*).

We developed a simple model in which a periodic sinusoidal membrane of bending stiffness B per unit depth interacts with a liquid film, exposing two free interfaces of surface tension γ (Fig. 2C and figs. S8 and S9) (*16*). This model ignores the effects of gravity on wavelength, but gravity only plays an important role in the orientation of the wrinkling pattern (*16*) (figs. S16 to S20). Under the constraint of constant liquid film volume and imposed compression ε, we minimize the total energy of the system, consisting of the membrane elastic energy $E_{el} = \frac{1}{2}B\int\kappa^2\,ds$ and the surface energy $E_\gamma = 2\gamma S$, where κ and S respectively denote the local membrane curvature and the exposed surface of the liquid film per unit depth (*16*). The model reveals h/L_{ec} as a relevant parameter of the system, where $L_{ec} = (B/\gamma)^{1/2}$ is the elastocapillary length (*21*). The limit $h/L_{ec} << 1$ typically corresponds to everyday-life soaked fibrous membranes (e.g., wet paper or cloth) that sag or buckle globally when compressed, irrespective of any surface tension effects (Fig. 2C). Conversely, our experiment is characterized by values of $h/L_{ec} >> 1$ for which the microstructure differs markedly from that of a common wet fabric: Interface energies can no longer be neglected, and the membrane now buckles under capillary confinement (Fig. 2C). In this regime, the ratio of surface energy to elastic energy scales as $E_\gamma/E_{el} \sim (h/L_{ec})^2 >> 1$; that is, any deformation of the liquid surface introduces a strong energetic penalty, which means that in-film wrinkling is a low-energy configuration. This phenomenon is therefore reminiscent of buckling under rigid confinement, for which the wavelength λ also scales linearly with the confinement gap h for a given compression ε (*22*), and this behavior is recovered by our model (Fig. 2D, inset, and fig. S10). In contrast to classic buckling under confinement (*22*), the experimentally measured wavelengths λ prove to be insensitive to compression. This

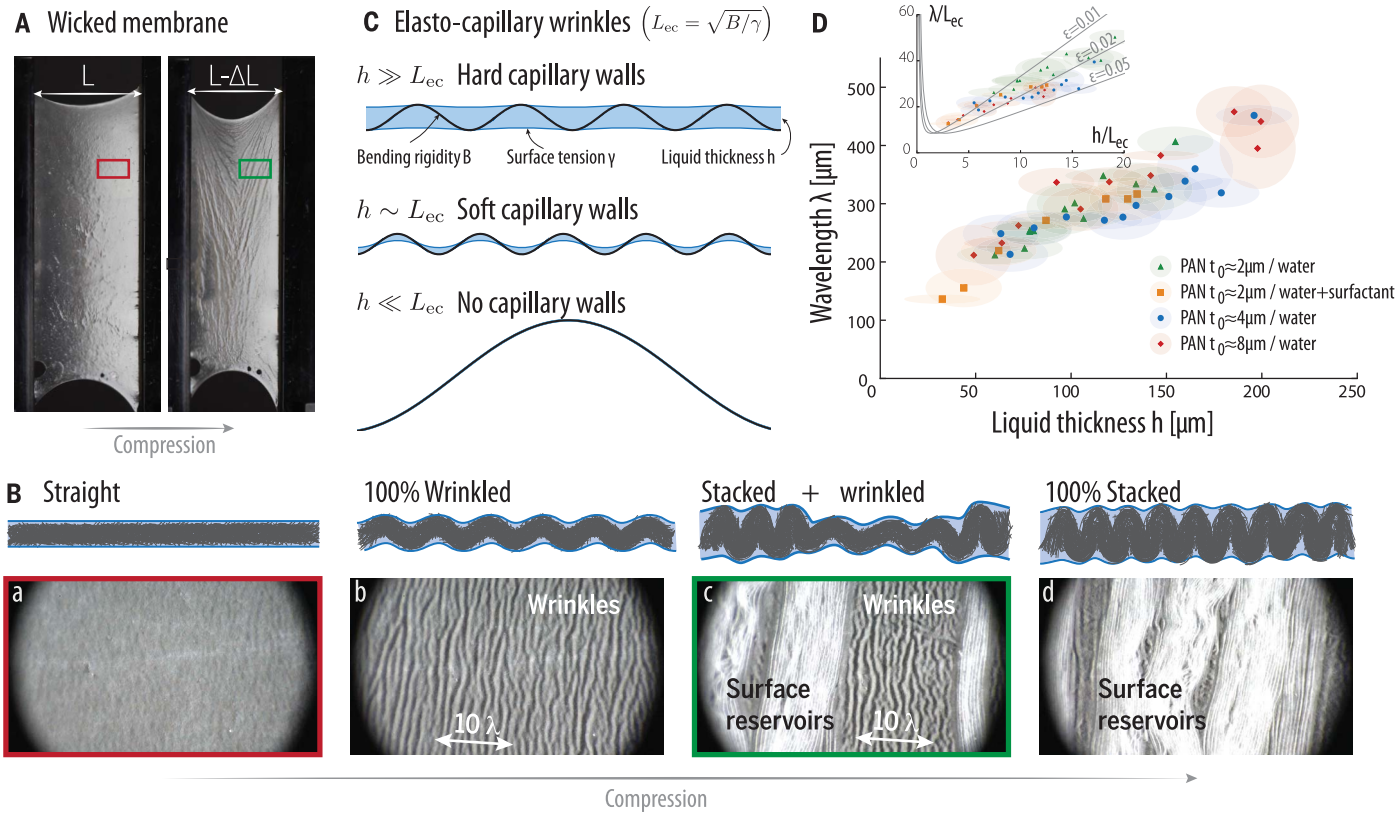


Fig. 2. Mechanics of the wicked membrane: Capillary-driven wrinkling and stacking. (**A**) Polyacrylonitrile (PAN) membrane wicked with water and attached to two straight mobile edges. Upon small compression, the wicked membrane exhibits a clear wrinkling pattern (L = 4 cm). (**B**) Close-up on the wicked membrane throughout compression. At its extended state, the wicked membrane is smooth (a) and a small compression generates a wrinkled surface of wavelength λ (b). Further compression then leads to a two-phase texture (c). One phase corresponds to the wrinkled texture (wavelength λ) and the other to a closely packed stack of folds, which gradually expands throughout compression. The whole membrane is packed in this stack of folds at the end of compression (d). Scale: 10λ = 3.1 mm. (**C**) Physical interpretation of the early wrinkling of the membrane inside the liquid film. Depending on the ratio of liquid film thickness to elasto-capillary length, h/L_{ec}, three different buckling scenarios emerge. (**D**) Experimental wavelength λ of the wrinkles observed at an early compression stage of the wicked PAN membrane as a function of the liquid film thickness h for different membrane thicknesses t_0 and wicking liquids. The inset provides the data normalized by the elasto-capillary length L_{ec} (*16*).

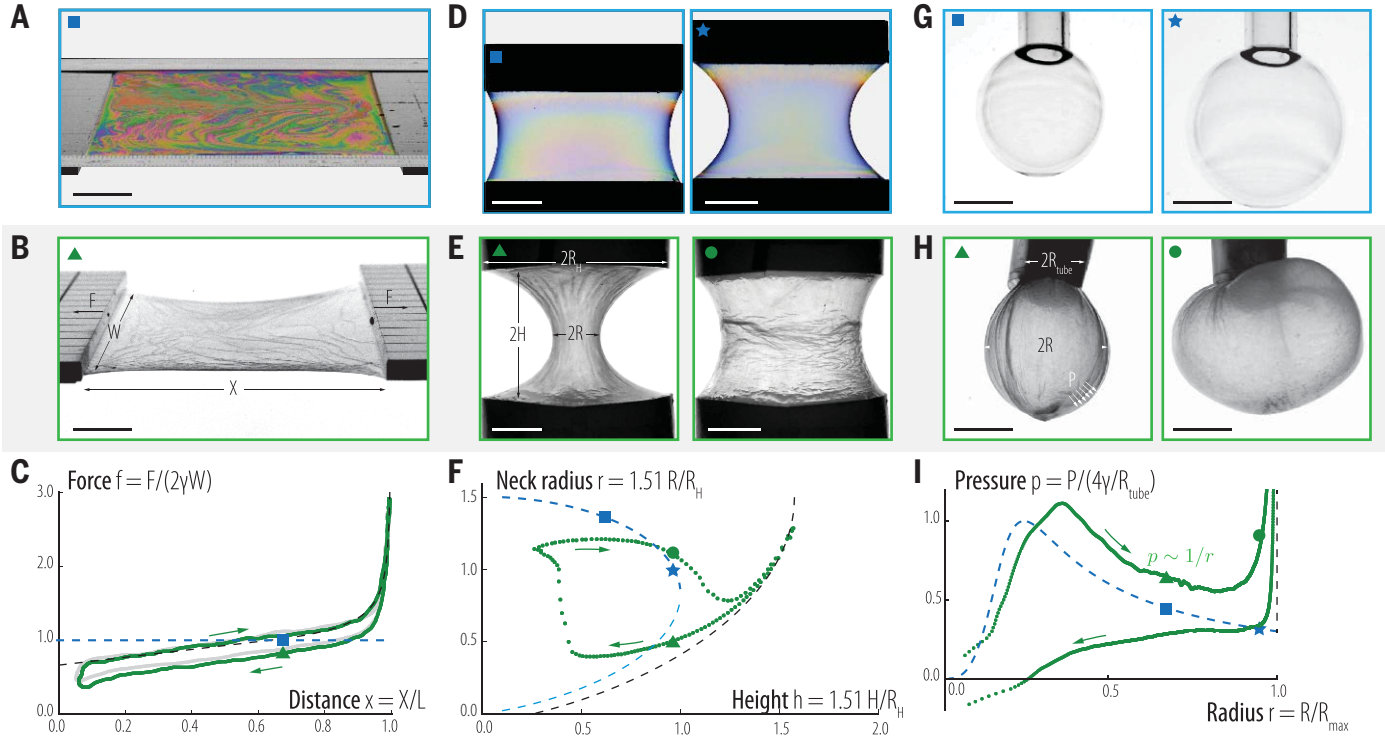


Fig. 3. Forms and forces for capillary-folded wicked membranes. (**A**) Soap liquid film on a frame. (**B**) Planar wicked membrane attached to two mobile supports. (**C**) The green curve corresponds to the force measurements during the first compression/extension cycle of a planar PVDF-HFP wicked membrane whereas the gray curve was obtained after imposing 100,000 compression/extension cycles on the membrane. The blue dashed line shows the force prediction on a planar soap film on a rigid frame; the black dashed line corresponds to the theoretical prediction with inextensibility constraint (*16*). (**D**) Soap liquid catenoid between two parallel circular rings. (**E**) Two states of the catenoid shape adopted by a wicked membrane attached to two parallel circular rings. (**F**) Neck radius of the catenoid versus distance between the two rings. The green points represent the experimental observation for a wicked membrane. The blue dashed line shows the soap liquid solution for the catenoid and the black dashed line shows the solution for a catenoid with inextensibility constraint (*16*). (**G**) Soap bubble. (**H**) Bubble generated by inflating a wicked membrane at two different inflating stages. (**I**) Pressure versus radius diagram. Here, the radius of a bubble is defined as $R = [(3/4\pi)V]^{1/3}$, where V is the volume of injected air. The blue dashed line represents the theoretical pressure for a soap bubble being inflated through a tube of radius R_{tube} (Laplace's law). The green points correspond to the pressure measurements of the wicked membrane. Scale bars, 1 cm.

behavior, not captured by the model, coincides with the emergence of a second phase in mechanical equilibrium with the first wrinkled phase. This second phase, corresponding to the membrane reservoir, consists of tightly stacked folds and is therefore characterized by a high membrane storage capacity (Fig. 2B). The coexistence between these phases allows continuous transfer of material from one phase to another and ensures the effectiveness of the wicked membrane as a stretchable material.

We next subjected wicked membranes to three different elementary deformations corresponding to the stretching of planar, cylindrically shaped, and spherically shaped membranes (movies S3 to S5, respectively). The equilibrium shape adopted by the wicked membrane in each configuration strongly resembles that of a liquid film under the same conditions: planar film (Fig. 3B), catenoid (Fig. 3E and fig. S5), and bubble (Fig. 3H and fig. S6). Once again, this behavior is made apparent by realizing that in the limit $h/L_{ec} >> 1$, the energy of the wicked membrane is dominated by its capillary contribution; the equilibrium shapes therefore essentially correspond to minimal surfaces. Although they differ strongly in longevity, fabrication methods, and internal structure, the wicked membranes and liquid films therefore present interesting similarities. Upon closer inspection, however, the shapes of the membranes appear to differ from their liquid counterparts in some respects. For example, a planar wicked membrane attached on only two straight edges adopts a stable shape (Fig. 3, A to C), whereas a liquid film in the same configuration would burst. To understand this stabilization mechanism for membranes, we must recognize that some regions of the membrane may undergo stretching up to a point where the membrane reservoirs are fully exhausted. For the thin fibers composing the membrane, pure stretching deformations represent a far higher energetical cost relative to bending deformations (*23*), and as a first approximation, this sharp energetical penalty can be seen as an inextensibility constraint. The shapes adopted by the planar configuration can therefore be captured with a surface area minimization under isoperimetric constraint (*16*) (figs. S11 to S13). This mixed liquid-solid behavior allows stabilization of the catenoid shape beyond its classic point of bursting to unveil new equilibria (*16*) (Fig. 3, D to F, and figs. S14 and S15), and is also responsible for considerable deviations from Laplace's law in the bubble configuration (Fig. 3, G to I). Such a hybrid mechanical behavior is again reminiscent of the response of cellular membranes, and indeed, whether for lymphocytes, fibroblasts (*3*, *6*), or wicked membranes (Fig. 3), the mechanical response switches from liquid-like to solid-like once all the membrane reservoirs have been smoothed out.

The peculiar behavior of our wicked membrane stems from its compound nature: Capillarity-induced folds allow it to undergo ample shape changes while remaining taut, while its solid underlying matrix provides mechanical robustness. Geometrical reorganizations at the microstructural level (reservoir folding or unfolding) are key in the mechanics of the wicked membrane, and in particular they prevent any notable

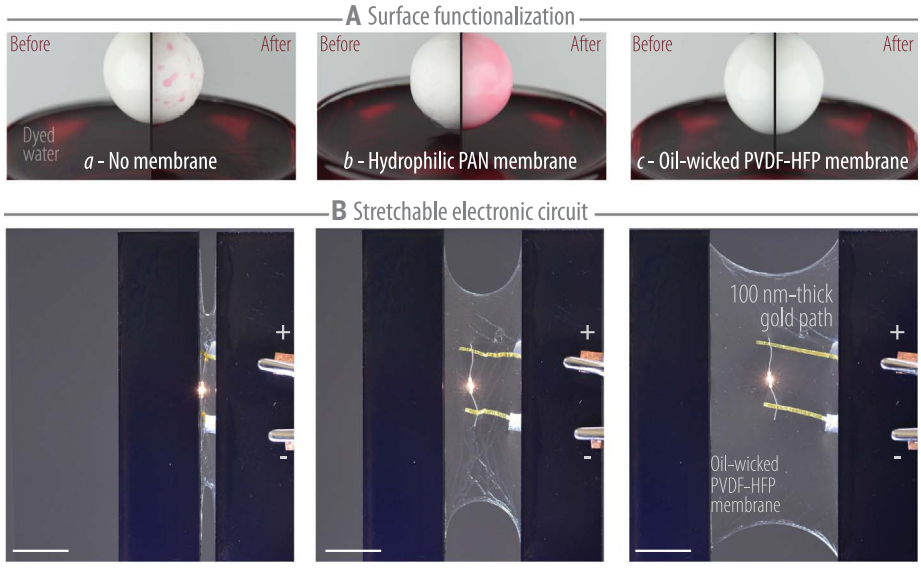


Fig. 4. Conformability and stretchability for chemical and electrical functionalization of the wicked membrane. (**A**) Without previous treatment (a), a clean zircon bead is dipped in a dyed water bath and subsequently pulled out. A water film is drawn at the bead surface, and because water only partially wets zircon, it rapidly disintegrates into droplets. A dry hydrophilic PAN membrane (b) is now applied on the bead and the experiment is repeated. After dipping and extracting the bead around 10 times, water has percolated through the membrane, and the bead is coated with a fairly homogeneous dyed water film, the PAN membrane adapts to the bead surface and secures the water film, thus providing a hydrophilic surface treatment. The bead is now covered with a silicone oil–wicked PVDF-HFP membrane (c), which acts as a water repellent coating; when the bead is dipped and pulled out of the bath, no water is drawn at its surface. Diameter of the zircon bead: 1.5 cm. (**B**) Two 100-nm-thick gold strips are affixed to a silicone oil–wicked PVDF-HFP membrane and are connected to a 1.55-V LED. Capillary adhesion secures the gold paths to the membrane and in the compressed state, they follow the folding of membrane reservoirs. Electricity runs through this elementary circuit when membrane reservoirs are smoothed out upon the reversible factor of 8 extension. Scale bar, 1 cm.

stretching at the molecular level. This results in a marked resistance of this material to fatigue, as illustrated by the unvarying mechanical response of the membrane after more than 100,000 cycles of extension and compression by a factor of 10 (*16*) (Fig. 3C and figs. S3 and S4).

The mechanics of reservoir folding and unfolding is a priori material-independent, as it relies only on a combination of elasticity, capillarity, and geometry (see tested polymers in table S1). To illustrate a few potential applications, we present some proofs of concept in Fig. 4. First, we demonstrate how the natural conformability of the wicked membrane can be used to confer instant chemical functions to nonplanar surfaces. As an example, we tune the chemical surface properties of a zircon bead. At its native state, the bead is dipped and pulled out from a dyed-water bath, thus withdrawing a liquid film. Zircon is only partially wet by water, and the liquid film coating the bead therefore rapidly disintegrates into a series of droplets (Fig. 4A, panel a, and movie S6). We now cover the bead with a hydrophilic (PAN) membrane and repeat the experiment. As soon as the bead is pulled out from the bath, the membrane folds within the drawn liquid film and therefore adapts to the bead curvature (Fig. 4A, panel b, and movie S7). Here, the membrane acts as an adaptable scaffold that secures and stabilizes the liquid film onto the bead. Conversely, this strategy can also be used to prevent contact between the bead and water: If the bead is covered with a silicone oil–wicked oleophilic membrane (a PVDF-HFP membrane), its surface inherits both the immiscible character and low surface energy of oil. After the bead is withdrawn from the bath, no water trace remains at the bead surface (Fig. 4A, panel c, and movies S8 and S9). We now show how stretchability can be combined with electrical function (fig. S7). In Fig. 4B, we affix 100-nm-thick gold paths to a wicked PVDF-HFP membrane to design an elementary electronic circuit. A LED is powered through the metallic tracks that follow the deformation of the membrane. When the membrane is compressed, the tracks are localized in the membrane reservoirs. Upon extension, these tracks are unfolded, thus exhibiting conductivity throughout a factor of 8 extension-compression cycle (*16*) (Fig. 4B and movie S10).

The use of membrane reservoirs to fuel large shape changes is encountered in a variety of animal cells. Our results show that capillarity-induced folding in thin wicked membranes endows conventional materials with large effective stretchability and conformability, and thereby constitutes a promising tool for the design of soft deformable materials.

REFERENCES AND NOTES

1. H. Elettro, S. Neukirch, F. Vollrath, A. Antkowiak, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6143–6147 (2016).
2. J. Lam, M. Herant, M. Dembo, V. Heinrich, *Biophys. J.* **96**, 248–254 (2009).
3. L. Guillou *et al.*, *Mol. Biol. Cell* **27**, 3574–3582 (2016).
4. D. Raucher, M. P. Sheetz, *Biophys. J.* **77**, 1992–2002 (1999).
5. J. Dai, M. P. Sheetz, *Biophys. J.* **68**, 988–996 (1995).
6. N. Groulx, F. Boudreault, S. N. Orlov, R. Grygorczyk, *J. Membr. Biol.* **214**, 43–56 (2006).
7. J. A. Nichol, O. F. Hutter, *J. Physiol.* **493**, 187–198 (1996).
8. C. A. Erickson, J. P. Trinkaus, *Exp. Cell Res.* **99**, 375–384 (1976).
9. S. Majstoravich *et al.*, *Blood* **104**, 1396–1403 (2004).
10. G. Salbreux, G. Charras, E. Paluch, *Trends Cell Biol.* **22**, 536–545 (2012).
11. J. A. Rogers, T. Someya, Y. Huang, *Science* **327**, 1603–1607 (2010).
12. W. Liu, M.-S. Song, B. Kong, Y. Cui, *Adv. Mater.* **29**, 1603436 (2017).
13. J. Hu, H. Meng, G. Li, S. I. Ibekwe, *Smart Mater. Struct.* **21**, 053001 (2012).
14. R. F. Shepherd *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20400–20403 (2011).
15. A. Lazarus, P. M. Reis, *Adv. Eng. Mater.* **17**, 815–820 (2015).
16. See supplementary materials.
17. E. Cerda, L. Mahadevan, *Phys. Rev. Lett.* **90**, 074302 (2003).
18. T. Tallinen *et al.*, *Nat. Phys.* **12**, 588–593 (2016).
19. H. Vandeparre *et al.*, *Phys. Rev. Lett.* **106**, 224301 (2011).
20. B. Davidovitch, R. D. Schroll, D. Vella, M. Adda-Bedia, E. A. Cerda, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18227–18232 (2011).
21. B. Roman, J. Bico, *J. Phys. Condens. Matter* **22**, 493101 (2010).
22. B. Roman, A. Pocheau, *Europhys. Lett.* **46**, 602–608 (1999).
23. B. Audoly, Y. Pomeau, *Elasticity and Geometry: From Hair Curls to the Nonlinear Response of Shells* (Oxford Univ. Press, 2010).

ACKNOWLEDGMENTS

We thank I. Genois for SEM pictures, F. Poydenot for electrical resistance characterization, and T. Bastien for the fabrication of the fatigue test setup. **Funding:** Supported by ANR grant ANR-14-CE07-0023-01 and by CNRS PEPS-PTI and PICS grants. **Author contributions:** P.G., A.A., and S.N. designed the study; N.K. and C.L.-R. designed the electrospinning setup; N.K. and P.G. fabricated the membranes; P.G. and A.H.-F. performed the experiments and acquired the data; A.A., P.G., A.H.-F., and S.N. interpreted the data and proposed the mechanical models; and A.A. and P.G. wrote an initial version of the manuscript and all authors reviewed it. **Competing interests:** None. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. **Patent:** P.G., A.A., N.K., and C.L.-R. are inventors on French patent application FR 1751950 submitted by Sorbonne Université on "composite membrane and manufacturing process of such a membrane."

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/296/suppl/DC1
Materials and Methods
Supplementary Text
Table S1
Figs. S1 to S20
Movies S1 to S10
Data Files
References (*24–26*)

27 September 2017; accepted 22 February 2018
10.1126/science.aaq0677

MATERIALS SCIENCE

Ultralarge elastic deformation of nanoscale diamond

Amit Banerjee,[1,2]* Daniel Bernoulli,[3]* Hongti Zhang,[1,4]* Muk-Fung Yuen,[2,5] Jiabin Liu,[1] Jichen Dong,[6] Feng Ding,[6,7] Jian Lu,[1,4] Ming Dao,[3]† Wenjun Zhang,[2,5]† Yang Lu,[1,2,4]† Subra Suresh[8]†

Diamonds have substantial hardness and durability, but attempting to deform diamonds usually results in brittle fracture. We demonstrate ultralarge, fully reversible elastic deformation of nanoscale (~300 nanometers) single-crystalline and polycrystalline diamond needles. For single-crystalline diamond, the maximum tensile strains (up to 9%) approached the theoretical elastic limit, and the corresponding maximum tensile stress reached ~89 to 98 gigapascals. After combining systematic computational simulations and characterization of pre- and postdeformation structural features, we ascribe the concurrent high strength and large elastic strain to the paucity of defects in the small-volume diamond nanoneedles and to the relatively smooth surfaces compared with those of microscale and larger specimens. The discovery offers the potential for new applications through optimized design of diamond nanostructure, geometry, elastic strains, and physical properties.

Diamond, the hardest natural material, is notably stiff and durable. It is also a member of the class of carbon materials with a multitude of applications in mechanics, biomedicine, electronics, and photonics (*1*–*7*). Controlled introduction of microstructures can further enhance the strength and hardness of diamond (*8*, *9*). However, the mechanical characteristics of diamond are offset by poor deformability and relatively high brittleness (*10*). These limitations of diamond motivated investigations (*11*, *12*) of pathways and mechanisms for inducing appreciable or even ultralarge elastic deformation. Such efforts with diamond and other nominally brittle materials also entail development of new methods for elastic strain engineering (*13*–*15*) whereby engineered band structures and corresponding enhancements in electronic and optical properties (*16*, *17*) can be attained. In this study, we demonstrate the occurrence of large, reversible elastic deformation in nanoscale needles of both single-crystalline and polycrystalline diamond. We investigated key characteristics of this reversible elastic deformation through nanomechanical bending experiments conducted in situ on the diamond nanoneedles inside a scanning electron microscope (SEM). We complemented the experimental studies with detailed computational simulations involving the finite element method (FEM) and molecular dynamics to quantify the local tensile and compressive stress and strain variations and the mechanisms underlying mechanical deformation.

We fabricated nanoscale diamond needles by plasma-induced etching of diamond thin films deposited on Si substrates through bias-assisted chemical vapor deposition (*5*). We controlled the shape, size, crystallinity, and density of the nanoneedles by the appropriate choice of growth and etching parameters during deposition. We etched the single-crystalline diamond nanoneedles from a <111>-oriented diamond film (figs. S1 and S2) (*18*).

We mounted the diamond nanoneedles on the Si substrate inside a SEM nanoindenter system (Fig. 1A) for quantitative in situ nanomechanical characterization (Fig. 1B). The downward motion of the indenter tip generated a sideward displacement from the inclined tip surface, thereby bending the nanoneedle after contact (Fig. 1, B and C). With the use of high-resolution transmission electron microscope (HRTEM) imaging, we found that our needles had a smooth surface and a pristine single-crystalline diamond structure along the <111> growth orientation aligned with the nanoneedle axis (Fig. 1D). The fracture mode, when the nanoneedle was bent to failure, was brittle, in line with expectations (Fig. 1E).

We simulated elastic deformation with a FEM model using the actual indenter tip and geometry for each individual nanoneedle (*18*). In FEM simulations, we invoked analyses of nonlinear elasticity and elastic anisotropy (fig. S3), as well as frictional contact between the indenter tip and the nanoneedle with friction coefficients f between 0.1 and 1.0 (fig. S4). We chose these values on the basis of previous observation (*19*)

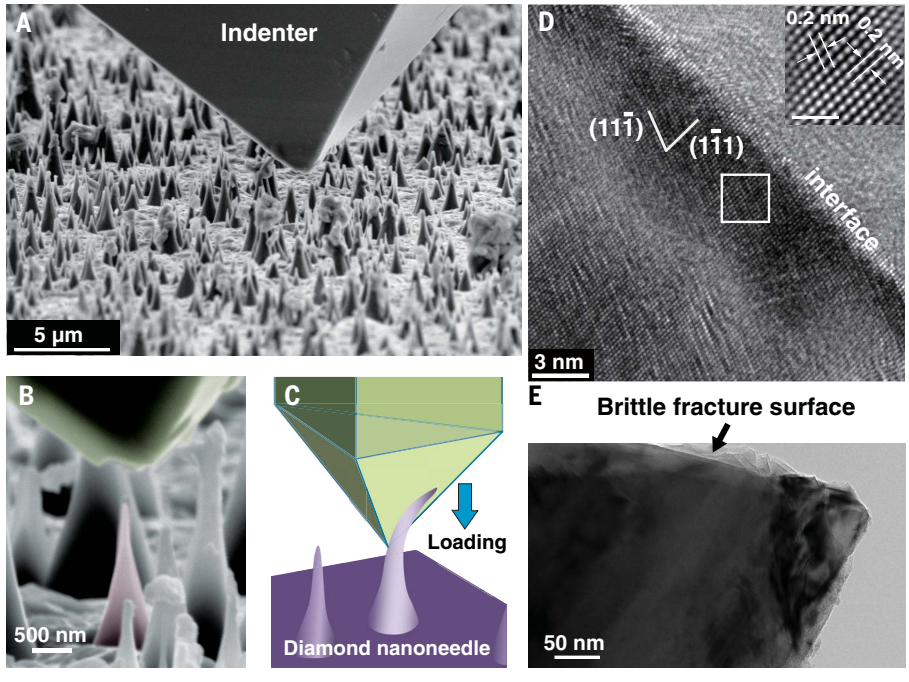


Fig. 1. Experimental setup and nanoneedle characterization. (**A** and **B**) SEM micrographs showing (A) the relative orientation of the diamond nanoneedles with respect to the indenter tip in the experimental configuration and (B) a close-up view of the alignment of the "cube corner" diamond tip of the nanoindenter and a freestanding diamond nanoneedle. (**C**) Schematic representation of the bending deformation of the nanoneedles by the indenter tip's downward motion. (**D** and **E**) TEM micrographs depicting (D) the atomic-scale structure of a pristine diamond nanoneedle sample (inset scale bar, 1 nm) and (E) the smooth fracture surface of a fractured nanoneedle. The inset corresponds to the boxed region in (D).

[1]Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Kowloon, Hong Kong, China. [2]Centre of Super-Diamond and Advanced Films (COSDAF), City University of Hong Kong, Kowloon, Hong Kong, China. [3]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Centre for Advanced Structural Materials, Shenzhen Research Institute of City University of Hong Kong, Shenzhen 518057, China. [5]Department of Materials Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong, China. [6]Center for Multidimensional Carbon Materials, Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea. [7]Department of Materials Science and Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea. [8]Nanyang Technological University, Singapore 639798, Republic of Singapore.
*These authors contributed equally to this work.
†Corresponding author. Email: mingdao@mit.edu (M.D.); apwjzh@cityu.edu.hk (W.Z.); yanglu@cityu.edu.hk (Y.L.); ssuresh@ntu.edu.sg (S.S.)

that the friction coefficient for diamond on diamond in low-pressure environments, such as the one existing inside the electron microscope, can be much larger than that measured in ambient air.

We found instantaneous full recovery of the nanoneedles to their original, undeformed shape (Fig. 2) when we retracted the nanoindenter tip before the onset of catastrophic fracture. An example of this process is shown in Fig. 2, which presents typical results for a single-crystalline diamond nanoneedle subjected to bending (Fig. 2A, A1 to A4, and movie S1). The maximum horizontal deflection (δ) of the needle tip in Fig. 2A was 442 nm (~19% of the needle's length) at a maximum indentation displacement of 700 nm.

We conducted experiments with larger indentation displacements (Fig. 2B, B1 to B3, and movie S2). We observed δ of 464 nm (~20% of the needle's length) (Fig. 2B, B1) at a maximum indentation displacement of 734 nm just before catastrophic fracture (Fig. 2B, B2 and B3). The needle in Fig. 2 had the largest amount of elastic bending. The smooth fracture surface (highlighted in red in Fig. 2B) indicates that catastrophic failure of the nanoneedle occurred immediately after elastic deformation by atomic separation along preferred crystallographic planes. We confirmed this assessment with ex situ TEM observation of the fractured nanoneedle (Fig. 1E), which showed that the fracture surface was oriented at ~70° with respect to the basal plane. This measurement matches closely with the angle between two {111} planes that are known to be the dominant shear surfaces in single-crystalline diamond (*20*). We found a close overlap between the loading portion of the force-versus-displacement curve (blue curve in Fig. 2C) for the elastic deformation illustrated in Fig. 2A and the subsequent reloading portion of that curve (the portion after complete unloading) for the same nanoneedle (red curve in Fig. 2C) loaded to final fracture as depicted in Fig. 2B. These results indicate fully recoverable elastic deformation during repeat experiments. The hysteresis exhibited by the fully reversible elastic loading-unloading curve (Fig. 2C) arises from frictional contact between the indenter tip and the nanoneedle.

Figure 2 shows FEM simulation of deformed geometry, along with predictions of the maximum principal strain distribution (Fig. 2D) for the monocrystalline nanoneedle just before the onset of catastrophic fracture for the experimental conditions (Fig. 2B, B1). We assumed f of 1.0 for contact between the indenter tip and the needle and Young's modulus E of 1100 GPa for diamond (*18*). We estimated a local maximum tensile stress of ~98 GPa from the simulation, corresponding to a local (tensile) maximum principal strain of 8.88% on the tension side of the bent needle at the maximum indentation displacement of 734 nm. Our simulation of the bent shape of the nanoneedle (Fig. 2D) matches the shape we observed experimentally (B1 in Fig. 2B). Our simulated δ of 478 nm matches well with our experimental observation of 464 nm. We also carried out FEM simulation with smaller friction coefficients, where the local (tensile) maximum principal strain was 8.92% for f = 0.1 and 8.93% for f = 0. These simulations confirm that the maximum strain is determined mainly by the extent of bending and that it is not sensitive to the friction coefficient (fig. S4) or the diamond elastic modulus (table S1). The estimated maximum local tensile stress, however, does depend on the assumed elastic modulus in FEM simulations and would be 89 to 98 GPa (table S1) for the case in Fig. 2B, corresponding to the commonly observed diamond modulus of 1000 to 1100 GPa (*10*).

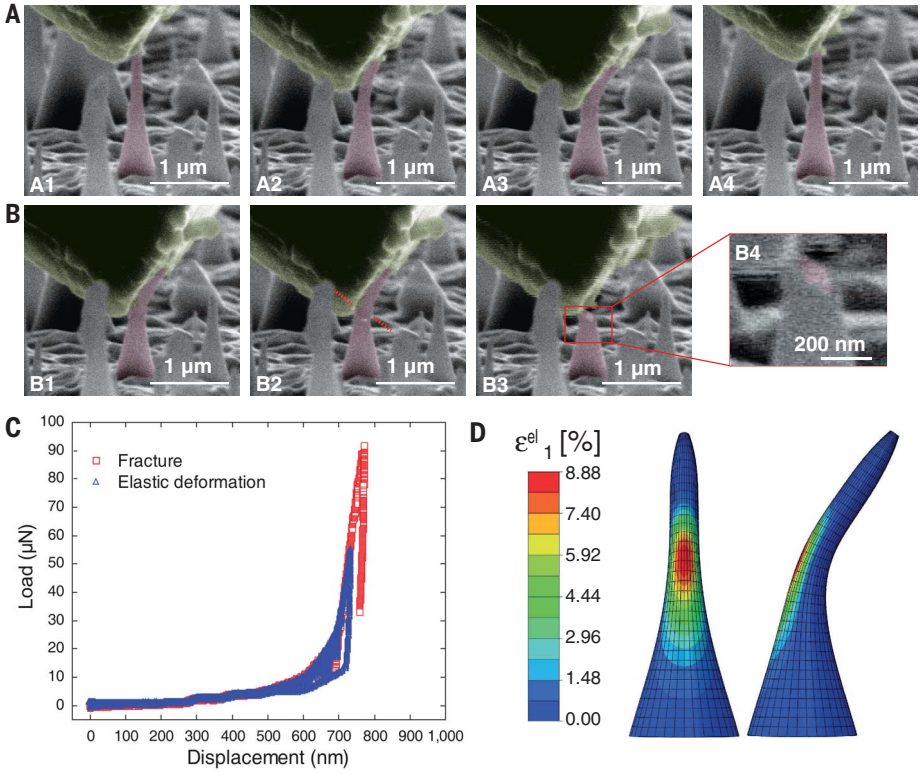


Fig. 2. Ultralarge elastic deformation achieved in a single-crystalline nanoneedle. (**A**) SEM micrographs of the bending deformation of a nanoneedle during loading and subsequent full recovery after unloading. (**B**) Subsequent rerun with the same nanoneedle and larger deformation to the point of catastrophic fracture (B2 shows the maximum deformation immediately before the fracture, B3 shows a moment after the fracture has occurred, and B4 provides a close-up view of the smooth fracture surface, with a characteristic angular orientation). (**C**) Load-displacement curves measured by pushing the nanoindenter tip onto the nanoneedle for the fully reversible elastic deformation (blue curve) and the final fracture run (red curve). (**D**) FEM simulation of the bending process, reproducing the shape of the bending and showing the local elastic (tensile) maximum principal strain (ε^{el}_1) distribution for the nanoneedle, with the frictional coefficient taken as f = 1.0.

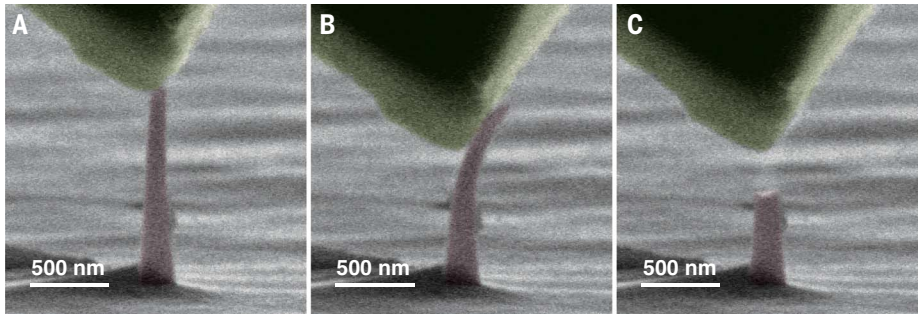


Fig. 3. Large elastic deformation achieved in a polycrystalline nanoneedle. (**A** to **C**) SEM image sequence of bending deformation of a typical polycrystalline nanoneedle, where (B) shows the maximum deformation before fracture and (C) shows the nanoneedle immediately after fracture has occurred.

We carried out five repeat experiments on five different single-crystalline diamond nanoneedles. We observed large, reversible, elastic deformation in each case. We loaded three out of five nanoneedles to the point of fracture. The peak local maximum principal strain from FEM simulation ranged from ~9 to 4% (Fig. 2B, fig. S5, and movie

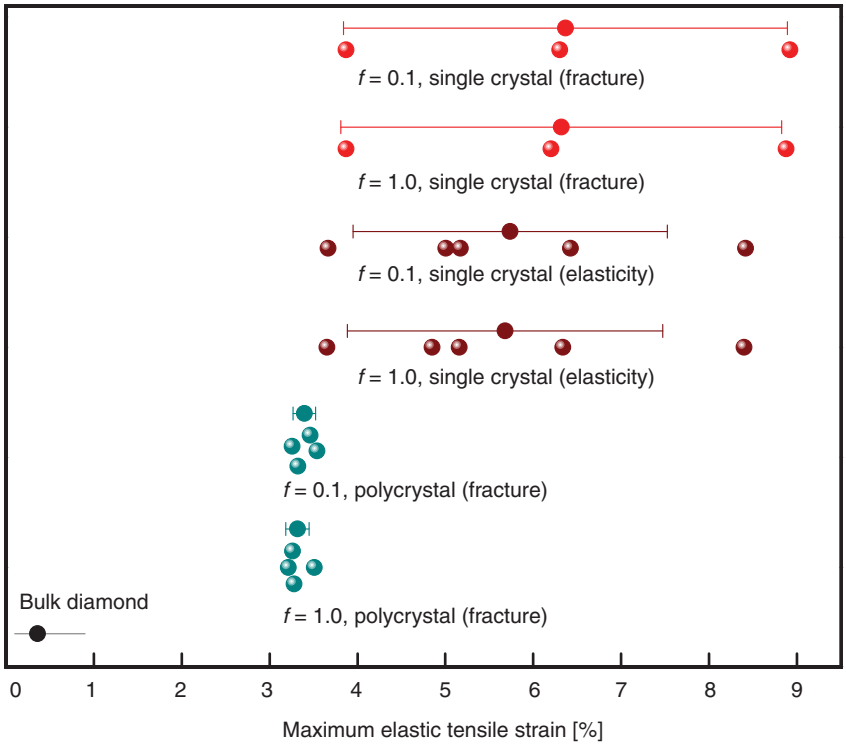


Fig. 4. Summary of the maximum elastic tensile strains. The results were extracted by FEM analysis on the basis of nanoneedle bending experiments, assuming friction coefficient f = 1.0 or 0.1. The data from eight experiments with five different single-crystalline samples (final fracture and fully reversible runs are shown by red and brown data points, respectively) and four experiments with four different polycrystalline samples (shown by cyan circles) are presented. These large elastic strains are much higher than the previously reported bulk diamond elastic limit (shown by the black data point) (*21*).

S3) for single-crystal samples loaded to the point of fracture. We also carried out additional experiments by bending the single-crystalline needles with sideward indenter tip movement, again achieving high tensile strains (fig. S6 and movie S4) (*18*).

We performed similar in situ bending experiments on polycrystalline diamond nanoneedles (Fig. 3, figs. S7 and S8, and movies S5 and S6). Unlike those in single-crystalline needles, fracture surfaces in polycrystalline needles were nearly parallel to the substrate (Fig. 3C). The local (tensile) maximum principal strain from FEM simulation was between 3.51% (f = 1.0) and 3.54% (f = 0.1) for the polycrystalline needle, which is approximately one-half the average for single crystals (*18*). The single-crystalline diamond nanoneedles generally achieved a much higher local maximum tensile strain than the polycrystalline ones (Fig. 4). For the single-crystalline nanoneedles, we achieved a mean maximum tensile strain of ~6% and a peak maximum tensile strain of ~9%. For the polycrystalline nanoneedles, the mean maximum tensile strain was ~3.3% and the peak maximum tensile strain was 3.5%. These large elastic strains are much higher than the previously reported bulk diamond elastic limit (*21*).

We conducted a hybrid density functional theory–molecular dynamics calculation (fig. S9, A and B), which showed the ideal maximum tensile strain to be ~13% along the <111> direction, with C–C bond fracture beyond this critical strain (*18*). We ascribed the ultrahigh elasticity of the diamond nanoneedles (~9% of the maximum tensile strain) to the paucity of internal defects and the relatively smooth surface (Fig. 1D). The nanoscale dimensions of the diamond needles (see the pristine single crystal in Fig. 1D) help avoid internal defects and improve smoothness, compared with that of microscale and macroscale specimens. These factors render it more difficult to initiate a surface crack at these heterogeneous nucleation sites (*18*).

In addition to the maximum elastic tensile strains shown in Fig. 4, we extracted the corresponding maximum elastic compressive strains (up to ~10% or more) observed on the compression side of each sample (fig. S10) (*18*). The related local maximum compressive stress reaches ~100 to 110 GPa (with a diamond modulus of 1000 to 1100 GPa), comparable to the values reported in (*22*).

We showed that single-crystalline diamond nanoneedles are capable of undergoing large elastic bending deformation approaching the theoretical elastic strain limit of diamond, with the corresponding local stress approaching the ideal strength. We demonstrated experimentally that single-crystalline needles are simultaneously ultrastrong and susceptible to large elastic deformation, with fully reversible mechanical deformability of up to a maximum of 9% of elastic tensile strain. Shrinking the diamond needles offers another pathway for achieving elastic strain engineering (*13*–*17*) of diamond. With a focus on physical properties, such as deformation, lattice strains, and band structure, diamond could be engineered to enhance the structural and functional performance of small-volume structures. Large elastic deformation in nanoscale diamond needles could lead to performance enhancements in applications involving bioimaging and biosensing (*7*), strain-mediated nanomechanical resonators (*23*), drug delivery (*7*), data storage (*24*), and optomechanical devices (*25*), as well as ultrastrength nanostructures (*13*, *14*).

REFERENCES AND NOTES

1. A. L. Robinson, *Science* **234**, 1074–1076 (1986).
2. W. Zhu, G. P. Kochanski, S. Jin, *Science* **282**, 1471–1473 (1998).
3. I. Aharonovich, A. D. Greentree, S. Prawer, *Nat. Photonics* **5**, 397–405 (2011).
4. V. N. Mochalin, O. Shenderova, D. Ho, Y. Gogotsi, *Nat. Nanotechnol.* **7**, 11–23 (2011).
5. X. Zhu *et al.*, *Adv. Healthc. Mater.* **5**, 1157–1168 (2016).
6. Y. Wang *et al.*, *Nat. Commun.* **5**, 4466 (2014).
7. X. Chen, W. Zhang, *Chem. Soc. Rev.* **46**, 734–760 (2017).
8. K. Tanigaki *et al.*, *Nat. Commun.* **4**, 2343 (2013).
9. Q. Huang *et al.*, *Nature* **510**, 250–253 (2014).
10. J. E. Field, *The Properties of Natural and Synthetic Diamond* (Academic Press, 1992).
11. M. Seal, *Nature* **182**, 1264–1267 (1958).
12. R. J. Hemley, H. Mao, G. Shen, J. Badro, P. Gillet, M. Hanfland, D. Häusermann, *Science* **276**, 1242–1245 (1997).
13. T. Zhu, J. Li, *Prog. Mater. Sci.* **55**, 710–757 (2010).
14. J. Li, Z. Shan, E. Ma, *MRS Bull.* **39**, 108–114 (2014).
15. D. Yu, J. Feng, J. Hone, *MRS Bull.* **39**, 157–162 (2014).
16. V. M. Pereira, A. H. Castro Neto, *Phys. Rev. Lett.* **103**, 046801 (2009).
17. H. Watanabe, C. E. Nebel, S. Shikata, *Science* **324**, 1425–1428 (2009).
18. Materials and methods are available as supplementary materials.
19. Z. Feng, Y. Tzeng, J. E. Field, *J. Phys. D Appl. Phys.* **25**, 1418–1424 (1992).
20. R. H. Telling, C. J. Pickard, M. C. Payne, J. E. Field, *Phys. Rev. Lett.* **84**, 5160–5163 (2000).
21. J. E. Field, *Rep. Prog. Phys.* **75**, 126505 (2012).
22. J. M. Wheeler *et al.*, *Nano Lett.* **16**, 812–816 (2016).
23. P. Ovartchaiyapong, K. W. Lee, B. A. Myers, A. C. B. Jayich, *Nat. Commun.* **5**, 4429 (2014).
24. S. Dhomkar, J. Henshaw, H. Jayakumar, C. A. Meriles, *Sci. Adv.* **2**, e1600911 (2016).
25. B. Khanaliloo *et al.*, *Phys. Rev. X* **5**, 041051 (2015).

ACKNOWLEDGMENTS

Funding: The work was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, under the projects CityU11216515 and CityU11207416 and by the National Natural Science Foundation of China under the project 51301147. W.Z. acknowledges the National Natural Science Foundation of China (grants 51372213 and 51672230). M.D. and D.B. acknowledge support by the Singapore-MIT Alliance for Research and Technology (SMART). J. Lu acknowledges support by the Shenzhen Science, Technology and Innovation Commission under grant JSGG20141020103826038. D.B. acknowledges support by the Advanced Postdoc Mobility fellowship from the Swiss National Science Foundation under grant P300P2_167604. J.D. and F.D. acknowledge the support of the IBS of South Korea (grant IBS-R019-D1). A.B. acknowledges a CityU International Transition Team–Graduate Teaching Assistants scheme (ITT-GTA) postdoctoral fellowship. S.S. thanks Nanyang Technological University, Singapore, for support through the Distinguished University Professorship. **Author contributions:** Y.L., W.Z., M.D., and S.S. supervised the research. Y.L. designed the experiments. A.B., D.B., H.Z., M.-F.Y., J. Liu, J.D., and F.D. performed the research. A.B., D.B., H.Z., J. Lu, M.D., W.Z., Y.L., and S.S. analyzed the data and discussed the results. A.B., D.B., H.Z., M.D., W.Z., Y.L., and S.S. wrote the manuscript. All authors reviewed and contributed to the final manuscript. **Competing interests:** None declared. **Data and materials availability:** All data are reported in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/300/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Table S1
References (*26*–*45*)
Movies S1 to S6

6 November 2017; accepted 5 March 2018
10.1126/science.aar4165

COLLOIDS

Five-dimensional imaging of freezing emulsions with solute effects

Dmytro Dedovets,[1] Cécile Monteux,[2] Sylvain Deville[1]*

The interaction of objects with a moving solidification front is a common feature of many industrial and natural processes such as metal processing, the growth of single crystals, the cryopreservation of cells, or the formation of sea ice. Interaction of solidification fronts with objects leads to different outcomes, from total rejection of the objects to their complete engulfment. We imaged the freezing of emulsions in five dimensions (space, time, and solute concentration) with confocal microscopy. We showed that the solute induces long-range interactions that determine the solidification microstructure. The local increase of solute concentration enhances premelting, which controls the engulfment of droplets by the front and the evolution of grain boundaries. Freezing emulsions may be a good analog of many solidification systems where objects interact with a solidification interface.

Solidification fronts interact with hard or soft objects such as particles, bubbles, droplets, or cells, in three different ways: Objects can be engulfed, rejected for some time and then encapsulated, or rejected for an extended period of time (*1*). Different outcomes are desired, depending on the context. Particle-reinforced metal alloys require homogeneous distribution of particles in the matrix, and immediate engulfment is therefore preferred (*2*). By contrast, for single-crystal growth, a complete rejection of impurities is essential (*3*). The rejection of dissolved hydrogen can form gas bubbles that are eventually trapped by solidification fronts, becoming a major source of defect in many metals and alloys (*4*). Reproductive and red blood cells can be physically damaged when engulfed between growing ice crystals during cryopreservation procedures (*5*, *6*). The processing of porous materials by ice templating relies on the complete segregation of particles by the solidification front (*7*). Our ability to control the solidification microstructure is essential in all these domains.

The dynamics of interfaces and objects are an inherent feature of these phenomena but are extremely difficult to observe because of the time scale (solidification fronts moving rapidly), spatial scale (submicrometer objects), and sometimes, high temperature. In addition, the role of solute is often important in these systems, from the gas and impurities in metals to the additives used in cryopreservation procedures or food engineering. Being able to follow in situ the concentration of the various phases in the presence of objects is crucial yet extremely challenging. Direct photometric measurements of the solute concentration in thin layers of solution during solidification have been reported (*8*), but only in two dimensions (2D), and not in the presence of particles. Confocal Raman microspectroscopy was used to measure the static concentration of a cryoprotective agent (*9*) in frozen cells encapsulated in an ice matrix.

Progress in solidification research has thus largely relied on the study of analogs—transparent materials that solidify near room temperature—by optical microscopy (*10*). However, 3D imaging cannot be performed and chemistry cannot be resolved. Thanks to recent progress in x-ray computed tomography, we can obtain 3D, time-lapse images of the microstructure evolution in situ during the solidification of metals (*11*). The volume investigated is nevertheless limited, particles can barely be resolved, the temporal resolution is not good enough to capture the interaction events (*12*), and quantifying the solute concentration in the presence of objects is still not possible. It is thus fundamentally difficult to track in situ the development of 3D solidification microstructures where objects interact with a solid/liquid interface (*13*), even more so in the presence of solute effects.

We studied the solidification process by confocal microscopy, translating a Hele-Shaw cell filled with an oil-in-water emulsion at constant velocity V through a fixed temperature gradient ΔT (Fig. 1A and movie S1). In this way, the solidification front propagates with a velocity V through the sample. A typical confocal image of the freezing emulsion is shown in Fig. 1B. The sulforhodamine B fluorescence served as a local proxy for the surfactant concentration (fig. S1).

We used the droplets' trajectories to measure their velocity in the observation frame, U', and deduce their velocity in the sample frame, $U = V - U'$ (fig. S2). An example of the variation of the droplet velocity U with time (Fig. 2A) shows that it reaches a maximum as the droplet approaches the solidification front. The droplets start being repelled from the front at a distance L_v on the order of 100 μm, accelerate up to a maximal velocity U_{max} of 1.2 μm/s, and rapidly decelerate as they cross the solidification front (Fig. 2B). The droplets are eventually completely engulfed and stop moving. The time during which the droplets are repelled by the front before being engulfed is called the interaction time, plotted as a function of interface velocity in Fig. 2D. During the interaction time, droplets are displaced by a large distance, from 2 μm to more than 100 μm (Fig. 2E).

To understand the origin of the droplet displacement, we measured the solute concentration profile as a function of the distance from the solidification front. The solute concentration (Fig. 2C) is constant far from the front and then progressively increases. The concentration abruptly drops as we cross the interface because of the low solubility of solute in ice. The droplet starts being repelled by the interface at a distance L_v

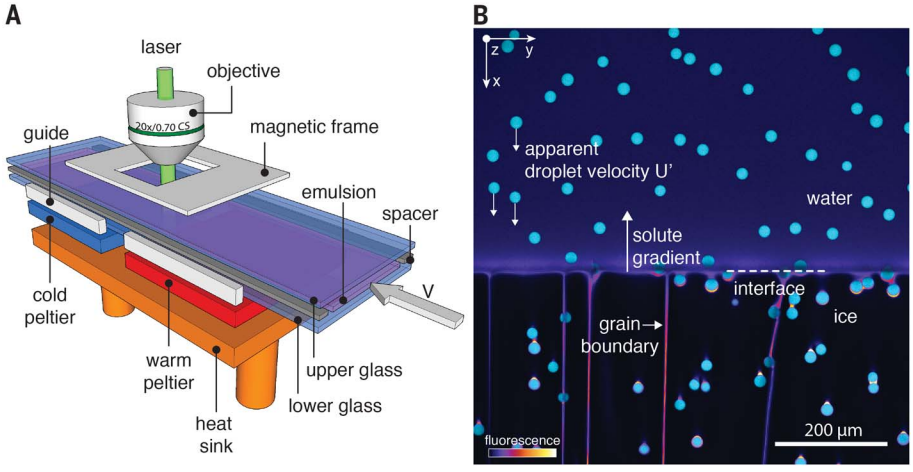


Fig. 1. Cooling stage and typical confocal image of a freezing emulsion. (**A**) Stage for the directional solidification experiment with a Hele-Shaw cell, placed under the objective of the confocal microscope. (**B**) Typical confocal image of the emulsion during freezing (front velocity of 3 μm/s, temperature gradient of 5°C/mm). The ice is shown in black, the liquid in magenta (see fluorescence color bar), and the oil droplets in cyan.

[1]Laboratoire de Synthèse et Fonctionnalisation des Céramiques, UMR 3080 CNRS/Saint-Gobain CREE, Cavaillon, France. [2]Sciences et Ingénierie de la Matière Molle, ESPCI Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ Paris 06, Paris, France.
*Corresponding author: Email: sylvain.deville@saint-gobain.com; Twitter: @DevilleSy

close to L_c—the distance at which the concentration of solute increases. This is more apparent when we plot L_v as a function of L_c for several front velocities (Fig. 2F), which shows that L_v increases linearly with L_c.

The presence of a surfactant in the system results in driving forces that can displace droplets over up to hundreds of micrometers from the solidification front, on a time scale of tens to hundreds of seconds. These results depart from the behavior observed and predicted by most physical models of the interaction of a particle with a solidification front (*1*) which, in the absence of solute, assumed a dominant role of thermomolecular forces in the interactions between the object and the front (*5*). However, such interactions are only effective over very short distances (nanometers). The distances measured in this study are several orders of magnitude larger. In the presence of a solute such as a surfactant, thermomolecular forces seem to play a minor role.

We suggest that the driving force for the motion of the droplets is diffusiophoresis or solutal Marangoni effect, which are known to induce micrometer-scale motion of droplets or particles in gradients of solute concentration (*14*, *15*). The droplet velocity increases with the gradient of surfactant concentration along the x axis (Fig. 2G). Diffusiophoresis is due to a gradient of osmotic pressure, whereas Marangoni flow is due to a gradient of surface tension along the droplet surface. In our case, the bulk surfactant concentration is two orders of magnitude above the critical micelle concentration and rises by a factor of 3 to 10 near the solidification front. Therefore, we expect a weak variation of the surfactant monomer concentration and of the surface tension along the x axis but a substantial variation in the concentration of micelles, which may lead to a diffusiophoretic motion.

The long-range interactions between the droplets and the front have a strong influence on the solidification microstructure. When the droplets' displacement (Fig. 2E) is larger than the interdroplet distance, the droplets form clusters as they get engulfed by the front (top of Fig. 2H, obtained for a front velocity of 1 μm/s). At 2 μm/s, the interaction time is shorter, and almost no aggregates are formed. For a front velocity of 3 μm/s, the displacement of droplets is significantly smaller than the mean interdroplet distance, which results in the preservation of the original spatial distribution of droplets upon solidification. In all these cases, the droplets are engulfed by the growing ice, but the microstructure of the solidified sample differs, depending on the solidification front velocity.

Most physical models developed over the past 60 years to describe the interaction of a particle with a solidification front focused on the criterion of critical particle size for a given velocity (or conversely, the critical velocity for a given size). The results shown here demonstrate how important the dynamics of interactions with the solidification front can be. In situations where

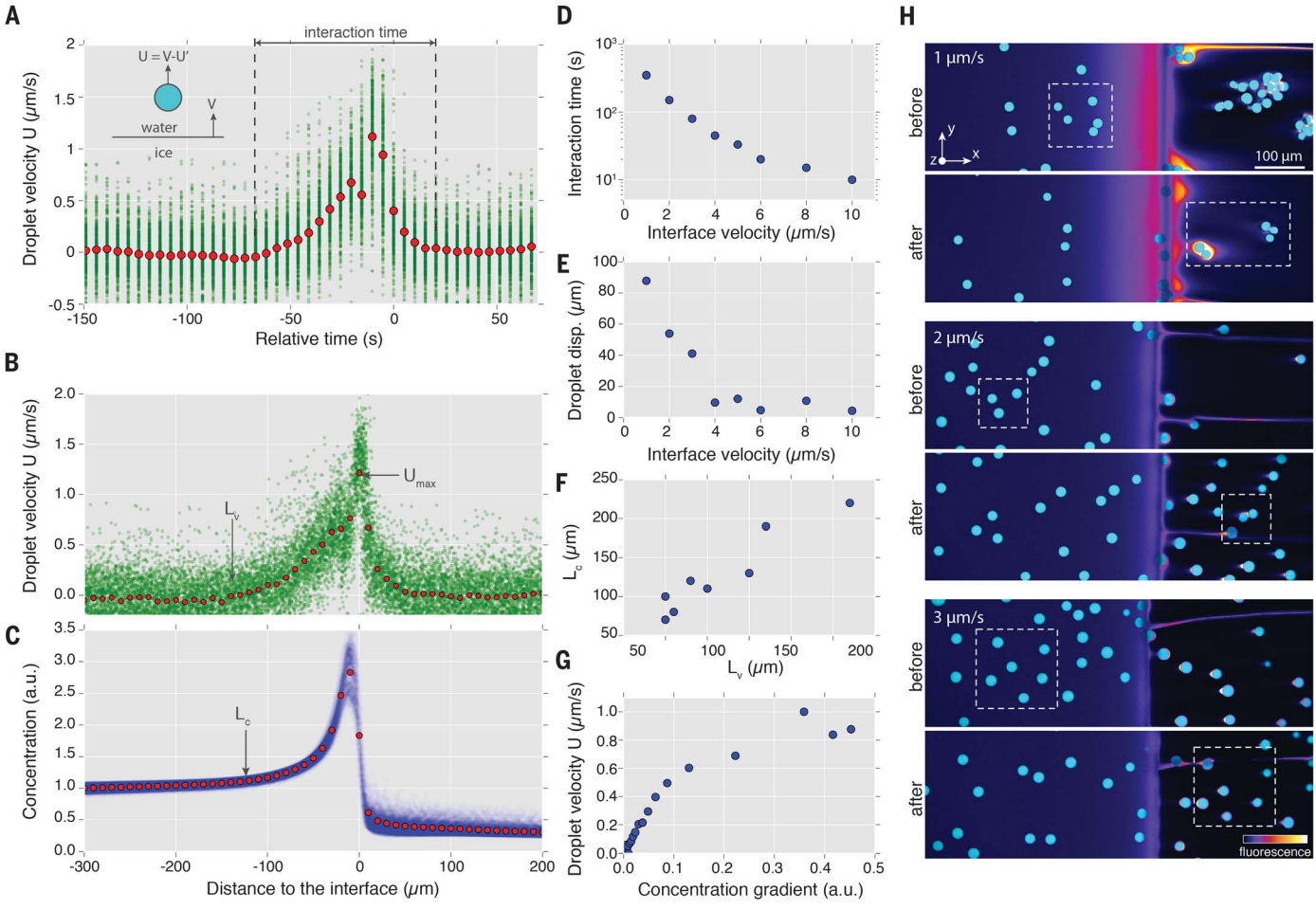


Fig. 2. Droplets dynamics and solidification microstructures (temperature gradient of 5°C/mm). (**A**) Droplet velocity versus relative time. Time zero is when the front edge of the droplet is at the solidification front position. (**B** and **C**) Droplet velocity (B) and solute concentration (C) versus distance to interface, for a solidification front velocity of 3 μm/s. (**D** and **E**) Interaction time (D) and droplet displacement (E) versus interface velocity. The error bars do not exceed the size of the circles. (**F**) L_c versus L_v. (**G**) Droplet velocity versus concentration gradient of solute, for a solidification front velocity of 3 μm/s. (**H**) Effect of the droplet dynamics and interaction time on the development of solidification microstructure. The dashed boxes indicate the same groups of droplets before and after passing the interface.

the distribution of objects should be retained in the solidified microstructure, such as particle reinforcements in a metal matrix, the displacement of objects by the solidification front is a much more relevant parameter for understanding and controlling the solidification microstructures.

We investigated the solute redistribution in the premelted film at the ice grains boundaries and around the droplets after their engulfment. Premelting describes the existence of liquid films at solid surfaces at temperatures below the bulk freezing temperature, a phenomenon common to many solids, including ice. Premelting has consequences for a wide range of biological, geophysical, and technological processes (*16*), such as the heaving of frozen ground (*17*), glacier motion, weathering, material transport through ice, using ice core as climate proxies (*18*), or the mobility of particles in solid materials. The driving force for melting is a reduction of interfacial energy between the particles and the matrix. The concentration of impurities or solute that depress the freezing point of the melt can greatly enhance the interfacial premelting, which is then called solute premelting. Because of their small dimensions, imaging premelted films remains a challenge.

Confocal microscopy allows in situ, 3D imaging of the solute redistribution in the solidifying sample with a semiquantitative estimation of its concentration. In the solidified part of the sample, the solute is concentrated in the premelted film at the grain boundaries of the ice crystals (Fig. 3A, region 1) and around the engulfed oil droplets (Fig. 3A, region 2). Because the concentrated solute broadens the thickness of these wetting films, they can be observed.

At low solidification front velocities (1 and 2 μm/s), the interface is planar. Increasing the interface velocity to 3 μm/s (Fig. 3B) results in the formation of grain boundaries—periodic liquid films of concentrated solute. Their number and spacing increase with the interface velocity (fig. S1).

The lateral growth of ice crystals results in a progressive thinning of the premelted films at the grain boundaries, eventually leading to their breakup into spherical droplets by an interfacial instability (Fig. 3B). These microstructures are analogous to those developed in metal alloys with a miscibility gap (*19*) (Fig. 3, C and D). The transition from a planar to a cellular front and the destabilization of the premelted film into droplets depend on both the temperature gradient and the solidification velocity (Fig. 3E), similar to that in metals (*20*).

We can capture the instability initiation and propagation dynamics (Fig. 3F). The growing geometrical instabilities results in the formation of side branches that elongate and break up into droplets. This process is quite slow, on the time scale of tens of seconds, which is a result of low interfacial tension between the solid and liquid phase. The development of these instabilities in 3D (fig. S4) reveals that they initiate where the liquid film is the thinnest. This instability is analogous to the pinch-off in dendrites observed in metal alloy solidification, which is driven by capillarity (*21*).

The premelted film also plays a critical role in the droplet engulfment by the solidification front (Fig. 4A). At the approach of the droplet, the solidification front bends away from it owing to the formation of a pocket of concentrated solute, which colligatively depresses the freezing point of water. Analytical models predict a deflection of the solidification front toward an object if the thermal conductivity of the latter is lower than that of the melt (*22*). The thermal conductivity of oil (propyl benzoate) is smaller than that of water (0.141 W m^{-1} K^{-1} versus 0.569 W m^{-1} K^{-1}) and thus the formation of a convex interface is expected, which is what we observe in the absence of solute (fig. S5). However, in the presence of solute, we observe only concave interfaces. The solute is thus able to reverse the bending of the front, which indicates the dominant role of the solute in droplet–interface interaction.

The ice continues to grow over and around the droplet, until the latter is completely engulfed. Three-dimensional imaging (Fig. 4B) provides a complete picture of the morphological evolution of the premelted film around the droplet. As the solidification proceeds, the solute-rich liquid pocket, initially located in front of the oil droplet, passes on the opposite side of it—away from growing ice. From the fluorescence intensity, we obtain semiquantitative information of the evolution of the solute concentration in the premelted film around the oil droplet (Fig. 4C and fig. S6). We then correlate the solute concentration with the thickness of the liquid film (Fig. 4D), to reveal the dynamics of the process and see how the premelted film evolves toward its equilibrium thickness (6 μm in this case).

For the droplet size investigated, comet tail–shaped segregation patterns were systematically observed (Fig. 4E). These observations may be a good analog for industrial processes where pores and particles can affect the segregation pattern (Fig. 4, F to H). Understanding the

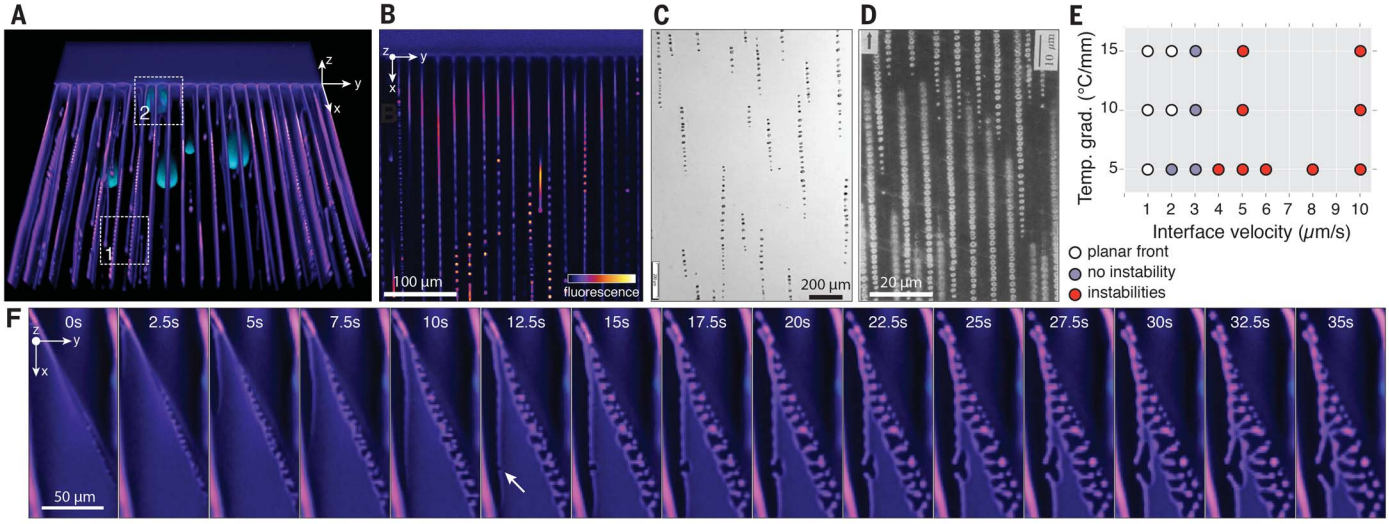


Fig. 3. Solute premelting around the droplet and at the grain boundaries. (**A**) Typical 3D view of the emulsion during freezing. Premelting plays a role on the destabilization of thin liquid film (concentrated solute) between ice crystals (grain boundaries, region 1), and on the evolution of the droplet/ice interface (region 2). (**B**) String-of-pearls microstructure (solidification front velocity of 3 μm/s, temperature gradient of 15°C/mm) similar to that developed in zinc-bismuth (*27*) (**C**) and aluminum-bismuth alloys (*19*) (**D**). (**E**) Stability diagram of the solidification front as a function of the temperature gradient and interface velocity. (**F**) Time-lapse sequence of the dynamics of a premelted film instability at a grain boundary during freezing. The white arrow indicates the location where a second instability is initiated. [Credits: (C) Reprinted from (*27*) by permission from Elsevier; (D) Reprinted from (*19*) by permission from Springer Nature]

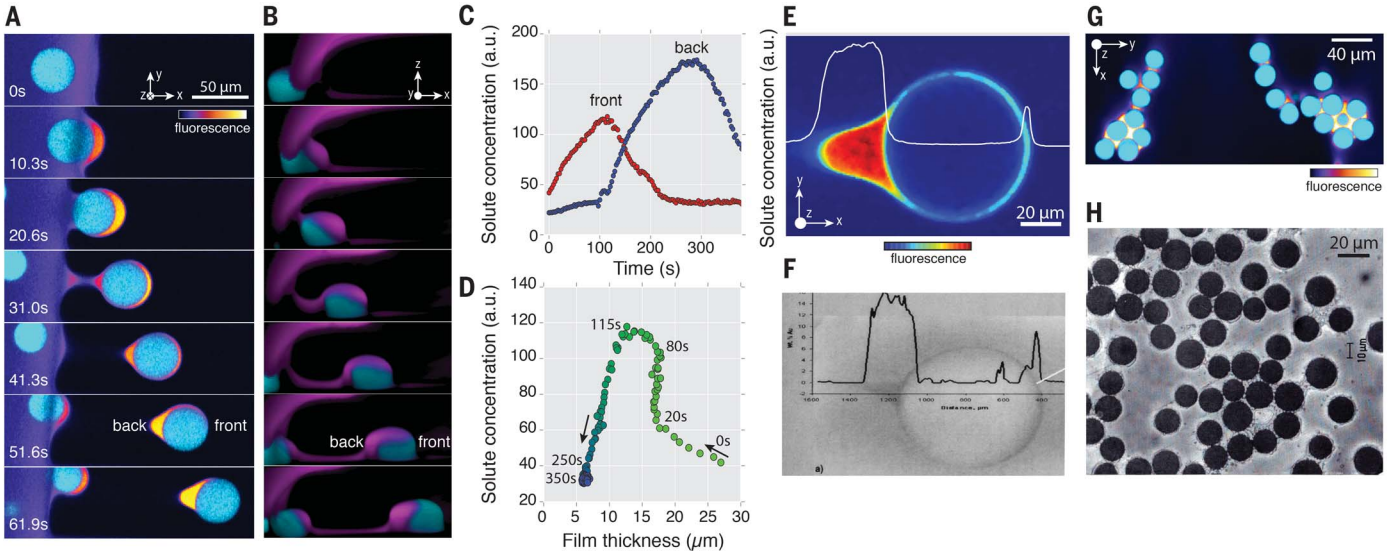


Fig. 4. Dynamics of premelted films around the droplets and effect on the solidification microstructure. (**A**) Typical interaction between a droplet and the solidification front, showing the accumulation of solute and its redistribution around the droplet. Horizontal cross section. Sample velocity: 2 µm/s; temperature gradient: 5°C/mm. (**B**) Same sequence, but in 3D and in vertical cross section. The meniscus formed by the solidification front can clearly be observed, as well as the premelted film dynamics. Most of the solute is eventually pushed below the crystal. (**C**) Time-lapse evolution of solute concentration ahead and behind the droplet. (**D**) Time-lapse evolution of solute concentration versus film thickness. (**E**) Solute-rich, comet tail–shaped region behind the droplet and corresponding solute concentration profile across the droplet in the frozen region. A similar solute segregation has been found in metal alloys (**F**), such as gold segregated around a porosity (*28*). The plot shows the gold concentration across the picture. A solute-rich, comet tail–shaped region is observed behind the pore. [Reprinted from (*28*) by permission from Springer Nature] (**G**) Agglomerates of droplets. The solute-rich regions around the droplets are the last regions to solidify. (**H**) An analogous behavior is observed in metals, where this segregation can lead to the formation of a secondary phase around objects, such as the $\theta(CuAl_2)$ phase in an Al–4.5 weight % Cu alloy formed around alumina fibers. [Micrograph courtesy of P. Bystricky and A. Mortensen (*29*)]

solute redistribution mechanisms could thus be an important step toward a better control of the solidification microstructures.

Solute also plays a critical role in the cryopreservation of red blood cells, platelets, fibroblasts, and reproductive cells (*23*). Cryoprotection agents are now systematically used to improve the survival rate. The crystal growth pattern and concentration of these agents between the growing crystals, along with the cells, affect their survival rates (*5*, *24*, *25*). Quantitative studies of solute concentration during freezing (but without cells) were done in thin films, by photometric measurements (*8*). However, when the integration of concentration is done through the depth of the sample, the peak solute concentration may be underestimated if the solidification microstructure is not homogeneous. Our results show that, because of the complex solute concentration and redistribution pathways, 5D imaging is required to better estimate the solute location and concentration. Our approach may lead to a better understanding of the physical damages inflicted on cells during the cryopreservation procedures and optimize such procedures.

The features observed during the freezing of emulsions seem thus to be good analogs of many solidification systems. The rapid 3D confocal microscopy imaging, along with the ability to map the solute concentration, provides new insights into the dynamics and mechanisms of solidification in the presence of foreign hard (particles) or soft (droplets, bubbles, cells) objects. The approach proposed here is versatile and can be easily adapted to a variety of solidification systems, from geophysics to food engineering or biology. We expect confocal microscopy to be a welcome addition to the toolbox available for the studies of interfacial evolution problems in 3D (*26*).

REFERENCES AND NOTES

1. R. Asthana, S. N. Tewari, *J. Mater. Sci.* **28**, 5414–5425 (1993).
2. L.-F. Zhang, *J. Iron Steel Res. Int.* **13**, 1–8 (2006).
3. O. M. Bunoiu, T. Duffar, I. Nicoara, *Prog. Cryst. Growth Charact. Mater.* **56**, 123–145 (2010).
4. A. K. Gupta, B. K. Saxena, S. N. Tiwari, S. L. Malhotra, *J. Mater. Sci.* **27**, 853–862 (1992).
5. C. Körber, *Q. Rev. Biophys.* **21**, 229–298 (1988).
6. K. L. Scott, J. Lecak, J. P. Acker, *Transfus. Med. Rev.* **19**, 127–142 (2005).
7. S. Deville, *Scr. Mater.* **147**, 119–124 (2018).
8. S. Kourosh, K. R. Diller, M. E. Crawford, *Int. J. Heat Mass Transfer* **33**, 39–53 (1990).
9. J. Dong, J. Malsam, J. C. Bischof, A. Hubel, A. Aksan, *Biophys. J.* **99**, 2453–2459 (2010).
10. I. Farup, J.-M. Drezet, M. Rappaz, *Acta Mater.* **49**, 1261–1269 (2001).
11. D. Tolnai *et al.*, *Acta Mater.* **60**, 2568–2577 (2012).
12. R. Daudin, S. Terzi, P. Lhuissier, L. Salvo, E. Boller, *Mater. Des.* **87**, 313–317 (2015).
13. S. Deville, in *Freezing Colloids: Observations, Principles, Control Use* (Springer International Publishing, 2017), pp. 47–90; https://link.springer.com/10.1007/978-3-319-50515-2.
14. J. L. Anderson, M. E. Lowell, D. C. Prieve, *J. Fluid Mech.* **117**, 107 (1982).
15. C. A. Weber, C. F. Lee, F. Jülicher, *New J. Phys.* **19**, 053021 (2017).
16. J. S. Wettlaufer, M. G. Worster, *Annu. Rev. Fluid Mech.* **38**, 427–452 (2006).
17. J. G. Dash, H. Fu, J. S. Wettlaufer, *Rep. Prog. Phys.* **58**, 115–167 (1999).
18. A. W. Rempel, *J. Glaciol.* **56**, 1122–1128 (2010).
19. B. Majumdar, K. Chattopadhyay, *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **27**, 2053–2057 (1996).
20. L. Ratke, *Mater. Sci. Eng. A* **413-414**, 504–508 (2005).
21. H. Neumann-Heyme, K. Eckert, C. Beckermann, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **92**, 060401 (2015).
22. M. S. Park, A. A. Golovin, S. H. Davis, *J. Fluid Mech.* **560**, 415 (2006).
23. A. Hubel, *Transfus. Med. Rev.* **11**, 224–233 (1997).
24. H. Ishiguro, B. Rubinsky, *Cryobiology* **31**, 483–500 (1994).
25. M. Uemura, H. Ishiguro, *Cryobiology* **70**, 122–135 (2015).
26. D. Rowenhorst, P. Voorhees, *Annu. Rev. Mater. Res.* **42**, 105–124 (2012).
27. L. Ratke, A. Muller, *Scr. Mater.* **54**, 1217–1220 (2006).
28. A. V. Catalina, S. Sen, D. M. Stefanescu, W. F. Kaukler, *Metall. Mater. Trans., A Phys. Metall. Mater. Sci.* **35**, 1525–1538 (2004).
29. P. Bystricky, thesis, Massachusetts Institute of Technology (1997).

ACKNOWLEDGMENTS

We acknowledge V. Saggiomo (@V_Saggiomo) for advice on microfluidics and F. Ginot for assistance with some of the Python coding. **Funding:** The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 278004. **Author contributions:** S.D. designed and supervised the project; S.D., D.D., and C.M. designed the experiments; D.D. designed and built the cooling stage, prepared the emulsions, and carried out the confocal microscopy; D.D. and S.D. analyzed the data. All authors discussed the results and implications and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability**: Additional data and code related to this paper are available on Figshare at https://doi.org/10.6084/m9.figshare.5746740.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/303/suppl/DC1
Figs. S1 to S6
Movie S1
References (*30*, *31*)

9 November 2017; accepted 26 February 2018
10.1126/science.aar4503

QUANTUM GASES

Recurrences in an isolated quantum many-body system

Bernhard Rauer,[1] Sebastian Erne,[1,2,3]* Thomas Schweigler,[1] Federica Cataldini,[1] Mohammadamin Tajik,[1] Jörg Schmiedmayer[1]†

The complexity of interacting quantum many-body systems leads to exceedingly long recurrence times of the initial quantum state for all but the smallest systems. For large systems, one cannot probe the full quantum state in all its details. Thus, experimentally, recurrences can only be determined on the level of the accessible observables. Realizing a commensurate spectrum of collective excitations in one-dimensional superfluids, we demonstrate recurrences of coherence and long-range order in an interacting quantum many-body system containing thousands of particles. Our findings will enable the study of the coherent dynamics of large quantum systems even after they have reached a transient thermal-like state.

The expectation that a nonequilibrium system evolves toward thermal equilibrium is deeply rooted in our daily experience and is one of the foundations of statistical mechanics (*1*). However, as formulated by Poincaré and Zermelo, a finite isolated physical system will recur arbitrarily close to its initial state after a sufficiently long but finite time (*2*, *3*). The reconciliation of these seemingly contradicting statements is at the heart of the emergence of irreversible processes from reversible microscopic mechanics (*4*).

The above discussion can be transferred to the quantum domain. In analogy to Boltzmann's conjecture, von Neumann formulated a quasi-ergodic theorem for the evolution of the wave function (*5*), and the equilibration of isolated quantum systems grew into an active field of research (*6*). However, a general recurrence theorem can be proven in quantum mechanics as well (*7*, *8*) and shows explicitly that the wave function returns arbitrarily close to its initial state. Experimentally, recurrent behavior has been observed only in small quantum systems. Examples are the revival dynamics in the Jaynes-Cummings model (*9*) or few interacting atoms in isolated sites of an optical lattice (*10*, *11*). These studies accomplished coherent dynamics over long times and thereby revealed properties of the underlying Hamiltonian spectrum. Enabling recurrent behavior in more complex quantum many-body systems opens a window into their long-term coherent evolution by providing a strong measurable signal at times far beyond the initial relaxation. Further, it is of fundamental interest as it provides insight into the emergence of statistical ensembles from coherent unitary evolution.

For larger systems, however, it becomes exponentially difficult to observe the eigenstates directly, and the complexity of the spectrum of eigenstates leads to exceedingly long recurrence times, in general prohibiting their observation. Nevertheless, for a large class of systems, the essential dynamical features can be described by effective field theories with a much simpler structure (*12*, *13*). This reflects the distinction between relevant and irrelevant operators in the microscopic model and sharply reduces the complexity of the problem from a large number of constituents to a much smaller number of populated modes. Within this level of description, quantum recurrences manifest through the return of observables dominated by these modes. Designing the whole system such that the collective excitations follow a simple commensurate spectrum makes the observation of recurrences feasible, even for many-body systems containing thousands of interacting particles.

Ultracold gases (*14*) are an ideal starting point for studying these fundamental phenomena as they can be well isolated from the environment, and excellent tools are available to prepare, manipulate, and probe them.

As a model system, we study coherence in one-dimensional (1D) superfluids. In a first approximation, the corresponding many-body physics can be mapped to an effective low-energy description: a bosonic Luttinger liquid (*15*–*18*) where the collective excitations are free phonons. These phonons are directly related to the phase fluctuations observed in the interference of two 1D superfluids (*19*).

The dephasing of these collective phononic excitations leads to a loss of coherence and long-range order, proceeding in a light-cone–like fashion (*20*–*23*). Whereas the fully dephased state is described by a generalized Gibbs ensemble (*24*), the long time behavior crucially depends on the spectrum of the collective phononic modes. In a harmonic longitudinal confinement with trap frequency ω_z, these phonon frequencies are noncommensurate: $\omega_j = \omega_z\sqrt{j(j+1)/2}$ (*25*), with j being the mode index. If the atoms are confined

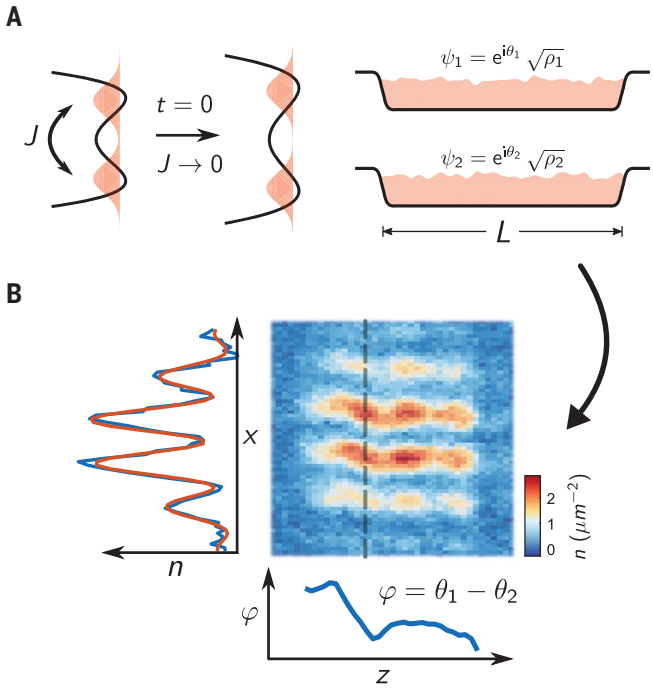


Fig. 1. Schematics of the experiment and the measurement process. (**A**) Two coupled 1D superfluids $\psi_{1(2)}$ in a double-well potential are taken out of equilibrium by a sudden quench of the tunneling coupling J to zero. Along the longitudinal direction z, the system is confined to a box-shaped potential of variable length L. (**B**) The evolution of the decoupled system is measured through matter-wave interference in time-of-flight. A typical interference picture showing the atomic density n is given. A sinusoidal function is fitted to the local interference fringes for each position z. On the left, we show an example of such a fit (red) for one slice (blue), indicated by the dashed line in the image; the x coordinate being the spatial direction of the double-well separation. These fits give access to the spatially resolved relative phase $\varphi(z) = \theta_1(z) - \theta_2(z)$ between the two superfluids (bottom).

[1]Vienna Center for Quantum Science and Technology, Atominstitut, TU Wien, Stadionallee 2, 1020 Vienna, Austria. [2]Institut für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany. [3]Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany. *Present address: School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK. †Corresponding author. Email: schmiedmayer@atomchip.org

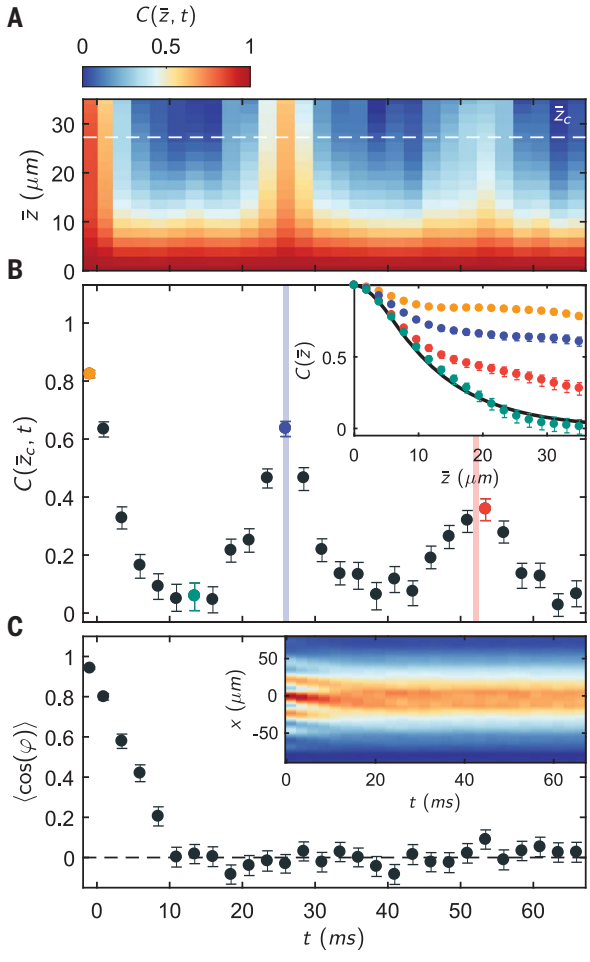


Fig. 2. Dynamics after decoupling. (**A**) Temporal evolution of the phase correlation function $C(\bar{z}, t)$ after decoupling in a 49-μm-long box potential. (**B**) A cut at $\bar{z}_c = 27.3$ μm, with the error bars giving the 68% confidence interval obtained from a bootstrap. The first and second recurrences are clearly visible and occur at the expected times $t_{rec} = L/c$ and $2t_{rec}$, as indicated by the vertical blue and red bars, respectively. The distance $\bar{z}_c$ is chosen such that it is considerably smaller than the size of the system but long enough that the correlations after the initial dephasing are low; this is to facilitate a high recurrence visibility. The inset shows the phase correlation function at the first (blue) and second (red) recurrence and compares them to the correlations in between the recurrences (green) and the correlations in the initial state (orange). The corresponding points in (B) are colored accordingly. A thermal fit to the correlations in between the recurrences is given by the solid black line. (**C**) Evolution of the coherence factor $\langle\cos(\varphi)\rangle$ in the center of the trap at $z = 0$, showing no sign of a recurrence (see main text for details). Owing to the near homogeneity of the system, the behavior is similar over the whole sample. Therefore, in contrast to the phase correlation function, the averaged interference picture shows no recurrent behavior, as shown in the inset (integrated along the z axis).

to a box-shaped trap, the phonon frequencies become commensurate with $\omega_j \propto j$, and recurrences should be observable at short times (*26*).

We implement a boxlike confinement in an atom chip double-well setup (*27*) by adding hard walls to the very weak longitudinal harmonic confinement with the help of a blue-detuned optical dipole potential (*28*). To observe recurrent dynamics, we prepare a thermal equilibrium state in a coupled double-well potential (*29*). Typical samples have a linear density of about 70 atoms per micrometer and, depending on the box size, between 2300 and 4800 atoms in each well, resulting in an interaction energy per atom close to 1 kHz. The tunneling coupling J is tuned to a regime where the phases $\theta_{1,2}$ of the two superfluids lock, creating a system with a strongly correlated relative phase field $\varphi = \theta_1 - \theta_2$, such that $\langle\cos(\varphi)\rangle \approx 1$. To initiate the nonequilibrium dynamics, the coupling is rapidly ramped to zero, leaving the two gases to evolve independently (Fig. 1).

We observe the subsequent dynamics by matter-wave interferometry (*27*), which gives direct access to the spatially resolved relative phase $\varphi(z)$ between the two superfluids. To study the coherence and long-range order, we evaluate the two-point correlation function (*23*)

$$C(\bar{z} = |z - z'|) = \langle\cos(\varphi(z) - \varphi(z'))\rangle \quad (1)$$

with the expectation value taken over many experimental realizations. Further, we average over all points z, z' within the central part of the clouds that fulfill $\bar{z} = |z - z'|$.

A typical temporal evolution of the phase correlations in a box trap of length $L = 49$ μm is shown in Fig. 2A. Before the quench at $t = 0$, the relative phase between the superfluids is locked and correlations are close to unity over the whole sample. Immediately after decoupling, this long-range order decays, reflecting the dephasing of the collective excitations. At the first minimum in Fig. 2B, the initial state is completely dephased, and the system is indistinguishable from a thermal state (see inset). For a system with incommensurately spaced modes, this dephased state persists for a long time, showing the emergence of statistical properties from the unitary quantum evolution (*22*). In our system, however, mode frequencies are designed to be commensurate, and two partial recurrences of phase coherence are clearly visible in the subsequent evolution.

To understand the recurrence time, we need to look at the dispersion relation of excitations. For a perfect hard-wall box confinement, $\omega_j = \frac{c\pi}{L} j$, where c is the speed of sound and j is the mode index (*28*). These modes are not only commensurate but also equally spaced, facilitating a recurrence at the earliest possible time $t = \frac{2\pi}{\Delta\omega} = \frac{2L}{c}$, with $\Delta\omega$ being the energy spacing between the modes. At this time, the lowest-lying mode has finished a full rotation whereas the higher-energy modes have all performed an integer number of turns, bringing all excitations back to their initial configuration. Half-way to this full recurrence, the system rephases to the mirrored initial state. As we initially start from a nearly flat relative phase profile and our observable C (Eq. 1) is insensitive to the transformation $\varphi(z) \rightarrow \varphi(-z)$, this point is equivalent to the full recurrence. Therefore, the expected recurrence time for the correlations is given by $t_{rec} = \frac{L}{c}$ (blue and red bars in Fig. 2B), which agrees well with the observed peaks in coherence.

Although two-point correlations return close to the initial state, the relative phase field does not recur. The time evolution of the coherence factor $\langle\cos(\varphi)\rangle$ shows no recurrence (Fig. 2C). It relaxes during the initial dephasing dynamics and stays close to zero from then on. This is equivalent to observing that the ensemble-averaged interference picture shows no revival of high-contrast fringes (Fig. 2C, inset). At the time of the recurrences, the interference fringes return close to their straight initial state, displaying long-range coherence throughout the system; however, for each distinct realization, they are shifted by a random global phase. The reason behind this global phase is a small random atom number imbalance between the two wells. This imbalance originates from the thermal fluctuations of the initial state, imperfections in the experiment, and quantum noise relevant for lower temperatures. It leads to an inevitable population of the $k = 0$ mode (*30*), resulting in a global phase accumulation that is different for each realization, and therefore to vanishing interference contrast in the ensemble average. In contrast, the phase correlations (Eq. 1) are insensitive to a global offset of the phase $\varphi(z)$, and the recurrences of excitations on top of the field can be observed. This illustrates that recurrent behavior can be hidden below a global phase diffusion, which necessitates the measurement of, at least, two-point correlations to reveal the underlying coherent dynamics.

To confirm the scaling of the recurrence time with the size of the system, we vary the length of the box potential by changing the position of the dipole trap walls. As the system size is increased, the recurrence is shifted to later times (Fig. 3A). For this comparison, the time axis was rescaled by the theoretical prediction for the speed of sound (*28*) to make measurements with slightly different atomic densities comparable. Extracting the exact times of the first and second recurrence by fitting

the peaks in the data (*28*) reveals the linear scaling with L (Fig. 3B).

Although the first two recurrences are clearly visible, for all these measurements, the height of the recurrences is rapidly damped (see Fig. 2, A and B), and observing a third recurrence becomes infeasible in most cases. To probe the recurrence decay in more detail, we studied the evolution for different initial temperatures in the $L = 49$ μm trap. Increasing the initial temperature, the height of the observed recurrence with respect to the correlations of the initial state decreases rapidly (Fig. 4). The temperatures for this analysis are extracted from the full distribution functions of interference contrasts (*31*) for the completely dephased state in between the recurrences (*28*).

To understand the origin of this damping with temperature, we considered different theoretical descriptions of our system. We first investigated the low-energy effective description by solving the Luttinger liquid Hamiltonian (dashed lines in Fig. 4). This model describes the free propagation of phononic excitations on top of a stationary background density. For a homogeneous background, it would give perfect recurrences of phase coherence. For the comparison to the experimental data, we consider an inhomogeneous background density that reflects our boxlike potential. In addition, we took into account the typical spread in total particle number and particle number imbalance between the wells, measured independently for the respective samples (*28*). The shot-to-shot fluctuations of the speed of sound induced by this spread together with the inhomogeneous background density constitute the only source of damping within the model. Although the recurrence times are well described, as seen in Fig. 3B, the damping in this description is too weak to explain the experimental observations even for the lowest temperatures.

As a second model (solid lines in Fig. 4), we numerically simulated the dynamics using the Gross-Pitaevskii equation (GPE) for finite-temperature initial states (*28*). This description goes beyond the free phononic excitations of the Luttinger liquid and takes interactions between these quasiparticles into account. It agrees well with our experimental findings (Fig. 4). Together with the failure of the Luttinger liquid model, this agreement shows that physics beyond the low-energy description becomes relevant and indicates that phonon-phonon interactions are responsible for the observed damping. As the temperature is increased, these processes become more important as higher-energy modes are populated.

This study illustrates that the observation of recurrences enables insight into the nonequilibrium dynamics of quantum many-body systems (*32*). Their shape and magnitude are sensitive probes into coherent dynamics far beyond the time scales of the initial relaxation. This feature makes them a versatile tool for the verification of coherence in quantum simulators and to test the regimes of validity for approximate models. Moreover, the long coherent evolution probed through recurrences magnifies the effects of small perturbations whose

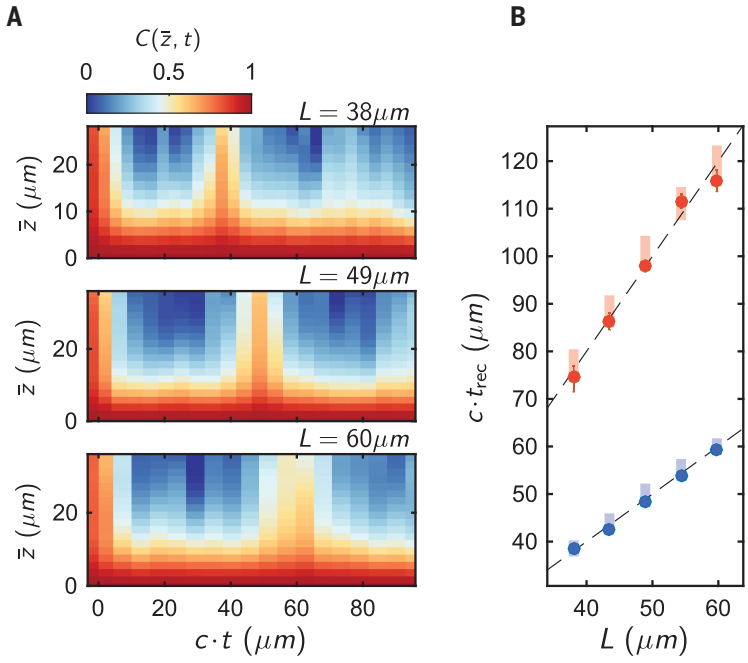


Fig. 3. Comparison of the recurrences for different box lengths. (**A**) Phase correlations for three different box lengths $L = 38$, 49, and 60 μm (top to bottom). The time axis is rescaled with the theoretical prediction for the speed of sound c. (**B**) Recurrence time over the box length extracted from the phase correlations at $\bar{z}_c = 27.3$ μm (*28*). For each box length, the time of the first (blue) and second (red) recurrence is plotted and compared to the ideal linear scaling (dashed lines). The error bars give the 68% confidence interval obtained from a bootstrap, whereas the shaded bars indicate the predictions of the Luttinger liquid model for our particular trap (*28*). The vertical extension of the bars corresponds to the uncertainty of the decoupling time, whereas the horizontal extension is chosen arbitrarily.

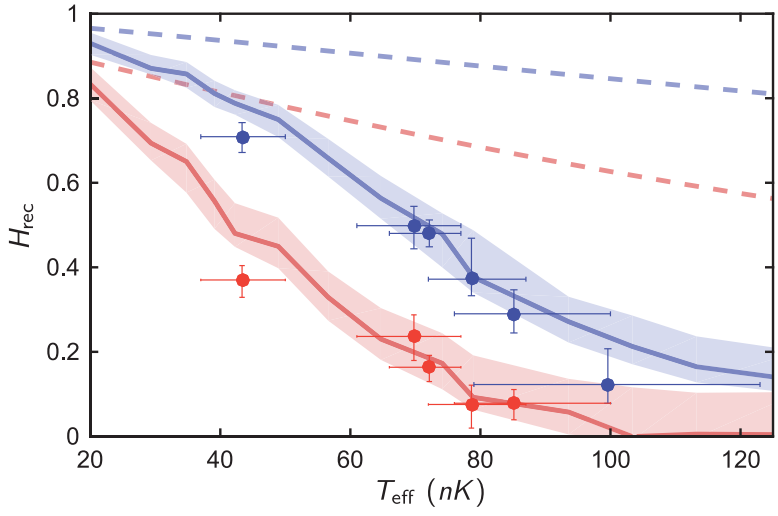


Fig. 4. Temperature dependence of the recurrence height. The data points represent measurements of the height H_{rec} of the first (blue) and second (red) recurrence in a box of length $L = 49$ μm for different effective temperatures T_{eff}. The height is extracted from fitting the peaks of the correlation function at $\bar{z}_c = 27.3$ μm and the error bars give the 68% confidence interval obtained from a bootstrap. A fit of the full distribution function of measured interference contrasts at $t = t_{rec}/2$ gives the effective temperature. The solid lines are results of GPE simulations analyzed in the same way as the experiment. The shaded area indicates the uncertainty caused by the limited experimental statistics (1σ deviation). The dashed lines are the predictions of the Luttinger liquid model for the experimental parameters (*28*). The experimental shot-to-shot fluctuations of the atom number as well as the inhomogeneous confinement are incorporated in both theoretical calculations and constitute the only source of damping in the Luttinger liquid model.

influence on the short-term dynamics is negligible. In our model case, the decay of recurrences clearly shows that physics beyond the Luttinger liquid model sets in rather fast. A combined study looking at recurrences and mode occupations will shed light onto the fundamental quantum processes in the relaxation of 1D systems.

REFERENCES AND NOTES

1. L. Boltzmann, *Sitz.-Ber. Akad. Wiss. Wien (II)* **66**, 275 (1872).
2. H. Poincaré, *Acta Math.* **13**, 1 (1890).
3. E. Zermelo, *Ann. Phys.* **293**, 485–494 (1896).
4. D. ter Haar, *Rev. Mod. Phys.* **27**, 289 (1955).
5. J. von Neumann, *Z. Phys.* **57**, 30 (1929).
6. C. Gogolin, J. Eisert, *Rep. Prog. Phys.* **79**, 056001 (2016).
7. P. Bocchieri, A. Loinger, *Phys. Rev.* **107**, 337–338 (1957).
8. I. C. Percival, *J. Math. Phys.* **2**, 235–239 (1961).
9. G. Rempe, H. Walther, N. Klein, *Phys. Rev. Lett.* **58**, 353–356 (1987).
10. M. Greiner, O. Mandel, T. W. Hänsch, I. Bloch, *Nature* **419**, 51–54 (2002).
11. S. Will *et al.*, *Nature* **465**, 197–201 (2010).
12. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, International Series of Monographs on Physics (Clarendon, Oxford, ed. 4, 2002).
13. T. Schweigler *et al.*, *Nature* **545**, 323–326 (2017).
14. N. P. Proukakis, D. W. Snoke, P. B. Littlewood, *Universal Themes of Bose-Einstein Condensation* (Cambridge Univ. Press, Cambridge, 2017).
15. S.-I. Tomonaga, *Prog. Theor. Phys.* **5**, 544–569 (1950).
16. J. M. Luttinger, *J. Math. Phys.* **4**, 1154–1162 (1963).
17. D. C. Mattis, E. H. Lieb, *J. Math. Phys.* **6**, 304–312 (1965).
18. T. Giamarchi, *Quantum Physics in One Dimension* (Clarendon, Oxford, 2004).
19. A. Polkovnikov, E. Altman, E. Demler, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6125–6129 (2006).
20. R. Bistritzer, E. Altman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9955–9959 (2007).
21. T. Kitagawa, A. Imambekov, J. Schmiedmayer, E. Demler, *New J. Phys.* **13**, 073018 (2011).
22. M. Gring *et al.*, *Science* **337**, 1318–1322 (2012).
23. T. Langen, R. Geiger, M. Kuhnert, B. Rauer, J. Schmiedmayer, *Nat. Phys.* **9**, 640–643 (2013).
24. T. Langen *et al.*, *Science* **348**, 207–211 (2015).
25. D. S. Petrov, G. V. Shlyapnikov, J. T. M. Walraven, *Phys. Rev. Lett.* **85**, 3745–3749 (2000).
26. R. Geiger, T. Langen, I. E. Mazets, J. Schmiedmayer, *New J. Phys.* **16**, 053034 (2014).
27. T. Schumm *et al.*, *Nat. Phys.* **1**, 57–62 (2005).
28. See supplementary materials.
29. T. Betz *et al.*, *Phys. Rev. Lett.* **106**, 020407 (2011).
30. M. Lewenstein, L. You, *Phys. Rev. Lett.* **77**, 3489–3493 (1996).
31. S. Hofferberth *et al.*, *Nat. Phys.* **4**, 489–495 (2008).
32. J. Cardy, *Phys. Rev. Lett.* **112**, 220401 (2014).

ACKNOWLEDGMENTS

We acknowledge discussions with I. Mazets, J. Berges, and T. Gasenzer. **Funding:** This work was supported by the European Union through the EU-FET Proactive grant AQuS, Project no. 640800, and the European Research Council advanced grant QuantumRelax, by the J. Templeton foundation Project no. 60478 and by the Austrian Science Fund (FWF) through the doctoral program CoQuS (W1210) (B.R., T.S., F.C.). This work is supported by the SFB 1225 'ISOQUANT' financed by the German Research Foundation (DFG) and the FWF. **Author contributions:** B.R., T.S., F.C., and M.T. conducted the experiments and the data analysis. S.E. and B.R. performed the theoretical calculations. J.S. designed the experiment and supervised the project. All authors contributed to interpreting the data and writing the manuscript. **Competing interests:** None declared. **Data and materials availability:** The data presented in this paper are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/307/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*33–50*)
Data File S1

23 May 2017; accepted 13 February 2018
Published online 22 February 2018
10.1126/science.aan7938

EXTINCTION

Body size downgrading of mammals over the late Quaternary

Felisa A. Smith,[1]* Rosemary E. Elliott Smith,[2] S. Kathleen Lyons,[3] Jonathan L. Payne[4]

Since the late Pleistocene, large-bodied mammals have been extirpated from much of Earth. Although all habitable continents once harbored giant mammals, the few remaining species are largely confined to Africa. This decline is coincident with the global expansion of hominins over the late Quaternary. Here, we quantify mammalian extinction selectivity, continental body size distributions, and taxonomic diversity over five time periods spanning the past 125,000 years and stretching approximately 200 years into the future. We demonstrate that size-selective extinction was already under way in the oldest interval and occurred on all continents, within all trophic modes, and across all time intervals. Moreover, the degree of selectivity was unprecedented in 65 million years of mammalian evolution. The distinctive selectivity signature implicates hominin activity as a primary driver of taxonomic losses and ecosystem homogenization. Because megafauna have a disproportionate influence on ecosystem structure and function, past and present body size downgrading is reshaping Earth's biosphere.

Wild mammals are in decline globally because of a lethal combination of human-mediated threats, including hunting, introduced predators, and habitat modification (*1*–*5*). Extinction risk is particularly acute for the largest mammals, which are more frequently in conflict with humans (*1*, *6*). The ongoing extirpation of large-bodied mammals is a major conservation concern because their decline can lead to the loss of ecological function within communities (*3*, *5*, *7*). Megafauna have crucial direct and indirect impacts on vegetation structure, biogeochemical cycling, ecological interactions, and climate (*7*–*10*). Although the current extinction rate is higher than earlier in the Cenozoic (*4*), the ongoing biodiversity crisis may be an acceleration of a long-term trend over the late Quaternary. For example, a striking feature of the Pleistocene was the abundance and diversity of extremely large mammals such as the mammoth, giant ground sloth, wooly rhinoceros, and sabretooth tiger on all habitable continents. The debate about the causes of the terminal Pleistocene megafauna extinction has been long and acrimonious, with particular controversy surrounding the role of humans (*11*–*13*).

Multiple hominins—including at a minimum Neandertals, Denisovans, and archaic/modern humans—have been part of ecosystems throughout the late Pleistocene. Genetic analyses reveal a complicated history, with substantial admixture between populations (*14*). Anthropologists remain divided about the routes, exact timing, and number of early migrations from Africa (*14*–*18*), but several hominin species were probably widespread across Africa and Eurasia around 80 thousand to 60 thousand years (ka) ago (*15*–*17*). Further expansion followed, with modern *Homo sapiens* reaching Australia ~60 to 50 ka ago and crossing into the Americas ~15 to 13 ka ago (*15*). Migrations were likely driven or facilitated by climatic factors (*17*, *18*) and were followed by rapid increases in population sizes (*17*, *19*). For example, hominin populations in western Europe increased 10-fold by the Neandertal-to-Modern human transition ~40 ka ago (*19*). Middle to Upper Paleolithic hominins were hunters who lived in groups and used both tools and fire (*20*); thus, it is plausible that their activities and rapid population growth influenced mammal biodiversity well before the terminal Pleistocene.

We investigated the influence of these emerging and increasingly sophisticated hominin predators on continental and global mammalian biodiversity over the late Quaternary (*21*). Ongoing biodiversity loss is robustly linked to human activities (*1*–*5*); and previous work linked extinction risk over the Holocene, terminal Pleistocene, and end-Pleistocene to human activities (*4*, *6*, *11*–*13*, *22*–*25*); but earlier influences remain poorly characterized. Although recent work on paleodemography exists for *H. sapiens* over the late Pleistocene and Holocene (*17*), a lack of data for other hominins precludes direct comparison of mammalian extinction risk over time against hominin population density. However, should we find significant differences between the pattern of late Quaternary extinction selectivity and the rest of the Cenozoic mammal record, this would strongly suggest a role of hominin activity (*13*, *24*, *25*).

We used two data sets to test the potential role of hominin activity on extinction selectivity, mammalian body size distributions, and patterns of biodiversity over time and into the future (*21*). First, we updated a spatially explicit global record of body size and trophic mode for nonvolant, terrestrial mammals for the late Quaternary (MOM). Second, we constructed a global data set of Cenozoic mammals with associated stratigraphic duration, body mass, and trophic mode. We categorized late Quaternary extinctions into five temporal bins: late Pleistocene (125 to 70 ka ago), which corresponded with the initial waves of migration of hominins out of Africa; end Pleistocene (70 to 20 ka ago), which represented the continued expansion of hominins into Eurasia and the colonization of Australasia; terminal Pleistocene (20 to 10 ka ago), which encompassed the migration of humans into the Americas; Holocene (10 to 0 ka ago), which represented further expansion of humans throughout the globe; and future (~+0.2 ka), where we assumed that all currently threatened mammals become extinct (*21*). We binned the Cenozoic fossil data set into intervals of 1 million years (Ma) as a reference standard and computed temperature metrics for each bin (*21*). For each time interval, we characterized the size selectivity of extinction using logistic regression and examined overall body size distribution and trophic guild structure (tables S1 to S7) (*21*). For the late Quaternary, we also characterized size selectivity by continent and trophic level.

Our analyses demonstrated a striking and significantly size-biased pattern of mammalian extinction over the late Quaternary, distinct in the Cenozoic record (Figs. 1 to 3 and fig. S1). We found a mass difference of two to three orders of magnitude between victims and survivors of late Quaternary extinction intervals (Fig. 2A and table S1), reflecting a significant association between size and extinction probability (Fig. 2B and table S5). This size bias occurred on each continent (Fig. 2, C and D) and within each major trophic group (Fig. 2, E and F), with the magnitude of the size difference and the statistical measure of size selectivity decreasing between the Pleistocene and Holocene (Fig. 2, A to F). The reduced selectivity of the Holocene and future extinctions likely reflects changes in the nature of threats. Today, many smaller-bodied animals are vulnerable because of habitat alteration, introduced predators, or urbanization (*5*–*7*, *11*, *26*).

Comparison of extinctions across the entirety of the Cenozoic demonstrated that body mass was rarely significantly associated with the probability of extinction before the late Pleistocene (Figs. 1 and 3, E and F), and further, size differences between victims and survivors never approached those observed in the Pleistocene (tables S1 and S3). There was a preferential loss of small-bodied species in the Oligocene that is perhaps linked to expansion of grasslands and prairies (~29 Ma ago) (Fig. 3E), although this value had high uncertainty. However, no interval over the past 65 Ma was as selective as the late Quaternary. Moreover, climate change did not increase extinction risk for large-bodied mammals before the spread of hominins. We found no relationship between temperature change over the Cenozoic and size bias of extinction; neither small nor large mammals were more vulnerable to extinction during times of high climate variability (table S3 and fig. S4). The probability that the late Pleistocene and Cenozoic selectivity

[1]Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA. [2]Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA. [3]School of Biological Sciences, University of Nebraska–Lincoln, Lincoln, NE 68588, USA. [4]Department of Geological Sciences, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: fasmith@unm.edu

coefficients came from the same distribution was very low ($P < 0.001$), given either log likelihood or nonparametric tests (fig. S3). Moreover, grouped as a single extinction event (as they would appear to a future paleontologist), the Quaternary extinction pulse was by far the most selective episode of extinction in the Cenozoic (Fig. 1). Such pronounced size selectivity is highly unusual in other fossil records; larger-bodied vertebrates and mollusks did not experience increased extinction risk over the Cenozoic or during the five mass extinction events (*27*). Because a reported signature of human hunting is size selectivity (*24*, *25*), our results are consistent with the hypothesis that hominin activities contributed to extinctions long before the terminal Pleistocene.

The late Quaternary biodiversity losses led to dramatic, time-transgressive shifts in both mean and maximum body mass on each continent (Fig. 3), which followed hominin dispersal patterns (*15*) and began much earlier than previously suspected. Because body size distributions are related to the size of the landmass (*28*), the largest average or maximum body mass would be expected on Eurasia, followed by Africa, then North and South America, and the smallest on Australia. This expectation was largely met in the late Pleistocene (Fig. 3), but Africa was a notable outlier, with a mean body mass ~50% less than that of Eurasia or the Americas before 125 ka ago (table S1). We hypothesize that the late Pleistocene size distribution in Africa reflects the long prehistory of hominin-mammal interactions (*29*). This finding suggests that the homogenization of natural ecosystems was a consequence of hominin behavior in general and not specific to *H. sapiens*. Over the following ~100 ka, mean body mass dropped dramatically—first by 50% in Eurasia, and then by an order of magnitude in Australia—while remaining largely unchanged in the Americas until the terminal Pleistocene. Thus, for most of the late Quaternary, mean and maximum body masses were larger in the Americas than elsewhere—a pattern largely exceptional in the mammalian fossil record (Fig. 3 and table S1) (*28*). By the terminal Pleistocene, other hominin species were extinct, and the remaining *H. sapiens* had developed efficient long-range weapons (*11*). The latter likely contributed to the severity of the extinction in the New World (Fig. 1), with 11.5 and 9.7% of nonvolant terrestrial species lost in North and South America, respectively (tables S1 and S2). The loss of biodiversity resulted in a greater than 10-fold drop in both mean and maximum body mass, which was a steeper decline than elsewhere (Fig. 3, B and D). For example, mean mass of nonvolant terrestrial mammals in North America fell from 98.0 to 7.6 kg (table S1).

Future extinctions will continue the pattern of biodiversity loss and body size downgrading (fig. S1). If all species currently at risk are eventually driven extinct, ~22.4 to 53.7% of mammals will be lost relative to 125 ka ago (table S2). This will further decrease mean body mass in North America from 7.7 to 4.9 kg (Fig. 3B and table S1); similar declines are predicted for other continents. Thus, the largest mammal on earth in a few hundred

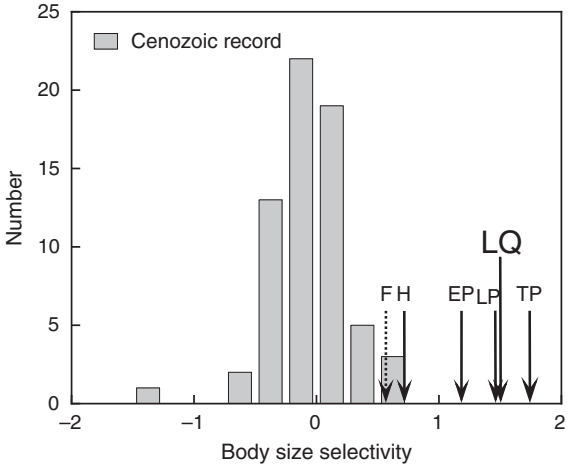


Fig. 1. Distribution of body size selectivity coefficients over the Cenozoic mammal record. All selectivity coefficients reflect change in the natural logarithm of the odds of extinction associated with a one-$\log_{10}$-unit change in body mass. Values of zero indicate no bias, positive values indicate bias against larger size, and negative values indicate bias against smaller size. LQ, average of all late Quaternary (LP to H) extinctions; LP, late Pleistocene; EP, end Pleistocene; TP, terminal Pleistocene; H, Holocene; and F, future extinctions.

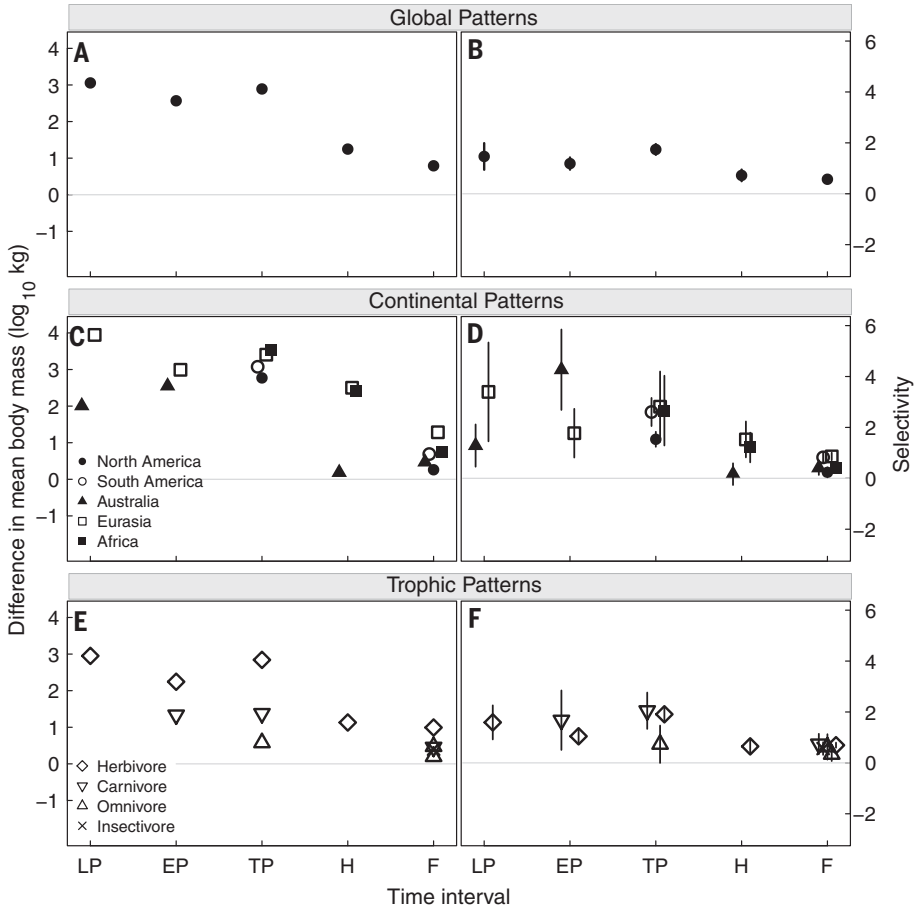


Fig. 2. Analyses of size bias in the mammalian fossil record. (**A** and **B**) Global patterns of extinction. (A) Difference in the mean of log-transformed sizes of victims versus survivors for intervals across the late Quaternary. (B) Selectivity coefficients measuring the association between body size and extinction probability derived from logistic regression of extinction status as a function of body mass. Multiple regressions controlling for the additive contributions of continental location and trophic guild yield even stronger associations between extinction status and body mass (table S5) (*21*). (**C** and **D**) Extinction patterns on each continent. (C) Size differences. (D) Size selectivity coefficients. (**E** and **F**) Influence of trophic guild on extinction risk. (E) Size differences. (F) Size selectivity coefficients. Bars indicate 95% confidence interval (CI).

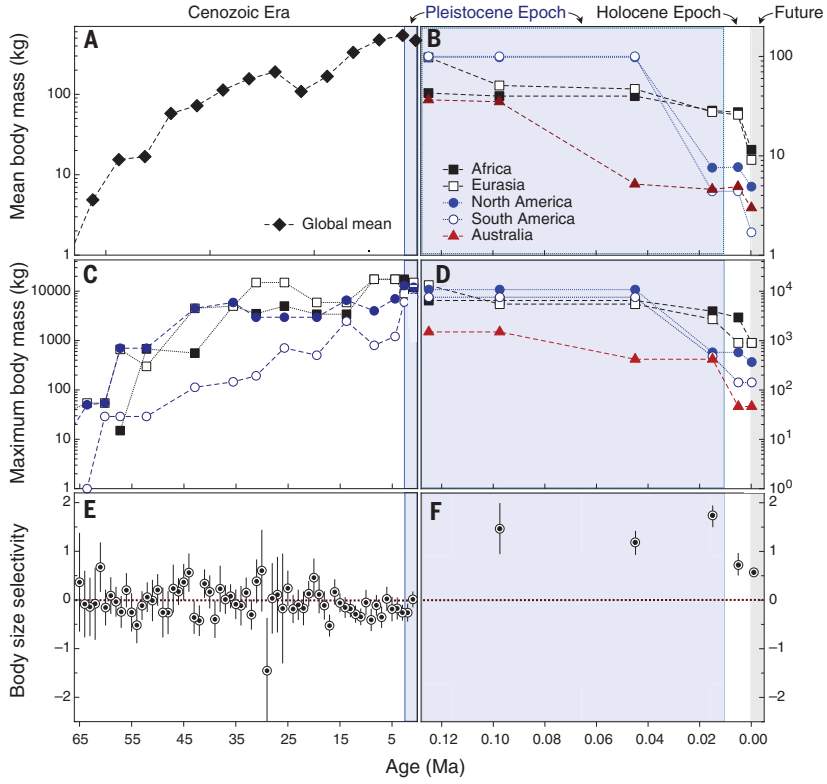


Fig. 3. Body size and its influence on extinction risk. (**A**) Global mean body size over the Cenozoic (65 to 1 Ma ago). (**B**) Mean body size by continent over the late Quaternary (past 125 ka). (**C**) Maximum body size across the Cenozoic by continent. (**D**) Maximum body size over the late Quaternary and into the future. (**E**) Size selectivity coefficients across the entire Cenozoic fossil record. (**F**) Size selectivity of late Quaternary extinctions. Bars indicate 95% CI. All masses are in kilograms. Light blue shading indicates late Pleistocene, white shading indicates Holocene, and gray shading indicates the future (+200 years). Ages here and elsewhere are plotted as midpoint of time interval.

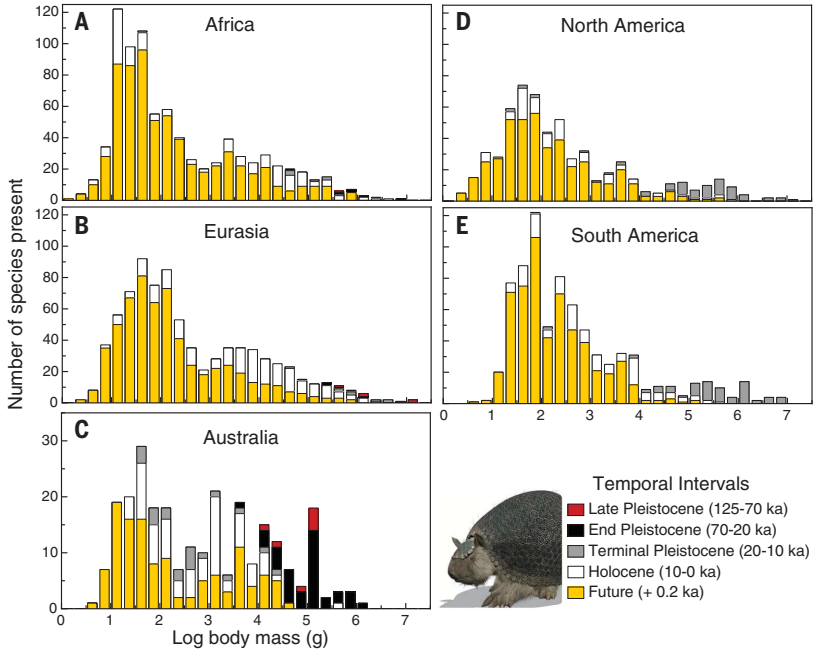


Fig. 4. The body size distribution of terrestrial, nonvolant mammals on each continent over the late Quaternary. (**A**) Africa. (**B**) Eurasia. (**C**) Australia. (**D**) North America. (**E**) South America. Body sizes for each temporal interval are plotted; distributions are overlaid from oldest to youngest. Yellow shading indicates the predicted distribution in the future, if vulnerable species go extinct.

years may well be a domestic cow (*Bos taurus*) at ~900 kg. Furthermore, the loss of currently endangered species would reduce terrestrial mammal body mass to the lowest values in the past 45 Ma (Fig. 3, A and C, as compared with Fig. 3, B and D). The last time the body size distribution of terrestrial vertebrates was similarly disrupted was ~66 Ma ago, during the end-Cretaceous mass extinction.

Because body size is strongly linked to most biological rates and processes (*30*), the extirpation of large mammals led to a fundamental restructuring of energy flow through mammal communities over the late Quaternary. The severe body size downgrading—a truncation of more than two orders of magnitude—resulted in substantial shifts from bimodal toward unimodal size distributions (Fig. 4 and fig. S1). Homogenization of distributions continued through the Holocene and is predicted to continue into the future (Fig. 4 and table S4). Extinctions also led to changes in the proportional representation of trophic guilds, especially herbivores (fig. S2). In the future, continental distributions will be severely skewed toward smaller mammals (Fig. 4)—in particular, rodents (fig. S2). Ecological principles suggest that changes in energy flow over the Pleistocene likely led to compensatory changes, potentially numerical responses by surviving smaller-bodied mammals to maintain ecosystem homeostasis (*31*). By the Holocene, however, humans were a strong influence on energy flow within ecosystems. Global expansion was accompanied by increased human densities (*17*) and animal domestication (*10*). By historical time, the terrestrial biosphere was transformed from one dominated by wild animals into one dominated by humans and their livestock, many provisioned with domesticated crops (*2*, *5*, *10*). Today, the biomass of the >4.5 billion domesticated animals on Earth exceeds estimates for wild mammals at the terminal Pleistocene (*10*).

Our study highlights the long and sustained influence of humans and other hominins on terrestrial ecosystems. As Neandertals, Denisovans, and humans spread across the globe over the late Quaternary, a highly size-biased extinction followed, a pattern distinct in the Cenozoic mammal record. The subsequent downgrading of body size was severe and differentially targeted herbivores. Thus, contemporary biodiversity loss is part of a trend spanning more than 125 ka, with expected future extinctions of greater magnitude, but reduced size selectivity, than in the past. The homogenization of ecosystems has dramatically influenced the past, present, and future role of wild mammals in the terrestrial biosphere.

REFERENCES AND NOTES

1. M. Cardillo *et al.*, *Science* **309**, 1239–1241 (2005).
2. A. D. Barnosky *et al.*, *Nature* **471**, 51–57 (2011).
3. J. A. Estes *et al.*, *Science* **333**, 301–306 (2011).
4. G. Ceballos *et al.*, *Sci. Adv.* **1**, e1400253 (2015).
5. R. Dirzo *et al.*, *Science* **345**, 401–406 (2014).
6. World Wildlife Fund, *Living Planet Report 2016. Risk and Resilience in a New Era* (Gland, 2016).
7. F. A. Smith, C. E. Doughty, Y. Malhi, J.-C. Svenning, J. Terborgh, *Ecography* **39**, 99–108 (2016).

8. C. N. Johnson, *Proc. Biol. Sci.* **276**, 2509–2519 (2009).
9. C. E. Doughty *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 868–873 (2016).
10. F. A. Smith *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 874–879 (2016).
11. S. K. Lyons *et al*., *Evol. Ecol. Res.* **6**, 339–358 (2004).
12. P. L. Koch, A. D. Barnosky, *Annu. Rev. Ecol. Syst.* **37**, 215–250 (2006).
13. W. Zuo, F. A. Smith, E. L. Charnov, *Am. Nat.* **182**, 524–531 (2013).
14. R. Nielsen *et al*., *Nature* **541**, 302–310 (2017).
15. S. Oppenheimer, *Philos. Trans. R. Soc. London B Biol. Sci.* **367**, 770–784 (2012).
16. Q. D. Atkinson, R. D. Gray, A. J. Drummond, *Mol. Biol. Evol.* **25**, 468–474 (2008).
17. A. Timmermann, T. Friedrich, *Nature* **538**, 92–95 (2016).
18. S. L. Carto, A. J. Weaver, R. Hetherington, Y. Lam, E. C. Wiebe, *J. Hum. Evol.* **56**, 139–151 (2009).
19. P. Mellars, J. C. French, *Science* **333**, 623–627 (2011).
20. W. Roebroeks, P. Villa, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5209–5214 (2011).
21. Materials and methods are available as supplementary materials.
22. D. A. Burney, T. F. Flannery, *Trends Ecol. Evol.* **20**, 395–401 (2005).
23. S. T. Turvey, Ed., *Holocene Extinctions* (Oxford Univ. Press, 2009).
24. T. Surovell, N. M. Waguespack, in *American Megafaunal Extinctions at the End of the Pleistocene*, G. Hayes, Ed. (Springer, 2009), pp. 77–105.
25. J. Alroy, *Science* **292**, 1893–1896 (2001).
26. R. E. Plotnick, F. A. Smith, S. K. Lyons, *Ecol. Lett.* **19**, 546–553 (2016).
27. J. L. Payne, A. M. Bush, N. A. Heim, M. L. Knope, D. J. McCauley, *Science* **353**, 1284–1286 (2016).
28. F. A. Smith *et al*., *Science* **330**, 1216–1219 (2010).
29. T. Surovell, N. Waguespack, P. J. Brantingham, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6231–6236 (2005).
30. R. H. Peters, *The Ecological Implications of Body Size* (Cambridge Univ. Press, 1983).
31. S. K. M. Ernest, J. H. Brown, *Ecology* **82**, 2118–2132 (2001).

ACKNOWLEDGMENTS

We thank the Smith laboratory for its enthusiasm and assistance and P. J. Wagner for comparing distributions of selectivity coefficients. **Funding:** Funding was provided by NSF grants DEB-1555525, DEB-1257625, and EAR-1151022. **Author contributions:** F.A.S., R.E.E.S., and S.K.L. compiled and analyzed data; J.L.P. performed additional analyses; F.A.S. wrote the first draft; and all coauthors provided comments. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/310/suppl/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S7
References (*32*–*197*)

7 August 2017; accepted 7 March 2018
10.1126/science.aao5987

NEURODEVELOPMENT

Synaptic transmission from subplate neurons controls radial migration of neocortical neurons

Chiaki Ohtaka-Maruyama,[1]* Mayumi Okamoto,[4] Kentaro Endo,[2] Minori Oshima,[1,5] Noe Kaneko,[1,5] Kei Yura,[6,7] Haruo Okado,[3] Takaki Miyata,[4] Nobuaki Maeda[1]*

The neocortex exhibits a six-layered structure that is formed by radial migration of excitatory neurons, for which the multipolar-to-bipolar transition of immature migrating multipolar neurons is required. Here, we report that subplate neurons, one of the first neuron types born in the neocortex, manage the multipolar-to-bipolar transition of migrating neurons. By histochemical, imaging, and microarray analyses on the mouse embryonic cortex, we found that subplate neurons extend neurites toward the ventricular side of the subplate and form transient glutamatergic synapses on the multipolar neurons just below the subplate. NMDAR (*N*-methyl-D-aspartate receptor)–mediated synaptic transmission from subplate neurons to multipolar neurons induces the multipolar-to-bipolar transition, leading to a change in migration mode from slow multipolar migration to faster radial glial-guided locomotion. Our data suggested that transient synapses formed on early immature neurons regulate radial migration.

The six-layered neocortex develops through radial migration of excitatory neurons in an inside-out manner (*1*, *2*). Defects in this construction process lead to brain disorders such as lissencephaly (*3*). In the developing neocortex, excitatory neurons are generated from progenitors known as radial glial cells in the ventricular zone (VZ) (Fig. 1A). Newly generated neurons exhibit a multipolar shape and meandering movement in the subventricular zone (SVZ) and intermediate zone (IZ) (multipolar migration) (*2*). These multipolar neurons (MpNs) then change to a bipolar shape (multipolar-to-bipolar transition) and migrate rapidly in the cortical plate (CP) toward the marginal zone (MZ) via radial glial-guided locomotion (locomotion). Subplate neurons (SpNs) and Cajal-Retzius cells are the first neurons born in the neocortex and reside in the subplate (SP) and MZ, respectively (*2*, *4*). Cajal-Retzius cells guide the inside-out layering of excitatory neurons by secreting reelin (*5*). SpNs build connections between the thalamus and the cortex (*4*). Here, we examined the possibility that SpNs also regulate radial migration.

To evaluate the roles of SpNs in radial migration, we observed the migration of green fluorescent protein (GFP)–labeled neurons by time-lapse recording of the slices from embryonic day 16 (E16) cortices (movie S1). We noticed that MpNs paused migration at the boundary between the IZ and the SP (Fig. 1A and fig. S1). After this stationary period, MpNs changed their morphology to a bipolar shape and began locomotion. Morphometric analysis revealed that the SP demarcated the boundary between the upper cortical region where bipolar neurons were distributed and the lower region where MpNs were distributed (fig. S2). In the SP, migrating neurons exhibited branched and twisted leading processes with features intermediate between bipolar and multipolar neurons (Fig. 1B). These observations raised the possibility that the SP plays roles in the multipolar-to-bipolar transition of migrating neurons.

The SP consists mainly of SpNs, developing axon tracts, and extracellular matrix (*6*). We hypothesized that MpNs receive signals from SpNs, which induce changes in their polarity. To investigate this possibility, we used two methods to visualize SpNs and migrating neurons separately in the same cortical slices: (i) in utero double electroporation and (ii) in utero single electroporation using Lpar1-EGFP (enhanced GFP) mice (*7*). In the VZ, SpNs are generated at E10 as one of the first-born neurons, and then excitatory neurons are generated in a sequential order. Therefore, SpNs and migrating neurons can be labeled separately by sequential in utero electroporations of GFP expression plasmid at E10 and red fluorescent protein (RFP) expression plasmid at E14, respectively, to the same embryos. In addition, various expression constructs can be introduced selectively into SpNs by in utero electroporation at E10 (fig. S3). In the embryonic cortex of the Lpar1-EGFP mouse, subsets of SpNs express EGFP (Fig. 1C). Thus, in utero electroporation of RFP expression plasmid into the cortex of this Tg mouse at E14 enables us to visualize GFP-positive SpNs and RFP-positive migrating neurons in the same slices.

Live imaging of the SpNs of Lpar1-EGFP mice revealed that they extended axon-like processes toward the ventricular side (Fig. 1D and movie S2). RFP labeling of migrating neurons indicated that these processes are closely associated with MpNs just below the SP (Fig. 1E and movies S3 and S4). The SpN processes had many varicosities, which were reminiscent of en passant synaptic boutons and often attached to the cell bodies of MpNs (Fig. 1, F and G). Similar contacts between SpNs and MpNs were observed in the in utero double-electroporated cortices (fig. S4). Immunohistochemical analysis showed that a part of the SpN neurites contacting with MpNs were vesicular glutamate transporter (VGLUT2)–positive, suggesting that they formed glutamatergic synapses (Fig. 2A). On average, 19 VGLUT2-positive puncta were detected per MpN (fig. S5). By using electron microscopy, we found synapse-like contacts on the cell bodies of MpNs in the upper IZ, which were characterized by a few vesicles of 43 ± 1 nm diameter, as well as electron dense structures reminiscent of active zones and postsynaptic densities (Fig. 2, B to E, and fig. S6). About 70 to 80% of the junctional structures around MpNs showed no vesicles, which are presumably immature or degenerating synaptic contacts (Fig. 2, F and G). The pre- and postsynaptic plasma membranes often showed a wavy shape (Fig. 2, F and G), which may reflect the transient nature of these synapses. We confirmed that this type of junctions was formed between MpNs and SpNs by double-label immunoelectron microscopy (fig. S7).

To investigate the roles of synapses between MpNs and SpNs, we examined the neuronal activity of SpNs by Ca^{2+} imaging using GCaMP3. SpNs exhibited Ca^{2+} transients (Fig. 3A and movie S5), suggesting that they were neuronally active as early as E15. We also tried to visualize the presynaptic activity of SpNs using cortical slices electroporated with synaptophysin-pHluorin constructs (*8*) at E10. Synaptophysin-pHluorin signals were observed at the upper IZ (Fig. 3B and movie S6), which were not observed when tetanus toxin (TeNT) was coexpressed (Fig. 3B). TeNT cleaves synaptobrevin and thus blocks exocytosis of synaptic vesicles, indicating that these signals were specific. Moreover, upon K^+ stimulation, synaptophysin-pHluorin signals were augmented around several cells, presumably MpNs, just below the SP (Fig. 3B and movie S7). These observations suggested that SpNs release neurotransmitters around MpNs.

We examined whether neuronal activities of SpNs were necessary for radial migration by overexpressing the inwardly rectifying potassium channel Kir2.1 (*9*), which causes a suppression of neuronal excitability. The embryos were doubly electroporated with Kir2.1 and GFP constructs at E10 and subsequently with RFP construct at E14. In contrast to the control cortices, where many RFP-labeled neurons entered the CP, they were accumulated just below SP when Kir2.1 was expressed in the SpNs (Fig. 3C, arrows). We further tried to suppress neurotransmitter release from

[1]Neural Network Project, Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan. [2]Center for Basic Technology Research, Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan. [3]Neural Development Project, Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan. [4]Anatomy and Cell Biology, Nagoya University Graduate School of Medicine, Aichi 466-8550, Japan. [5]Department of Biology, Ochanomizu University, Tokyo 112-8610, Japan. [6]Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo 112-8610, Japan. [7]School of Advanced Science and Engineering, Waseda University, Tokyo 169-0072, Japan.
*Corresponding author. Email: maruyama-ck@igakuken.or.jp (C.O.-M.); maeda-nb@igakuken.or.jp (N.M.)

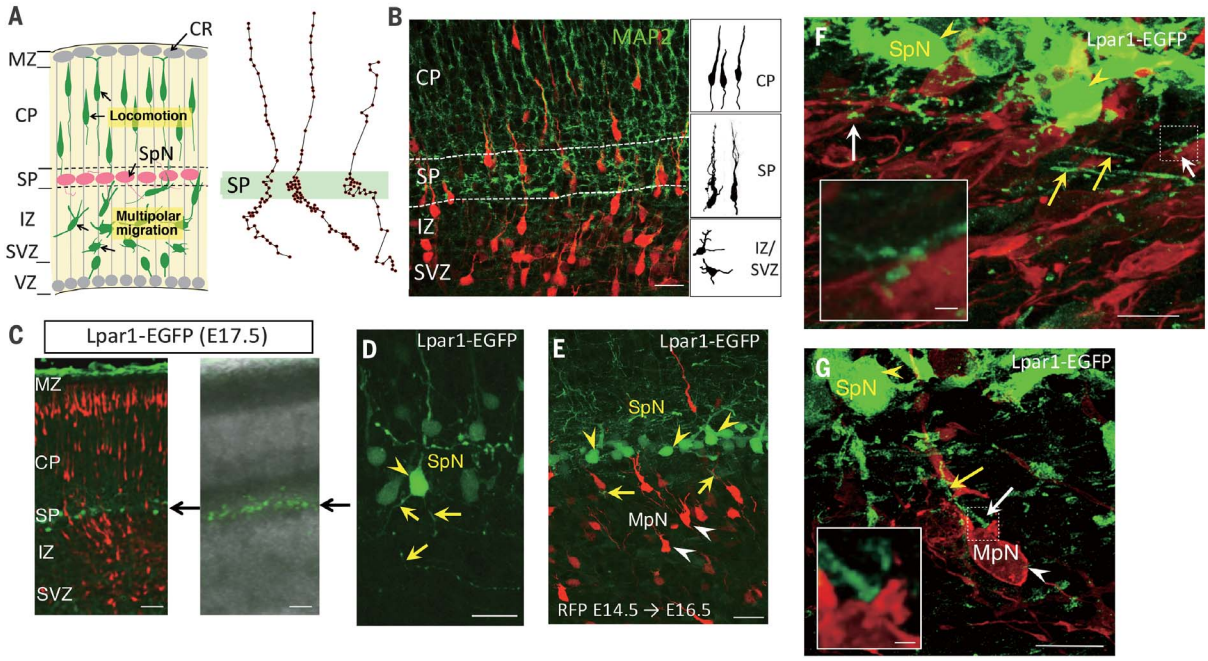


Fig. 1. SP as a strategic position for radial migration. (**A**) Newborn neurons first show multipolar migration. After multipolar-to-bipolar transition, they initiate locomotion (left). Three migrating neurons shown in movie S1 were tracked (right). (**B**) In the SP, migrating neurons showed a transitional form between multipolar and bipolar morphologies. (**C**) SpNs and migrating neurons were doubly labeled with GFP (green) and RFP (red), respectively, in the Lpar1-EGFP mouse cortex. The SP layer (arrow) can be visualized as the dark zone under bright-field microscopy. (**D**) SpNs (yellow arrowheads) extend axon-like processes (arrows). (**E** to **G**) SpNs (yellow arrowheads) and MpNs (white arrowheads) were visualized in cortical slices of Lpar1-EGFP mice. Yellow arrows indicate axon-like processes of SpNs. Enlarged images showed that neurites of SpNs contacted with migrating MpNs just under the SP. The contact sites often showed a varicosity-like structure (white arrows). Scale bars: 50 μm for (C); 20 μm for (B), (D), and (E); and 1 and 10 μm for (F) and (G).

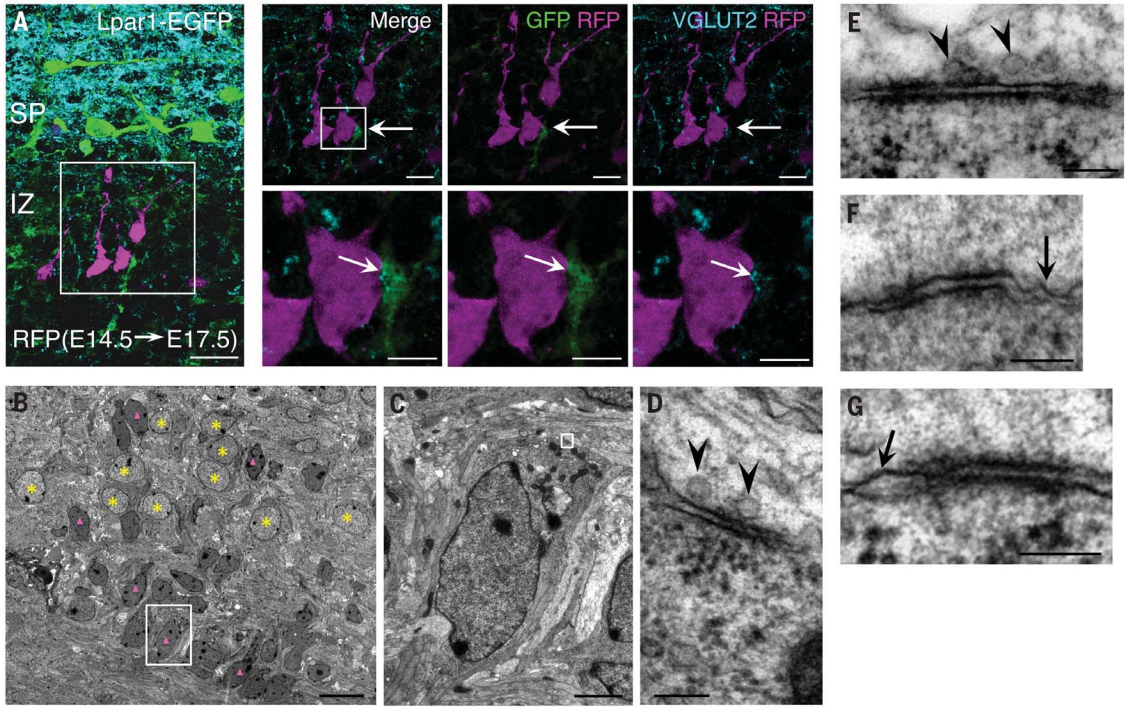


Fig. 2. Glutamatergic synaptic contacts between SpNs and MpNs. (**A**) SpNs and migrating neurons were doubly labeled with GFP (green) and RFP (magenta), respectively, in the cortex of Lpar1-EGFP mice. The SpN neurites showed VGLUT2 immunoreactivity (cyan) at the contact sites with MpNs (arrows). The insets show the enlarged areas indicated by rectangles. (**B**) E16 cortical sections were observed by electron microscopy. SpNs and migrating neurons are marked with asterisks and triangles, respectively. (**C** to **E**) The areas indicated by rectangles in (B) and (C) are enlarged in (C) and (D), respectively. Arrowheads in (D) and (E) show synaptic vesicles. (**F** and **G**) The junctional structures without synaptic vesicles. Arrows indicate wavy plasma membranes. Scale bars: 20 μm for (A), left; 10 μm for (A), right upper layer, and (B); 5 μm for (A), right lower layer; 2 μm for (C); and 100 nm for (D) to (G).

SpNs using TeNT knock-in mice, in which TeNT is expressed in a Cre-dependent manner (*10*). Using this mouse, in utero double electroporation of Cre expression plasmid at E10 and RFP constructs at E14 was performed. As in the case of Kir2.1 experiments (Fig. 3C), radial migration was inhibited just below the EGFP/TeNT-expressing SpNs (Fig. 3D, arrows). These results suggested that the synaptic transmission from SpNs to MpNs is critical for MpNs to initiate locomotion, although some deep CP neurons might also contribute to the synaptic transmission (fig. S3).

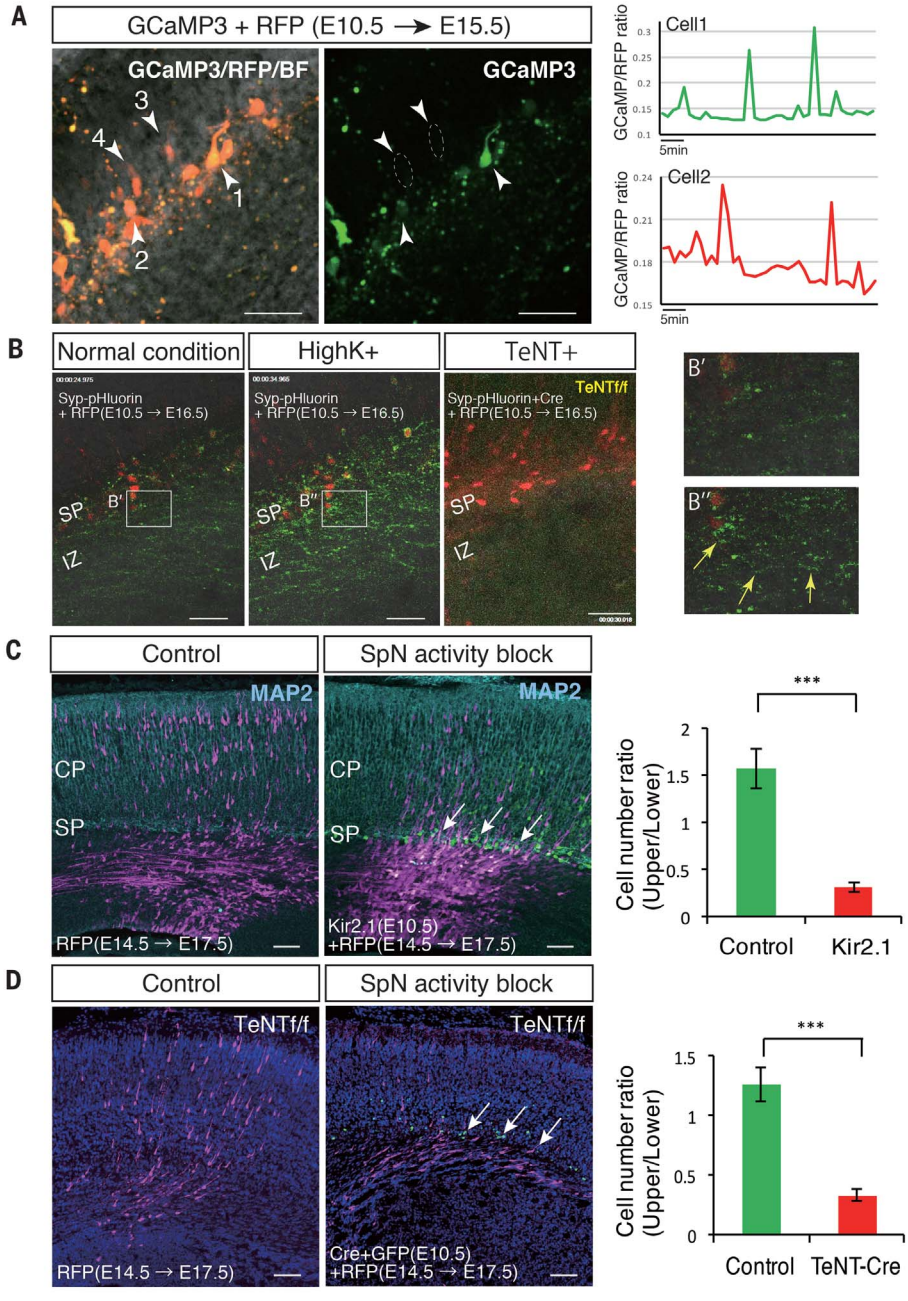


Fig. 3. Critical roles of neuronal activity of SpNs in radial neuronal migration. (**A**) Calcium imaging of SpNs (cells 1 and 2) electroporated with the GCaMP3 construct. Deep CP neurons scarcely showed Ca^{2+} rises (cells 3 and 4). (**B**) SpNs were electroporated with the synaptophysin-pHluorin construct, and the cortical slices were applied for imaging analyses. Synaptophysin-pHluorin signals were augmented after K^+ stimulation around many cells (**B′** and **B″**, arrows). The signals were suppressed when TeNT was coexpressed. (**C**) Neuronal migration was blocked when neuronal activity of SpNs was suppressed by Kir2.1. (**D**) Neuronal migration was blocked when neurotransmitter release from SpNs was impaired by EGFP-TeNT. Three brains of independent experiments were used, and two sections from each brain were analyzed for Kir2.1 experiments ($N = 6$). Two brains of independent experiments were used, and three sections from each brain were analyzed for EGFP-TeNT experiments ($N = 6$). Scale bars, 50 μm. ***$P < 0.0005$. Error bars, mean ± SEM.

Glutamate and *N*-methyl-D-aspartate receptors (NMDARs) are involved in radial migration (*11*–*14*). In these studies, glutamate in the CP has been considered to act as a chemoattractant for migrating neurons via nonsynaptic NMDARs. On the other hand, if glutamatergic synaptic transmission induces the change of migration mode, local application of glutamate should enhance the initiation of locomotion. Thus, we performed experiments using caged glutamate (Fig. 4A and movie S8). MpNs in the uncaged area, which were marked by the red fluorescence of Kikume green-red, rapidly initiated locomotion and passed through the SP without a stationary phase. The displacement of neurons from the uncaged areas and the migration toward the MZ cannot be explained by simple chemotactic effects of glutamate.

To estimate the glutamate receptors involved in the synaptic transmission between SpNs and MpNs, we first performed a gene expression profiling of newborn excitatory neurons (fig. S8). After in utero electroporation of GFP expression plasmid into E14 cortices, RNAs were purified from GFP-labeled cells collected at E15, E16, and E17 by fluorescence-activated cell sorting (FACS). Because GFP-labeled neurons passed through the multipolar to bipolar stages during these 3 days, we could show changes in gene expression that occurred as the migration stages proceeded. Enrichment analyses revealed that the expression of genes associated with gene ontology terms containing "synaptic contact" increased during migration. In particular, gene expressions of NR1 and NR2B subunits of NMDARs and postsynaptic density proteins such as PSD95 increased (fig. S9, A and B). Furthermore, NMDAR antagonists AP5 and MK801 inhibited radial migration in the cultured cortical slices, as reported previously using rat tissues (*12*) (fig. S9C).

We thus investigated the postsynaptic mechanism of MpNs focusing on PSD95 and NMDAR. Knockdown of PSD95 resulted in impaired radial migration, which was rescued by the short hairpin RNA (shRNA)–resistant expression construct of PSD95 (Fig. 4B and fig. S10, A and B). Observation of the morphology of migrating neurons revealed that multipolar-to-bipolar transition was impaired by knockdown of PSD95 (fig. S10, C to E). We next examined the involvement of NMDARs using $NR1^{flox/flox}$ mice (*15*). In utero electroporation of Cre expression plasmid resulted in accumulation of NR1 knockout neurons just under the SP (Fig. 4C). Because NR1 and PSD95 are essential components for NMDAR-mediated synaptic transmission, these results support the idea that glutamatergic synaptic transmission from SpNs to MpNs via NMDARs induces multipolar-to-bipolar transition.

Finally, we performed Ca^{2+} imaging of migrating neurons. We observed that the Ca^{2+} concentration in MpNs transiently increased just before initiation of locomotion, suggesting that Ca^{2+} entry through NMDARs plays roles in multipolar-to-bipolar transition (fig. S11 and movies S9 and

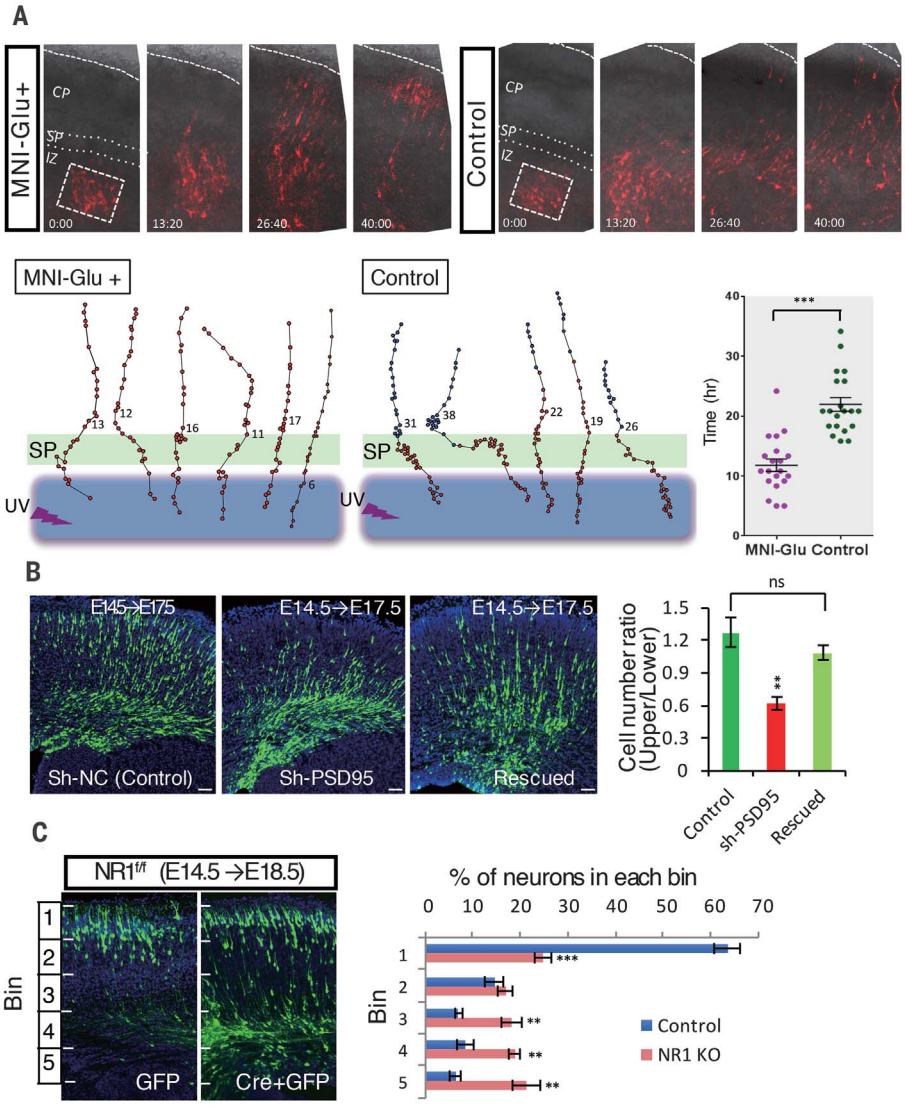


Fig. 4. NMDAR and PSD95 in MpNs are involved in radial migration. (**A**) The cortical slices were cultured in the presence (MNI-Glu^{+}) or absence (control) of caged glutamate, and the areas indicated by rectangles were laser-irradiated. Tracking of neuronal migration indicated that neurons in the uncaged area rapidly initiated locomotion. The time it took for red neurons to initiate locomotion after laser irradiation was plotted (N = 20). (**B**) Knockdown of PSD95 resulted in accumulation of MpNs below the SP. Three brains of independent experiments were used, and three sections from each brain were analyzed (N = 9). (**C**) The cortices of NR1$^{flox/flox}$ mice were electroporated with Cre and GFP expression plasmids at E14. Knockout of NR1 resulted in delay of neuronal migration. Three sections from one embryo and four sections from another embryo of independent experiments (N = 7) were used for quantification. Scale bars, 50 μm. $^{**}P < 0.005$. $^{***}P < 0.0005$. Error bars, mean ± SEM.

S10). However, we cannot exclude the possibility that AMPA receptors are also involved in Ca^{2+} signaling because their expressions increased during migration (fig. S9B). Increased Ca^{2+} transients in migrating neurons in the CP, which may be induced by radial glial fibers (*16*), cause migration arrest and maturation (*17*), suggesting that the roles of Ca^{2+} signaling change depending on the stage of migration.

Our data suggested that subplate neurons facilitate multipolar-to-bipolar transition of migrating excitatory neurons by NMDAR-mediated synaptic transmission, although the possibility that there are also contributions from deep layer neurons cannot be excluded (fig. S12). Impairment of multipolar-to-bipolar transition and accumulation of multipolar neurons below the subplate are observed in mice mutant for a variety of genes, including RP58 (*2*, *18*); some of these genes are involved in human neurodevelopmental disorders such as fragile X syndrome (*19*). Subplate neurons might regulate the expression and function of these genes in multipolar neurons through NMDAR-mediated Ca^{2+} signaling. The subplate has been associated with autism and schizophrenia by gene expression profiling (*6*). Disruption of these transient synapses during development may contribute to these and other psychiatric disorders.

REFERENCES AND NOTES

1. P. Rakic, *Science* **183**, 425–427 (1974).
2. C. Ohtaka-Maruyama, H. Okado, *Front. Neurosci.* **9**, 447 (2015).
3. J. G. Gleeson, C. A. Walsh, *Trends Neurosci.* **23**, 352–359 (2000).
4. K. L. Allendoerfer, C. J. Shatz, *Annu. Rev. Neurosci.* **17**, 185–218 (1994).
5. K. Sekine, K. Kubo, K. Nakajima, *Neurosci. Res.* **86**, 50–58 (2014).
6. A. Hoerder-Suabedissen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3555–3560 (2013).
7. A. Hoerder-Suabedissen, Z. Molnár, *Cereb. Cortex* **23**, 1473–1483 (2013).
8. B. Granseth, B. Odermatt, S. J. Royle, L. Lagnado, *Neuron* **51**, 773–786 (2006).
9. J. Burrone, M. O'Byrne, V. N. Murthy, *Nature* **420**, 414–418 (2002).
10. M. Sakamoto *et al.*, *J. Neurosci.* **34**, 5788–5799 (2014).
11. T. N. Behar *et al.*, *J. Neurosci.* **19**, 4449–4461 (1999).
12. K. Hirai *et al.*, *Brain Res. Dev. Brain Res.* **114**, 63–67 (1999).
13. H. Jiang *et al.*, *Brain Res.* **1610**, 20–32 (2015).
14. J. J. LoTurco, M. G. Blanton, A. R. Kriegstein, *J. Neurosci.* **11**, 792–799 (1991).
15. T. Iwasato *et al.*, *Nature* **406**, 726–731 (2000).
16. B. G. Rash, J. B. Ackman, P. Rakic, *Sci. Adv.* **2**, e1501733 (2016).
17. Y. Bando *et al.*, *Cereb. Cortex* **26**, 106–117 (2016).
18. C. Ohtaka-Maruyama *et al.*, *Cell Reports* **3**, 458–471 (2013).
19. G. La Fata *et al.*, *Nat. Neurosci.* **17**, 1693–1700 (2014).

ACKNOWLEDGMENTS

We acknowledge Y. Nishito for microarray analyses, T. Hara for support with FACS, A. Odajima for image processing, J. Horiuchi for critical reading of the manuscript, and S. Okabe, S. Takamori, and Y. Yoshimura for helpful discussions. **Funding:** This work was supported in part by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (KAKENHI) (17K07428 to C.O.-M., 16K07077 to N.M., and 16H02457 to T.M.), Ministry of Education, Culture, Sports, Science and Technology KAKENHI (22111006 to T.M.), and the Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS) from the Japan Agency for Medical Research and Development (AMED) to K.Y. **Author contributions:** C.O.-M., N.M., T.M., and M.Ok. designed the experiments; C.O.-M., M.Ok., K.E., M.Os., and N.K. collected the data; C.O.-M., M.Os., and K.Y. performed the statistical analysis of microarray data; H.O. provided administrative support; and C.O.-M., and N.M. wrote the paper. **Competing interests:** None declared. **Data and material availability:** The microarray data are available in Gene Expression Omnibus database under accession number GSE102911. 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/360/6386/313/suppl/DC1
Materials and Methods
Figs. S1 to S12
Movies S1 to S10
References (*20*, *21*)

22 October 2017; accepted 22 February 2018
10.1126/science.aar2866

PLANT ECOLOGY

Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment

Peter B. Reich,[1,2]* Sarah E. Hobbie,[3] Tali D. Lee,[4] Melissa A. Pastore[3]

Theory predicts and evidence shows that plant species that use the C_4 photosynthetic pathway (C_4 species) are less responsive to elevated carbon dioxide (eCO_2) than species that use only the C_3 pathway (C_3 species). We document a reversal from this expected C_3-C_4 contrast. Over the first 12 years of a 20-year free-air CO_2 enrichment experiment with 88 C_3 or C_4 grassland plots, we found that biomass was markedly enhanced at eCO_2 relative to ambient CO_2 in C_3 but not C_4 plots, as expected. During the subsequent 8 years, the pattern reversed: Biomass was markedly enhanced at eCO_2 relative to ambient CO_2 in C_4 but not C_3 plots. Soil net nitrogen mineralization rates, an index of soil nitrogen supply, exhibited a similar shift: eCO_2 first enhanced but later depressed rates in C_3 plots, with the opposite true in C_4 plots, partially explaining the reversal of the eCO_2 biomass response. These findings challenge the current C_3-C_4 eCO_2 paradigm and show that even the best-supported short-term drivers of plant response to global change might not predict long-term results.

The idea that C_4 plants are less limited by ambient atmospheric CO_2 concentrations than C_3 plants, and will thus respond less to increasing CO_2 concentrations, has a long history (*1*–*3*) and is deeply embedded in models of past, present, and future vegetation-climate interactions (*2*–*7*). The hypothesis has proven useful, if not always entirely predictive, in describing C_3 and C_4 plant distributions (*2*–*4*, *8*–*11*) and biomass responses to environmental variation (*12*–*15*).

There is strong logic for this hypothesis. C_3 plants, which use the carboxylase enzyme RuBisCO (ribulose-1,5-bisphosphate carboxylase-oxygenase) to fix CO_2 from the air and obtain 3-carbon intermediate molecules as the first step in photosynthesis, lose a portion of their fixed CO_2 to oxidative photorespiration under present CO_2:O_2 ratios because RuBisCO is also an oxygenase (*1*–*3*). Thus, C_3 plants should exhibit increased leaf-level net photosynthesis as increasing CO_2:O_2 ratios reduce rates of photorespiration and increase rates of carboxylation (*1*–*3*). By contrast, in C_4 plants, a different enzyme (phosphoenolpyruvate carboxylase) with a high affinity for CO_2 and lacking oxygenase activity first incorporates CO_2 into a 4-carbon intermediate, which is then shuttled to specialized bundle sheath cells where CO_2 is released, resulting in locally high CO_2 concentrations. Here, RuBisCO catalyzes carboxylation, but with low rates of photorespiration because of the high CO_2:O_2 ratios in the bundle sheath cells. As a result, eCO_2 levels in air have little impact on photosynthetic rates for C_4 plants (*1*–*3*). Globally, most plants are C_3; graminoids are the only major group with substantial abundance of both C_3 and C_4 species (*2*, *16*). Given that C_4 grasslands may constitute one-fifth of global terrestrial net primary productivity (*16*), a better understanding of C_4 performance under rising CO_2 vis-à-vis C_3 grasses is needed for global ecology and for improved ecosystem and Earth system modeling (*5*–*7*, *17*). C_3 and C_4 grasses (and sedges) are also distinguished in terms of ecological success by different affinities for temperature, rainfall, and nutrient supply (as well as CO_2); as a consequence, they can cooccur or shift their relative abundances depending on the mix of conditions (*8*–*11*, *14*, *15*, *18*–*20*).

Experimental evidence has strongly supported the theoretical prediction that at current CO_2 levels, C_3 grasses are more CO_2-limited than C_4 species and thus will respond more to rising CO_2. Two influential meta-analyses (*12*, *13*) reported greater stimulation (by factors of 2 to 4) of aboveground biomass by eCO_2 for C_3 species than for C_4 species. One of those publications focused on published studies of graminoids, where plants were mostly grown in pots for less than a year (*12*), whereas the other focused on free-air CO_2 enrichment (FACE) field studies (mostly 3 years or shorter) with plants grown in the ground (*13*). Another more recent meta-analysis also showed much greater responses of biomass growth for C_3 than for C_4 plants (*21*). One of the few longer-term studies comparing C_3 and C_4 grasses under eCO_2 found that C_4 species were, surprisingly, apparent "winners" after 3 years of eCO_2 and warming (*20*). However, by 6 years of treatment, the situation had reversed, with C_3 species becoming more positively responsive to the simulated global changes and the dominant plant type, in accordance with expectations (*22*).

In sum, current physiological theory and short- and medium-term studies support the paradigm that C_4 species benefit much less from rising CO_2 than C_3 species. Given both theoretical and empirical support for the differences in eCO_2 response of C_3 and C_4 species, their differential responsiveness to eCO_2 comes as close to "accepted fact" as exists in ecology and, as such, is incorporated into many ecosystem and Earth system models (*3*–*6*, *8*–*13*). However, it is not known whether findings from short- and medium-term studies apply over ecologically realistic time frames (>10 years) in field settings where complex feedbacks might influence response to eCO_2. Understanding long-term responses of C_3 and C_4 species is especially germane given that the limited available evidence suggests strong nonlinearity of responses of ecosystems to eCO_2 over time (*23*).

Here, we report results from a long-term (20-year) FACE experiment in Minnesota, USA, that support the long-held paradigm for the early part of the experiment but reveal a gradual reversal to a much more positive response to eCO_2 by C_4 than by C_3 grasses. The study uses 88 plots that are components of several different but overlapping global change experiments within the BioCON project (*24*, *25*) and by themselves constitute a fully factorial 2 × 2 × 2 × 2 experiment of CO_2 levels (ambient or +180 parts per million), nitrogen levels (ambient or +4 g N m^{-2} $year^{-1}$), species richness (one or four species), and functional group identities (C_3 or C_4 grasses) (*26*). Eight species of temperate perennial grasses (four each of C_3 and C_4) were used in the study (*26*) and were equally weighted in the original plantings of those 88 plots; that is, there are equal numbers of replicated monocultures of all species, and the four-species plots contain all species within each functional group. Annually over 20 years, we sampled both aboveground and belowground (0 to 20 cm) biomass late in each growing season in every plot, and also made an independent measure of fine root production (0 to 20 cm). We also measured in situ soil net N mineralization in every plot for a 1-month period each year just prior to biomass sampling (*26*, *27*). Leaf-level net photosynthetic rates were measured midseason (*28*) for a subset of these eight grass species in 16 of the 20 years.

Over the 20-year experimental period, total biomass of C_4 grasses became increasingly enhanced by eCO_2 exposure, with the reverse true for C_3 grasses (Figs. 1 and 2 and figs. S1 and S2). During approximately the first 12 years (1998–2009), results were as expected (Fig. 1): C_3 plots averaged a 20% increase in total biomass (+136 g/m^2) at eCO_2 relative to ambient CO_2, in contrast to C_4 plots that averaged a 1% increase (+12 g/m^2). During the subsequent 8 years (2010–2017), the pattern reversed: C_3 plots averaged 2% less (−12 g/m^2) and C_4 plots 24% more (+233 g/m^2) biomass in eCO_2 than in ambient CO_2 (Fig. 1).

Repeated-measures analyses of variance (Table 1) support these conclusions, which are illustrated for successive 5-year periods in Fig. 2. Significant main effects on total biomass were found for N addition (higher biomass than at ambient N), species richness (higher biomass in plots with four species than in plots with one species), and functional group (higher biomass on average in

[1]Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA. [2]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW 2753, Australia. [3]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA. [4]Department of Biology, University of Wisconsin, Eau Claire, WI 54701, USA.
*Corresponding author. Email: preich@umn.edu

C_4 plots than in C_3 plots) (Table 1). Additionally, on average across treatments, biomass of C_3 plots was originally greater than that of C_4 plots, but over time this ranking reversed (interaction of functional group × year, $P < 0.0001$; Table 1 and Fig. 1). Most germane was the significant functional group × year × CO_2 interaction ($P = 0.007$; Table 1), showing that C_3 and C_4 functional groups responded differently to eCO_2 over time (Fig. 2). For example, in each of the first two 5-year periods, C_3 grasses increased biomass under eCO_2 by ~20% (+140 g/m^2); this declined to a 10% enhancement (+40 g/m^2) in years 11 to 15 and a 2% decline in years 16 to 20 (–15 g/m^2). In contrast, under eCO_2, biomass of C_4 grasses was reduced by 2% (–23 g/m^2) in years 1 to 5, enhanced by ~7% (+60 g/m^2) in years 5 to 10 and 11 to 15, and enhanced by 31% (+298 g/m^2) in years 16 to 20. These different responses of functional groups to CO_2 and time were unaffected by N treatment ($P = 0.76$ for functional group × year × CO_2 × N interaction) and were slightly more pronounced in four-species plots than in one-species plots ($P = 0.048$ for functional group × year × CO_2 × species richness interaction) (Table 1 and fig. S3). Results were generally similar for aboveground and belowground biomass viewed separately, as well as for annual net primary production (estimated as the sum of annual aboveground biomass production and fine root production).

We explored several potential mechanisms for this long-term reversal of C_3 versus C_4 responsiveness to eCO_2, including a temporal switch in leaf-level photosynthetic response, differential CO_2 sensitivity associated with potential climate variation over the 20 years, and potential feedbacks from changing N cycle responses to eCO_2 over time. Measurements of light-saturated net photosynthesis were made for one to three C_3 and one to three C_4 grass species (mean of 2.2) in monocultures in 16 of the 20 years of the study, at all combinations of CO_2 and N treatment. There was no evidence of a shift over time in the enhancement of net photosynthesis as observed for biomass (no interactions of functional group × CO_2 × year; fig. S4). Moreover, there was no correspondence between years when eCO_2 enhancement of net photosynthesis was high and years when eCO_2 enhancement of biomass was high, in either functional group (compare Fig. 1 and fig. S4). Although we lack data for all species in all treatments in all years, the available data provide no evidence to suggest that the rank reversals of biomass responses to eCO_2 were driven by parallel rank reversals in leaf-level photosynthetic responses.

We then asked whether the shifting responsiveness of C_3 versus C_4 grass plots could be related to interannual variation in temperature or rainfall (*2–5*, *15*, *19*, *21*). Responses of C_3 and C_4 grasses did not depend on year-to-year variations in mean or lagged spring, summer, or growing-season daily average temperature. The only significant effect involved summer rainfall [May to July (MJJ)]: There was a significant ($P = 0.0264$) interaction of CO_2 × functional group × MJJ rainfall on the biomass response (table S1); C_4 grasses were slightly more responsive to eCO_2 when rainfall was higher, whereas C_3 grasses were more responsive in low rainfall. These results are inconsistent with C_3 and C_4 grass responses in many studies (*2–5*, *15*, *19*). However, MJJ rainfall was only weakly correlated with year, and the CO_2 × year × functional group interaction was significant in the model ($P = 0.0347$) even after accounting for differential responses to rainfall for the two functional

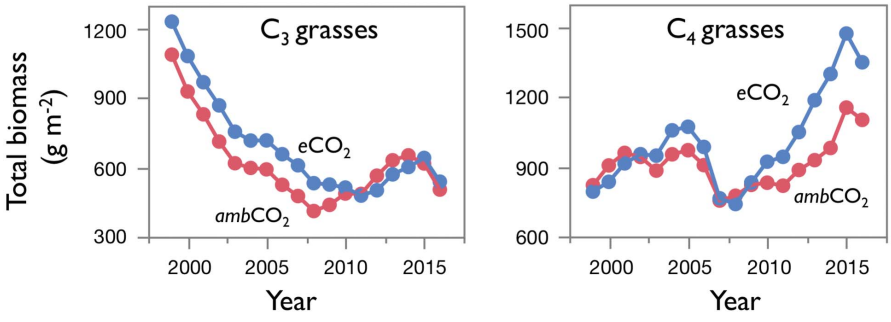


Fig. 1. Biomass over time of C_3 grasses and C_4 grasses at ambient and elevated CO_2. Total biomass (aboveground + 0 to 20 cm belowground) of plots comprising C_3 grasses and C_4 grasses in ambient CO_2 (red) and elevated CO_2 (blue) from 1998 to 2017. Data are shown as moving 3-year averages centered over the middle of each 3-year group. Each point represents data pooled across N treatments, and across monoculture and four-species plots (equally weighted), for each functional group ($n = 22$ plots for each functional group at each CO_2 level). See Table 1 for statistical analysis and fig. S1 for annual data and information on variation in response within treatments.

Table 1. Summary of repeated-measures analysis of variance of year, CO_2, N, species richness (SR), and C_3 versus C_4 functional group (FuncGroup) effects on total biomass and soil net N mineralization. Three-way or higher interactions involving treatments and two-way interactions or higher involving covariates shown only if significant. Five-way interactions were not tested. N mineralization data were missing for 2008 and 2017. Biomass was log-transformed prior to analysis. Year was a continuous term; Year and Year × Year terms were included in the model to assess linear and nonlinear changes over time involving CO_2 and functional groups. Significant terms ($P < 0.05$) in bold font.

Effect	Total biomass (g m^{-2})		Net N mineralization (mg kg^{-1} day^{-1})	
Whole model	$R^2 = 0.609$, $P < 0.0001$, $n = 1760$		$R^2 = 0.086$, $P < 0.0001$, $n = 1582$	
	F	*P* > *F*	*F*	*P* > *F*
Year	**158.95**	**<0.0001**	0.31	0.5787
CO_2	2.92	0.1521	0.75	0.4020
N	**13.14**	**0.0005**	**12.19**	**0.0008**
SR	**10.11**	**0.0022**	0.01	0.9187
FuncGroup	**42.32**	**<0.0001**	**5.52**	**0.0195**
Year × CO_2	1.96	0.1614	0.39	0.5327
Year × N	2.26	0.1326	2.08	0.1493
Year × SR	**23.64**	**<0.0001**	0.00	0.96865
Year × FuncGroup	**296.56**	**<0.0001**	1.46	0.2269
CO_2 × N	0.01	0.9342	0.67	0.4170
CO_2 × SR	0.01	0.9265	0.36	0.5505
CO_2 × FuncGroup	0.02	0.8965	0.49	0.4853
N × SR	1.42	0.2379	0.73	0.3953
N × FuncGroup	0.48	0.4885	**5.48**	**0.0221**
SR × FuncGroup	0.40	0.5314	0.56	0.4574
Year × CO_2 × FuncGroup	**7.24**	**0.0072**	**3.99**	**0.0461**
Year × CO_2 × FuncGroup × SR	**3.89**	**0.0486**	1.16	0.2827
Year × Year	**104.34**	**<0.0001**	**20.87**	**<0.0001**
Year × Year × CO_2	0.23	0.6325	0.42	0.5189
Year × Year × CO_2 × FuncGroup	0.08	0.7709	0.27	0.6034

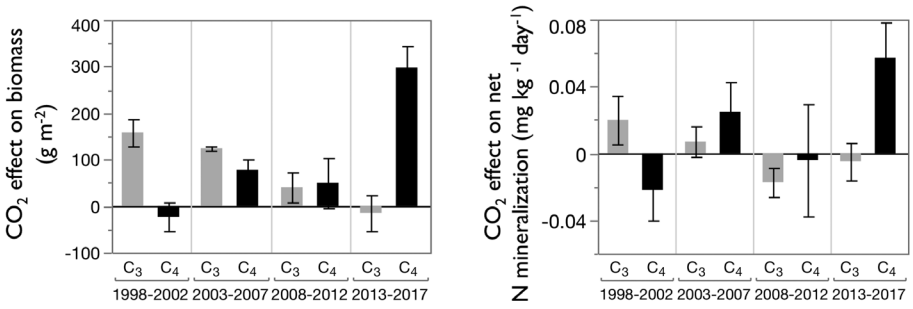


Fig. 2. Elevated CO_2 effect on total biomass and soil net N mineralization rates in plots comprising C_3 grasses and C_4 grasses. Left: Mean annual difference in biomass (mean biomass in eCO_2 – mean biomass in ambient CO_2) for C_3 and C_4 plots for four time periods during the study (1998–2002, 2003–2007, 2008–2012, and 2013–2017). Right: Mean difference in mean soil N mineralization rate (eCO_2 – ambient CO_2) for C_3 and C_4 plots for the same four time periods. Each bar represents data pooled across N treatments, and across monoculture and four-species plots (equally weighted), for each functional group (n = 22 plots for each functional group at each CO_2 level). Error bars represent SE among years.

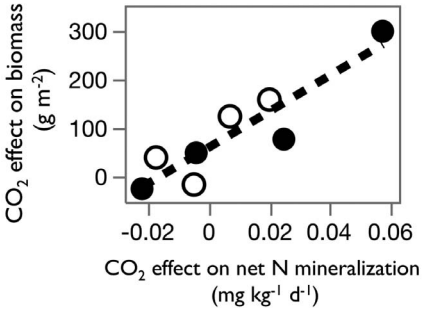


Fig. 3. Correspondence between biomass and net N mineralization responses to elevated CO_2. Relationship of biomass response to eCO_2 (effect size = biomass in eCO_2 – biomass in ambient CO_2) versus net N mineralization response to eCO_2 (defined similarly) in plots comprising C_3 grasses (open circles) and C_4 grasses (solid circles). R^2 = 0.82, P = 0.0018. Each data point represents effect sizes for each functional group for the four 5-year periods during the study (as in Fig. 2), based on the average biomass and net N mineralization across years in each period at each CO_2 level.

groups by including rainfall and rainfall interactions in the model (table S1). Thus, the reversal of responsiveness of C_3 and C_4 plots to eCO_2 over time was not explained by interannual variation in precipitation.

We also considered soil processes that might have played a role in the shifting responses of the C_3 and C_4 assemblages. Soil N supply has shaped the dynamics of biomass response to eCO_2 in the wider BioCON experiment (including the 9- and 16-species mixtures) because, as predicted by multiple resource limitation theory, responses to eCO_2 were greater when N supply levels were high [e.g., (*27*)]. Hence, we asked what role soil N availability (using soil net N mineralization as an index) might play here. The response of soil net N mineralization to eCO_2 changed over time (Table 1, Fig. 2, and figs. S2 and S5), mirroring responses of biomass for the C_3 and C_4 groups (Table 1). There was a significant (P = 0.046) interaction of CO_2 × year × functional group: The response of net N mineralization to eCO_2 in C_4 grass plots became more positive over time, whereas that of C_3 grass plots became more negative (Table 1 and Fig. 2).

Moreover, relationships between biomass and soil net N mineralization rate, in concert with the shifts in the response of net N mineralization to eCO_2 over time, help to explain the shifting biomass response to eCO_2 of both functional groups. Biomass and its response to eCO_2 were both positively related to net N mineralization rate: Across years, biomass in both functional groups increased with net N mineralization (P = 0.0031; table S1), more so at eCO_2 than at ambient CO_2 (interaction of CO_2 × net N mineralization, P = 0.024; table S1), and more so in C_4 than in C_3 plots (interaction of CO_2 × functional group, P = 0.038; table S1). Given that biomass is positively related to net N mineralization, and that the net N mineralization response to eCO_2 was increasingly positive over time in C_4 grass plots and increasingly negative in C_3 grass plots, shifting soil N biogeochemistry partially explains the shifting biomass responses to eCO_2.

These effects can be illustrated by the significant positive linear relationship between the eCO_2 enhancement of biomass and the eCO_2 enhancement of net N mineralization for the four 5-year periods of the study (Fig. 3): In periods when net N mineralization rates were higher under eCO_2 than under ambient CO_2, biomass tended to be higher in eCO_2 as well (Fig. 3 and table S1). These results are consistent with prior results showing that response to eCO_2 in this ecosystem is partially contingent on N supply (*27*), with greater N availability tending to promote greater eCO_2 response. Overall, 20 years of observations in this FACE experiment suggest that the opposite directional responses of net N mineralization to eCO_2 over time in C_3 versus C_4 grass plots (Table 1 and Fig. 2) may have contributed to the reversal of C_3 and C_4 biomass responses to eCO_2 over time. Why these soil N cycling responses played out in this fashion remains an open question, however.

Models that simulate future carbon cycling responses at ecosystem, regional, and global scales assume differing sensitivities of C_3 versus C_4 species to CO_2 based on differences in their photosynthetic physiology (*5*, *6*, *8*–*11*, *17*). Although those assumptions have major impacts on vegetation dynamics under varying climate and CO_2 scenarios (*8*–*11*, *29*, *30*), they do not match up well with the dynamic results of this long-term study. Our results thus serve as a reminder that even the best-predicted short-term ecosystem responses to global change can yield mid-term (decades) to long-term (centuries) surprises, as complex responses and interactions may occur over time. Determining whether the mid- to long-term responses demonstrated here are themselves broadly predictable represents a major unmet challenge for experimental and observational studies.

REFERENCES AND NOTES

1. J. Ehleringer, O. Björkman, *Plant Physiol.* **59**, 86–90 (1977).
2. J. R. Ehleringer, T. E. Cerling, B. R. Helliker, *Oecologia* **112**, 285–299 (1997).
3. R. F. Sage, D. S. Kubien, *Plant Cell Environ.* **30**, 1086–1106 (2007).
4. G. J. Collatz, J. A. Berry, J. S. Clark, *Oecologia* **114**, 441–454 (1998).
5. A. B. Harper *et al.*, *Geosci. Model Dev.* **9**, 2415–2440 (2016).
6. K. W. Oleson *et al.*, *Technical Description of Version 4.5 of the Community Land Model (CLM)* (National Center for Atmospheric Research, 2013).
7. Y. Luo *et al.*, *Glob. Change Biol.* **14**, 1986–1999 (2008).
8. E. J. Edwards *et al.*, *Science* **328**, 587–591 (2010).
9. L. Bremond, A. Boom, C. Favier, *Glob. Change Biol.* **18**, 2324–2334 (2012).
10. J. M. Cotton, T. E. Cerling, K. A. Hoppe, T. M. Mosier, C. J. Still, *Sci. Adv.* **2**, e1501346 (2016).
11. D. M. Griffith, J. M. Cotton, R. L. Powell, N. D. Sheldon, C. J. Still, *J. Biogeogr.* **44**, 2564–2574 (2017).
12. S. J. E. Wand, G. Y. F. Midgley, M. H. Jones, P. S. Curtis, *Glob. Change Biol.* **5**, 723–741 (1999).
13. E. A. Ainsworth, S. P. Long, *New Phytol.* **165**, 351–371 (2005).
14. H. W. Polley, J. D. Derner, R. B. Jackson, B. J. Wilsey, P. A. Fay, *J. Exp. Bot.* **65**, 3415–3424 (2014).
15. Z. Xu *et al.*, *Planta* **239**, 421–435 (2014).
16. C. J. Still, J. A. Berry, G. J. Collatz, R. S. DeFries, *Global Biogeochem. Cycles* **17**, 6-1–6-14 (2003).
17. A. Ali, B. Medlyn, K. Crous, P. B. Reich, *Funct. Ecol.* **27**, 1011–1022 (2013).
18. J. A. Langley, J. P. Megonigal, *Nature* **466**, 96–99 (2010).
19. S. H. Taylor *et al.*, *Glob. Change Biol.* **20**, 1992–2003 (2014).
20. J. A. Morgan *et al.*, *Nature* **476**, 202–205 (2011).
21. D. Wang, S. A. Heckathorn, X. Wang, S. M. Philpott, *Oecologia* **169**, 1–13 (2012).
22. K. E. Mueller *et al.*, *Ecol. Lett.* **19**, 956–966 (2016).
23. L. C. Andresen *et al.*, *Adv. Ecol. Res* **55**, 437–473 (2016).
24. P. B. Reich *et al.*, *New Phytol.* **150**, 435–448 (2001).
25. P. B. Reich *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10101–10106 (2004).
26. See supplementary materials.
27. P. B. Reich, S. E. Hobbie, *Nat. Clim. Chang.* **3**, 278–282 (2013).
28. T. D. Lee, S. H. Barrott, P. B. Reich, *Glob. Change Biol.* **17**, 2893–2904 (2011).
29. M. Baudena *et al.*, *Biogeosciences* **12**, 1833–1848 (2015).
30. R. Whitley *et al.*, *Biogeosciences* **14**, 4711–4732 (2017).

ACKNOWLEDGMENTS

We thank K. Worm, D. Bahauddin, and numerous summer interns for assistance with experimental operation and data collection and management. **Funding:** Supported by NSF Long-Term Ecological Research grants DEB-0620652 and DEB-1234162, Long-Term Research in Environmental Biology grant DEB-1242531, and Ecosystem Sciences grant DEB-1120064 and by U.S. Department of Energy Programs for Ecosystem Research grant DE-FG02-96ER62291. **Author contributions:** P.B.R. designed the study and supervised the overall experiment and measurements; P.B.R. analyzed the data with assistance from S.E.H., T.D.L., and M.A.P.; T.D.L. and M.A.P. collected the photosynthetic data;

P.B.R. wrote the first draft; and all authors jointly revised the manuscript. **Competing interests:** None. **Data and materials availability:** The data reported in this paper are available at the Environmental Data Initiative (EDI) (net nitrogen mineralization, DOI 10.6073/pasta/2ac4677a929290462877fd0df375ffa4; net nitrogen mineralization, DOI 10.6073/pasta/2ac4677a929290462877fd0df375ffa4; aboveground biomass, DOI 10.6073/pasta/8524be9f00b40a9e71b73a8ba2dc9ed0; belowground biomass, DOI 10.6073/pasta/c00662959002e588597bd77e0c7dbdbb). 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/360/6386/317/suppl/DC1

Materials and Methods
Figs. S1 to S5
Table S1
References (*31*, *32*)

9 January 2018; accepted 21 March 2018
10.1126/science.aas9313

EARLY EARTH

Two-billion-year-old evaporites capture Earth's great oxidation

C. L. Blättler,[1]* M. W. Claire,[2,3,4] A. R. Prave,[2] K. Kirsimäe,[5] J. A. Higgins,[1] P. V. Medvedev,[6] A. E. Romashkin,[6] D. V. Rychanchik,[6] A. L. Zerkle,[2,3] K. Paiste,[7] T. Kreitsmann,[5] I. L. Millar,[8] J. A. Hayles,[9] H. Bao,[10] A. V. Turchyn,[11] M. R. Warke,[2] A. Lepland[12,7,5,13]

Major changes in atmospheric and ocean chemistry occurred in the Paleoproterozoic era (2.5 to 1.6 billion years ago). Increasing oxidation dramatically changed Earth's surface, but few quantitative constraints exist on this important transition. This study describes the sedimentology, mineralogy, and geochemistry of a 2-billion-year-old, ~800-meter-thick evaporite succession from the Onega Basin in Russian Karelia. The deposit consists of a basal unit dominated by halite (~100 meters) followed by units dominated by anhydrite-magnesite (~500 meters) and dolomite-magnesite (~200 meters). The evaporite minerals robustly constrain marine sulfate concentrations to at least 10 millimoles per kilogram of water, representing an oxidant reservoir equivalent to more than 20% of the modern ocean-atmosphere oxidizing capacity. These results show that substantial amounts of surface oxidant accumulated during this critical transition in Earth's oxygenation.

The geological record preserves evidence of Earth's dynamic surface oxygenation [reviewed in (*1*, *2*)], but quantifying this history remains a challenge. The presence or absence of red beds, banded iron formations, and detrital grains of pyrite and uraninite (*1*, *3*) qualitatively indicate increasing oxidation during the Paleoproterozoic era, and the disappearance of large-magnitude mass-independent fractionation (MIF) of sulfur isotopes at 2.4 to 2.3 billion years ago (*4*, *5*) shows that the atmosphere exceeded a redox threshold of ~1 part per million P_{O_2} (partial pressure of oxygen) (*6*, *7*). However, this limit reflects only a tiny fraction of the potential surface oxidant budget and does not capture subsequent redox changes in the Earth system. Today, marine sulfate ($[SO_4^{2-}{}_{(aqueous)}]$ = 28 mmol/kg) constitutes one of the largest surface oxidant reservoirs, equivalent to almost twice the modern atmospheric O_2 inventory. Therefore, quantitative bounds on marine sulfate concentrations are essential for constraining the net electron balance and accumulation of oxidants on Earth's surface.

Sedimentary evaporite minerals are one of the best archives of ancient seawater chemistry [e.g., (*8*, *9*)] and specific isotopic signals [e.g., (*10*, *11*)]. Unfortunately, most Precambrian evaporite deposits consist of pseudomorphic replacements (*12*), and, until recently, the oldest known preserved evaporitic halite and bedded sulfates dated from ~830 million years ago (*13*, *14*) and ~1.2 billion years ago (*15*), respectively. This study presents analyses from a remarkably preserved ~2.0-billion-year-old marine evaporite succession bearing carbonates, sulfates, halites, and bittern salts. This succession was discovered during the 2007–2009 drilling of the Onega Parametric Hole (OPH), which intersected 2.9 km of Paleoproterozoic sedimentary and volcanic rocks and 600 m of Archean gneiss in the Onega Basin, Karelia, Russia (*16*, *17*). By extending the record of extensive marine evaporites by almost a billion years, core samples from the OPH provide a window into surface conditions and redox balance in the aftermath of the initial rise of oxygen on Earth.

The interval of the OPH presented here lies between 2940 and 2115 m depth and corresponds to the ~2.0-billion-year-old Tulomozero Formation (age discussed in the supplementary materials). In other cores and outcrop exposures, this formation contains abundant pseudomorphic replacements of evaporite minerals (*18*, *19*); original evaporite minerals are only preserved in the OPH where they define three major units. Unit 1 (2940 to 2833 m; 40% average core recovery in cored intervals) comprises dark red-pink, recrystallized halite with intraclasts of anhydrite, magnesite, and mudstone (Fig. 1, A and B), including ~10% various magnesium- and potassium-sulfate salts (Fig. 2). Unit 2 (2833 to 2330 m; 56% average core recovery in cored intervals) consists of decimeter- to meter-scale interlayered anhydrite, magnesite, and laminated dark gray to red mudstone (Fig. 1C) with minor glauberite, gypsum, and halite in its lower part. Unit 3 (2330 to 2115 m; 44% average core recovery in cored intervals) is typified by pink-tan, commonly microbially laminated dolostone (Fig. 1D), laminated red-brown-gray mudstone, and variable amounts of magnesite; quartz and dolomite pseudomorphs of calcium-sulfate minerals occur throughout the lower half

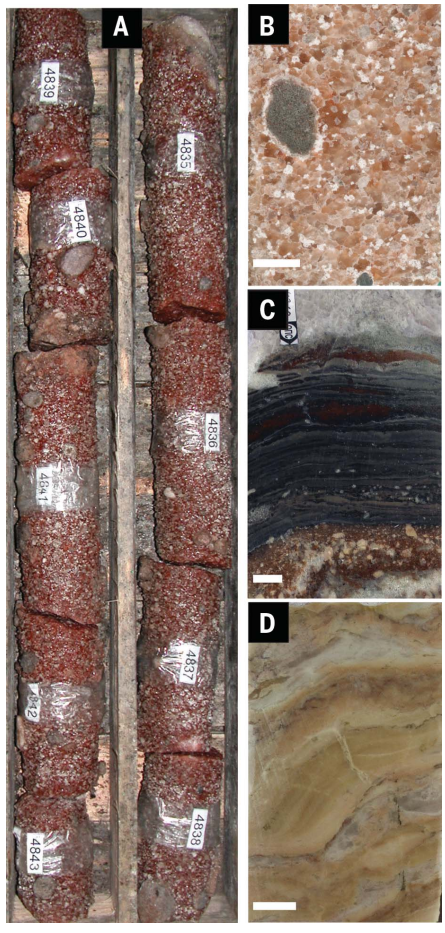


Fig. 1. Representative evaporite rocks of the Tulomozero Formation in the OPH. (**A**) Cored intervals of halite with rounded gray intraclasts consisting of mudstone, anhydrite, and magnesite (box length is 1 m; 2900 m depth). (**B**) Halite with felted anhydrite grains and anhydrite-magnesite intraclasts (2854 m depth). (**C**) Magnesite-anhydrite (white) and halite overlain sharply by laminated red-gray mudstone with desiccation cracks; a magnesite-anhydrite bed is infilling a compacted desiccation crack at the top of the image (2528 m depth). (**D**) Laminated fine-grained dolostone (2304 m depth). White scale bars are 1 cm in length.

[1]Department of Geosciences, Princeton University, Princeton, NJ 08544, USA. [2]School of Earth and Environmental Sciences, University of St Andrews, St Andrews KY16 9AL, Scotland, UK. [3]Centre for Exoplanet Science, University of St Andrews, St Andrews KY16 9AL, Scotland, UK. [4]Blue Marble Space Institute of Science, 1001 4th Avenue, Suite 3201, Seattle, WA 98154, USA. [5]Department of Geology, University of Tartu, 50411 Tartu, Estonia. [6]Institute of Geology, Karelian Research Centre, Pushkinskaya 11, 185610 Petrozavodsk, Russia. [7]Centre for Arctic Gas Hydrate, Environment and Climate, Department of Geosciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway. [8]NERC (Natural Environment Research Council) Isotope Geosciences Laboratory, British Geological Survey, Keyworth, Nottingham NG12 5GG, England, UK. [9]Department of Earth Science, Rice University, 6100 Main Street, Houston, TX 77005, USA. [10]Department of Geology and Geophysics, E235 Howe-Russell Geoscience Complex, Louisiana State University, Baton Rouge, LA 70803, USA. [11]Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, England, UK. [12]Geological Survey of Norway, 7491 Trondheim, Norway. [13]Institute of Geology, Tallinn University of Technology, 19086 Tallinn, Estonia.
*Corresponding author. Email: blattler@princeton.edu

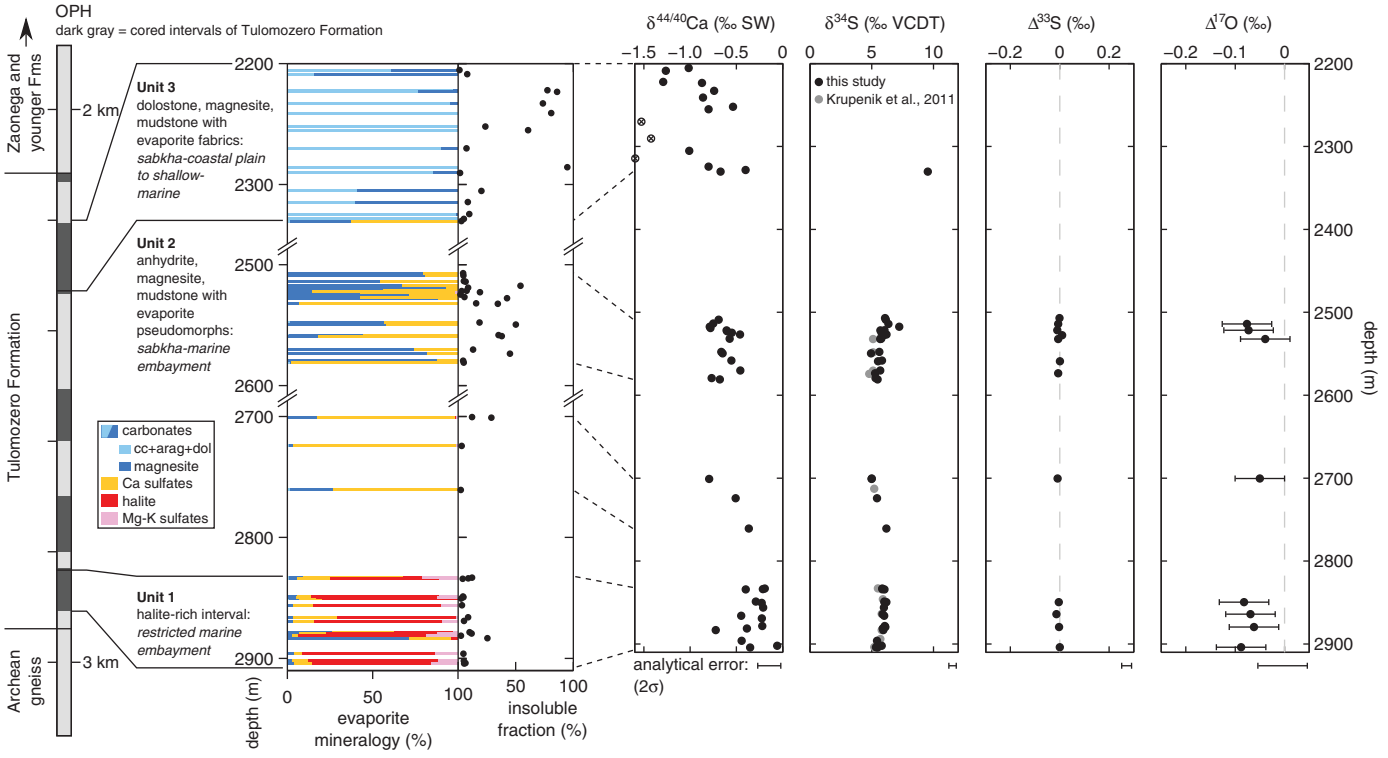


Fig. 2. Interpretive stratigraphy of the Tulomozero Formation in the OPH and associated mineralogical and geochemical data. Calcium isotope data for samples influenced by former aragonite are shown by open circles with crosses. Additional $\delta^{34}S$ data (gray filled circles) are from (*30*). Fms, formations; cc, calcite; arag, aragonite; dol, dolomite; SW, seawater; VCDT, Vienna Canyon Diablo Troilite. Methods are described in the supplementary materials.

of this unit, forming laths, nodules, discs, swallow-tail crystals, and chicken-wire fabric (fig. S2).

Considering the Tulomozero Formation's features in the OPH and its development across the 18,000 km^2 of the Onega Basin (*18*, *19*), the interpreted depositional setting is a restricted marine embayment with sabkha–coastal plain and shallow-marine environments (fig. S3). The OPH succession captures a decreasing degree of evaporation, from a state of halite and magnesium- and potassium-sulfate saturation (unit 1) through calcium-sulfate deposition (unit 2) and then to dolostone precipitation typical of a more open-marine setting (unit 3). The extent, thickness (>800 m in the OPH core), and mineral sequence of the evaporite succession are comparable to those of Phanerozoic evaporite basins.

The isotope geochemistry of the OPH evaporites presents an opportunity to investigate Earth's ocean-atmosphere system ~2.0 billion years ago (Fig. 2). Quadruple sulfur isotope analyses of samples from units 1 and 2 reveal $\Delta^{33}S$ and $\Delta^{36}S$ values that are indistinguishable from zero, confirming that the atmosphere was oxic (*6*, *7*) and that production of the Tulomozero Formation sulfate occurred well after that atmospheric transition. Triple oxygen isotope measurements of sulfates yield resolvably negative $\Delta^{17}O$ values. Oxygen MIF derived from atmospheric O_3/O_2 photochemistry (*20*) cannot be ruled out, but the small-magnitude $\Delta^{17}O$ signals preclude a unique interpretation (supplementary materials), and a quantification of P_{O_2} is not possible.

The mass-dependent behavior of sulfur and calcium isotopes provides compositional constraints on ancient seawater. Sulfate $\delta^{34}S$ values lie between 5 and 7 per mil (‰), except for the uppermost sample in unit 2 (discussed further in the supplementary materials). Given the small ^{34}S enrichment during sulfate evaporite formation (*10*), the seawater sulfate $\delta^{34}S$ composition is estimated to have been 4 to 6‰ during deposition of units 1 and 2. The homogeneous sulfur isotopic composition across ~400 m of OPH stratigraphy and the composition and sheer volume of evaporite minerals suggest that the OPH evaporites must have derived from seawater and preserve robust, primary isotopic signals. Additionally, the presence of halite and highly soluble bittern salts in unit 1 argues against interaction with large volumes of fluid and supports the interpretation of primary isotopic ratios for the major mineral-forming elements. Calcium isotope ratios show a clear stratigraphic relationship following mineralogical trends, with the highest $\delta^{44/40}Ca$ values in unit 1, decreasing values in unit 2, and even lower values in unit 3. Three samples in unit 3 with the lowest $\delta^{44/40}Ca$ values (–1.6 to –1.4‰, relative to modern seawater) also exhibit relatively high strontium content, with one sample containing minor relict aragonite; these observations indicate that the bulk sediment likely contained primary aragonite that has now largely been converted to other carbonate minerals (supplementary materials). Excluding those samples associated with aragonite, where mineralogy rather than evaporitic processes is the first-order control on calcium isotope ratios, the increase in $\delta^{44/40}Ca$ values in the more highly evaporated facies is consistent with evaporite precipitation driving isotopic distillation of calcium, the magnitude of which is sensitive to the initial composition of seawater (*21*).

An estimate for seawater sulfate concentrations ~2.0 billion years ago can be derived from the observed sequence, mineralogy, and calcium isotope ratios of the OPH evaporites (Fig. 3). Constraints were assessed by comparison with batch evaporation simulations with varying initial ion concentrations. The relative concentrations of calcium and sulfate are the principal controls governing the precipitation sequence that is expressed in the OPH—as in modern evaporites—of carbonates (unit 3) followed by calcium sulfates (unit 2), halite, and eventually magnesium sulfates (unit 1). As such, the OPH preserves a reversed evaporite sequence where the degree of evaporation decreases stratigraphically upward, progressing from the most evolved brine at the base of the Tulomozero Formation toward more open-marine conditions. During a forward evaporite sequence, calcium precipitates

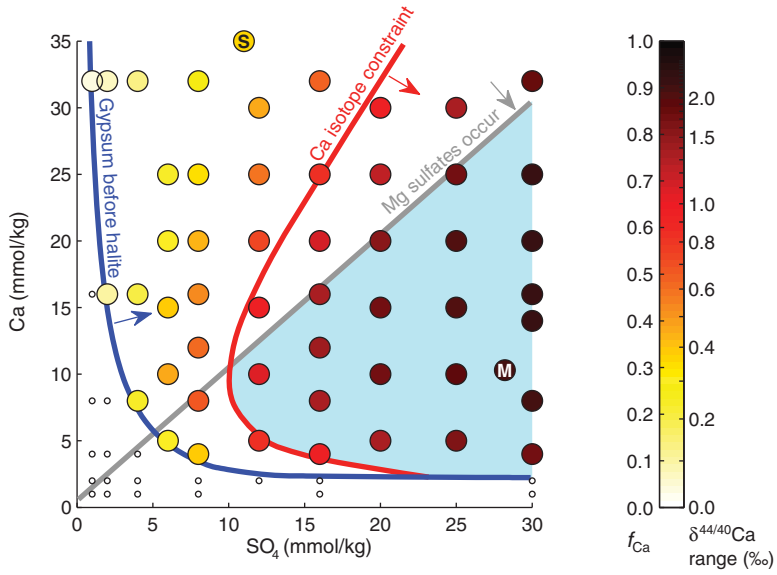


Fig. 3. Constraints on seawater chemistry during deposition of the Tulomozero Formation. Filled circles show results from batch evaporation simulations with variable calcium and sulfate concentrations; all other ions were as in modern seawater. Small open circles indicate failure to precipitate gypsum before halite. The color of filled circles indicates the fraction of initial calcium (f_{Ca}) removed at halite saturation. The conversion of f_{Ca} to $\delta^{44/40}$Ca range is based on a Rayleigh distillation model with $\alpha = 0.99905$ (supplementary materials). M and S identify the compositions of modern and estimated Silurian seawater (*8*), respectively. Lines indicate constraints from OPH observations (arrows give directionality), and the blue shaded region shows the range of seawater compositions consistent with these constraints.

as sulfate minerals and minor carbonate with an isotopic fractionation, so that calcium in the remaining brine becomes enriched in the heavier isotopes through Rayleigh distillation. If sufficient sulfate is present to remove a large fraction of the original calcium content of seawater, later calcium-bearing phases can record large $\delta^{44/40}$Ca enrichments (*21*). The ~1‰ $\delta^{44/40}$Ca range captured in the OPH between shallow-marine carbonates in unit 3 and halite-hosted anhydrite in unit 1 therefore places a lower limit on sulfate concentrations. Together with mineralogical constraints, and assuming modern concentrations of other major ions and a conservative interpretation of $\delta^{44/40}$Ca values (but see the supplementary materials for further discussion), the minimum sulfate concentration consistent with these observations is ~10 mmol/kg (Fig. 3).

The OPH core provides quantitative evidence that marine sulfate concentrations ~2.0 billion years ago were at least a third those of modern seawater. This constraint validates assertions that a large Paleoproterozoic sulfate reservoir existed (*18*, *22*) and increases fourfold the previous lower bound of 2.5 mmol/kg, derived from observing that gypsum evaporites precipitated before halite (*22*, *23*). Although the ancient ocean volume is unknown, a sulfate concentration of 10 mmol/kg in a modern-sized ocean represents an oxidant reservoir equivalent to 23% of the present ocean-atmosphere oxidizing capacity (or 62% of the present atmospheric O_2 inventory). The growth of such a reservoir from a sulfate concentration of <200 μmol/kg in the Neoarchean (*24*) would account for a net redox imbalance of at least 8 to 24×10^{10} mol/year in equivalent moles of O_2 produced or organic carbon buried over 100 to 300 million years. The accumulation of such a sizable fluid oxidant reservoir within the given time constraints, compared with estimates of modern organic carbon burial [5×10^{12} mol/year (*25*)], can be explained by either a large and rapid decline in reductant sinks (i.e., sulfide) or a prolonged 2 to 5% imbalance over 10^8-year time scales. In either case, the observations suggest that a sustained increase in net O_2 production occurred in the Paleoproterozoic.

The geologically rapid growth of a massive sulfate reservoir, with or without a commensurate increase in atmospheric O_2, also has implications for feedbacks between the global biogeochemical cycles of O_2 and CO_2 and Earth's climate. In particular, the oxidation of large amounts of reduced sulfur requires additional sources of carbon to offset the inferred organic carbon burial (the initial source of oxygen) and prevent catastrophic cooling (*26*). Models for Earth's oxidation must balance these considerations, as well as the new evidence for a substantial oxidant reservoir in the form of marine sulfate. Additionally, although sulfate likely represented the largest oxidant reservoir during deposition of the Tulomozero Formation, its concentration subsequently decreased, so that evaporites ~1.9 billion years ago no longer precipitated gypsum before halite (*12*, *27*) and a fundamental change in the sedimentary sulfur isotopic composition occurred (*28*). The apparently transient accumulation of surface oxidants is not yet well understood (*2*, *29*) but implies a protracted reorganization of the global redox budget on the time scales of sedimentary recycling of pyrite and organic carbon (i.e., hundreds of millions of years). Regardless of the mechanisms involved, the observations presented here from the OPH core document a large oxidant pool ~2.0 billion years ago—a pivotal constraint on the history of Earth's oxidation.

REFERENCES AND NOTES

1. D. E. Canfield, *Annu. Rev. Earth Planet. Sci.* **33**, 1–36 (2005).
2. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, *Nature* **506**, 307–315 (2014).
3. H. Holland, in *Early Life on Earth: Nobel Symposium, No. 84* (Columbia Univ. Press, 1994).
4. J. Farquhar, H. Bao, M. Thiemens, *Science* **289**, 756–759 (2000).
5. G. Luo *et al.*, *Sci. Adv.* **2**, e1600134 (2016).
6. A. A. Pavlov, J. F. Kasting, *Astrobiology* **2**, 27–41 (2002).
7. K. Zahnle, M. Claire, D. Catling, *Geobiology* **4**, 271–283 (2006).
8. T. K. Lowenstein, M. N. Timofeeff, S. T. Brennan, L. A. Hardie, R. V. Demicco, *Science* **294**, 1086–1088 (2001).
9. J. Horita, H. Zimmermann, H. D. Holland, *Geochim. Cosmochim. Acta* **66**, 3733–3756 (2002).
10. G. E. Claypool, W. T. Holser, I. R. Kaplan, H. Sakai, I. Zak, *Chem. Geol.* **28**, 199–260 (1980).
11. H. Bao, J. R. Lyons, C. Zhou, *Nature* **453**, 504–506 (2008).
12. M. C. Pope, J. P. Grotzinger, *J. Sediment. Res.* **73**, 280–295 (2003).
13. N. Spear *et al.*, *Geology* **42**, 103–106 (2014).
14. J. F. Lindsay, *Geol. Soc. Am. Bull.* **99**, 852 (1987).
15. L. C. Kah, T. W. Lyons, J. T. Chesley, *Precambrian Res.* **111**, 203–234 (2001).
16. A. F. Morozov *et al.*, *Dokl. Earth Sci.* **435**, 1483–1486 (2010).
17. V. A. Krupenik, K. Y. Sveshnikova, in *The Onega Palaeoproterozoic Structure* (*Geology, Tectonics, Deep Structure and Minerogeny*) (in Russian), L. V. Glushanin, N. V. Sharov, V. V. Shchiptsov, Eds. (Institute of Geology, Karelian Research Centre of the Russian Academy of Sciences, 2011), pp. 190–195.
18. V. A. Melezhik, A. E. Fallick, D. V. Rychanchik, A. B. Kuznetsov, *Terra Nova* **17**, 141–148 (2005).
19. A. T. Brasier, A. E. Fallick, A. R. Prave, V. A. Melezhik, A. Lepland, *Precambrian Res.* **189**, 193–211 (2011).
20. X. Cao, H. Bao, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14546–14550 (2013).
21. C. L. Blättler, J. A. Higgins, *Geology* **42**, 711–714 (2014).
22. M. Reuschel *et al.*, *Precambrian Res.* **192–195**, 78–88 (2012).
23. S. Schröder, A. Bekker, N. J. Beukes, H. Strauss, H. S. van Niekerk, *Terra Nova* **20**, 108–117 (2008).
24. K. S. Habicht, M. Gade, B. Thamdrup, P. Berg, D. E. Canfield, *Science* **298**, 2372–2374 (2002).
25. R. A. Berner, D. E. Canfield, *Am. J. Sci.* **289**, 333–361 (1989).
26. A. Bachan, L. R. Kump, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6562–6567 (2015).
27. J. P. Grotzinger, J. F. Kasting, *J. Geol.* **101**, 235–243 (1993).
28. N. J. Planavsky, A. Bekker, A. Hofmann, J. D. Owens, T. W. Lyons, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18300–18305 (2012).
29. A. Bekker, H. D. Holland, *Earth Planet. Sci. Lett.* **317–318**, 295–304 (2012).
30. V. A. Krupenik, A. M. Akhmedov, K. Y. Sveshnikova, in *The Onega Palaeoproterozoic Structure* (*Geology, Tectonics, Deep Structure and Minerogeny*) (in Russian), L. V. Glushanin, N. V. Sharov, V. V. Shchiptsov, Eds. (Institute of Geology, Karelian Research Centre of the Russian Academy of Sciences, 2011), pp. 250–255.

ACKNOWLEDGMENTS

Thanks to T. H. Bui for assistance with sulfur isotope analyses and B. A. Wing for helpful discussions. This work was supported by a grant from the Simons Foundation (SCOL 339006 to C.L.B.) and by the European Research Council (ERC Horizon 2020 grant 678812 to M.C.), the Research Council of Norway (RCN Centres of Excellence funding scheme project 223259 to K.P. and A.L.), the Estonian Science Agency (PUT696 to K.K., A.L., K.P., and T.K.), and Princeton University (to J.A.H.). Core material from the OPH is maintained by the Institute of Geology, Karelian Research Centre, Petrozavodsk, Russia. Data presented in this study are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/320/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Table S1
References (*31*–*79*)
Data S1 and S2

19 October 2017; accepted 7 March 2018
Published online 22 March 2018
10.1126/science.aar2687

STRUCTURAL BIOLOGY

Structure of a prehandover mammalian ribosomal SRP·SRP receptor targeting complex

Kan Kobayashi,[1]* Ahmad Jomaa,[1]* Jae Ho Lee,[2] Sowmya Chandrasekar,[2] Daniel Boehringer,[1] Shu-ou Shan,[2] Nenad Ban[1]†

Signal recognition particle (SRP) targets proteins to the endoplasmic reticulum (ER). SRP recognizes the ribosome synthesizing a signal sequence and delivers it to the SRP receptor (SR) on the ER membrane followed by the transfer of the signal sequence to the translocon. Here, we present the cryo–electron microscopy structure of the mammalian translating ribosome in complex with SRP and SR in a conformation preceding signal sequence handover. The structure visualizes all eukaryotic-specific SRP and SR proteins and reveals their roles in stabilizing this conformation by forming a large protein assembly at the distal site of SRP RNA. We provide biochemical evidence that the guanosine triphosphate hydrolysis of SRP·SR is delayed at this stage, possibly to provide a time window for signal sequence handover to the translocon.

In eukaryotes, nascent secretory and membrane proteins are cotranslationally targeted to the endoplasmic reticulum (ER) membrane by the universally conserved signal recognition particle (SRP) and its receptor (SR) (*1*–*3*). SRP recognizes the N-terminal signal sequence of the nascent chain on ribosomes synthesizing membrane or secretory proteins (*4*–*6*). Subsequently, through interactions with the membrane-anchored SR, the ribosome–nascent chain complex (RNC) is delivered to the Sec translocon (Sec61p) on the ER membrane (*7*, *8*).

The eukaryotic SRP targeting machineries are considerably more complex than their bacterial counterparts (*9*, *10*). The eukaryotic SRP contains a larger SRP RNA (7SL for mammals) composed of S and Alu domains (*11*) and six eukaryotic proteins (SRP9, SRP14, SRP19, SRP54, SRP68, and SRP72) (*1*), out of which only SRP54 is universally conserved. SRP54 binds the ribosomal tunnel exit and the signal sequence through its NG (*12*) and M domains (*13*), respectively. Eukaryotic SR is a heterodimer of SRα and eukaryotic-specific SRβ integrated into the ER membrane (*14*, *15*). SRα contains a universally conserved NG domain, a flexible linker, and a eukaryotic-specific SRX domain that complexes with SRβ (*16*). The NG domains of SRP54 and SRα interact in a guanosine triphosphate (GTP)–dependent manner to form the NG heterodimer (*17*, *18*), thus delivering translating ribosomes to the ER.

By analogy to the bacterial SRP pathway (*19*, *20*), it is assumed that the GTP-bound NG heterodimer of SRP·SR relocates from the ribosomal tunnel exit to the distal region of the SRP RNA where GTP is hydrolyzed. This conformational change might be necessary for the attachment and signal sequence handover to the Sec61p translocon (*18*, *21*). Both SRP68 and SRP72 proteins are likely to be involved in distal site interactions and GTP hydrolysis based on their positions on the SRP RNA (*18*, *22*) and because chemical modification of these proteins severely represses membrane targeting activity (*23*). Previous studies also suggested roles of the eukaryotic-specific SR components SRX and SRβ in guanosine triphosphatase (GTPase) regulation (*24*) and signal sequence handover to the Sec61p translocon (*25*, *26*), respectively. However, the implications of the mechanistic and architectural differences between the bacterial and eukaryotic SRP systems have been difficult to rationalize. We set out to obtain structural information on the complete eukaryotic protein targeting complex on the translating ribosome.

The mammalian SRP·SR·RNC targeting complex was assembled in the presence of 5′-guanylyl imidodiphosphate (GMPPNP), and its structure was determined using single-particle cryo–electron microscopy (cryo-EM) at 3.7-Å resolution on average (Fig. 1 and figs. S1 and S2). The densities of SRP and SR were resolved around 4.5- to 10-Å resolution (fig. S2, C and D), and the SRP RNA is visualized in its entirety (Fig. 1A and fig. S2C). At the Alu domain, the SRP RNA and SRP9/14 contact ribosomal subunit interface to arrest translation as previously observed (*27*) (fig. S3, A and B). In the S domain, the ribosomal tunnel exit has the density for the SRP54 M domain bound to the signal sequence (Fig. 1B and figs. S3, C and D, and S4A) but lacks that for its NG domain (Fig. 1, A and B), consistent with previous cryo-EM structures of the SRP·SR·RNC complex (*21*, *28*, *29*). Instead, a large density is visible at the distal region of the SRP RNA (Fig. 1A), where we can dock previously reported crystal structures of SRP and SR proteins (Fig. 1B and figs. S4 and S5) (*18*, *22*, *30*, *31*). This structure reveals an architecture of the entire mammalian SRP·SR·RNC complex (Fig. 1, B and C).

Compared with the structure of the SRP·RNC complex without the receptor (*27*), the SRP RNA lifts away from the ribosomal tunnel exit by ~12 Å (fig. S6), which would facilitate the transfer of the signal sequence to the Sec61p translocon. At the distal site of the SRP RNA, the NG heterodimer binds to the SRP RNA at a different angle compared with its bacterial counterpart (*29*) (fig. S7), leading to additional contacts with the SRP RNA, the ribosome, the SRP proteins, and SR (Fig. 2 and fig. S8, A to C). The linker between NG and M domains (GM-linker) of SRP54 connecting its NG and M domains, which forms a helix in the bacterial complex, is flexibly disposed and is not visible in the map (Fig. 1B and fig. S7). The resolution of the SRP RNA at the contact site with the NG heterodimer is ~4.5 Å (fig. S2C), resolving the phosphate backbone of the SRP RNA along with a flipped-out base stacking with Phe^{456} of the NG domain of SRα in the GTPase active site (figs. S2D and S9). The position of this flipped-out base corresponds to that of the universally conserved G232 (bacterial G83) (*22*, *31*, *32*). The interactions between the SRα NG domain and the SRP RNA, including the contacts formed by the flipped-out base and the surrounding helices of the NG domain, are similar to the interactions seen for the GTPase-activated bacterial complex (*32*) (figs. S8C and S9). The SRα NG domain also interacts with the SRP68 RBD (Fig. 2 and fig. S8C), whereas the extended loop of SRP68 RBD makes contact with the 28*S* ribosomal RNA (rRNA) (fig. S8D).

In the mammalian targeting complex, the NG heterodimer at the SRP RNA distal site is locked in position by extensive interactions with SRX·SRβ, which bridges the SRP68 RBD and the NG heterodimer (Figs. 1, B and C, and 2, and figs. S4, C and D, and S7). The cytosolic domain of SRβ belongs to the Ras GTPase superfamily (*14*) and forms a heterodimer with SRX in the GTP-bound form (*33*, *34*). The interaction interface of SRβ to SRX has similarity to the binding site of Ras to its GTPase activator GAP-334, although SRX alone does not activate the GTPase of SRβ (*30*). In the structure of the targeting complex, SRα NG further extends the interaction interface of SRβ in a similar manner as observed in the structure of the Ras·GAP-334 complex (*35*) (fig. S10), but, in contrast to GAP-334, no elements from either SRX or the SRα NG domain are in direct contact with the SRβ-bound GTP (fig. S10). It appears that SRβ does not hydrolyze GTP even in this conformation and that GTP-bound SRβ may rather provide the eukaryotic-specific stabilizing effect on the NG heterodimer binding at the SRP RNA distal site. Thus, SRX·SRβ and SRP68 RBD form a platform to stably dock the NG heterodimer at the distal site of SRP RNA to enable signal sequence handover to the Sec61p translocon at the exposed ribosomal tunnel exit.

[1]Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Otto-Stern-Weg 5, Zurich CH-8093, Switzerland. [2]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: ban@mol.biol.ethz.ch

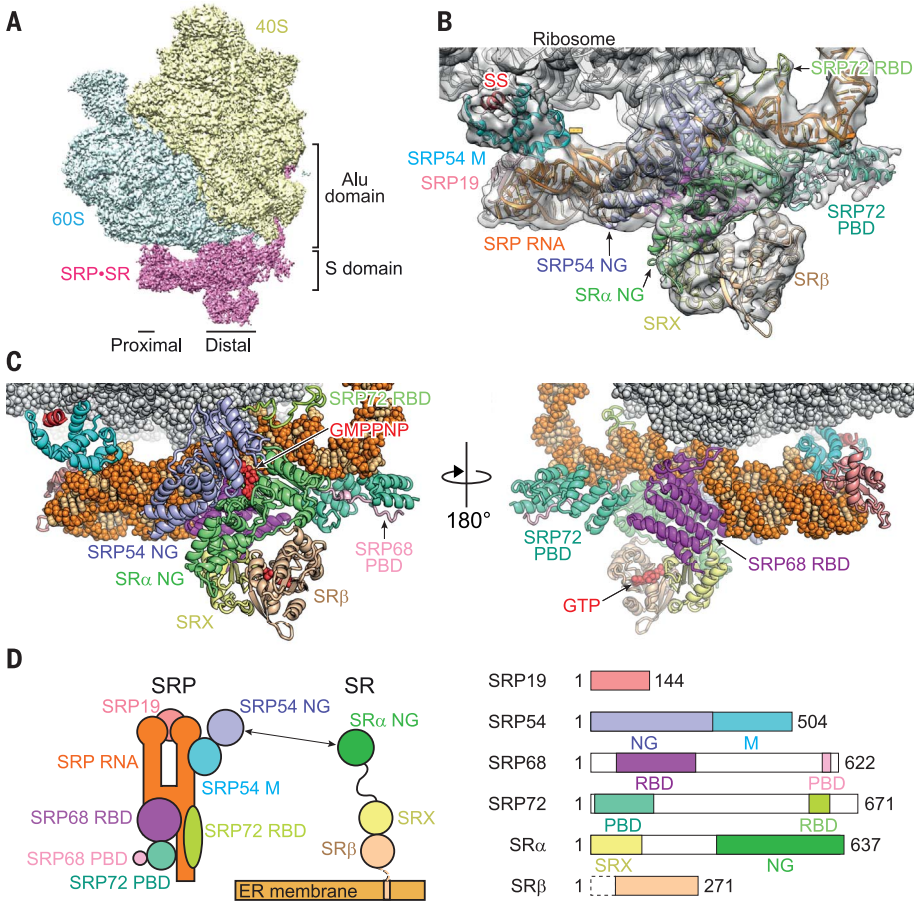


Fig. 1. Architecture of the mammalian SRP·SR·RNC complex. (**A**) The cryo-EM map of SRP·SR·RNC complex low-pass filtered to 3.7-Å resolution is shown. SRP·SR and ribosomal 40*S* and 60*S* subunits are colored magenta, yellow, and light blue, respectively. (**B**) Overall structure model of SRP S domain bound to SR on RNC fitted into the map. The cryo-EM map low-pass filtered to 6-Å resolution is colored gray. Models are shown in the cartoon. The ribosome and SRP RNA are colored gray and orange, respectively. The signal sequence (SS) is colored red, and protein components of SRP and SR are colored based on their domain architectures [SRP19, salmon; SRP54 NG, light blue; SRP54 M, cyan; SRP68 RNA binding domain (RBD), magenta; SRP68 protein binding domain (PBD), light pink; SRP72 PBD, green cyan; SRP72 RBD, limon; SRα SRX, yellow; SRα NG, lime; SRβ, wheat]. (**C**) The structure model of SRP S domain and SR on the RNC from two opposite directions. The ribosome and SRP RNA are shown as spheres, and the signal sequence and protein components of SRP and SR are shown in the cartoon. Models are colored as in Fig. 1B, but bases of SRP RNA are colored light orange. The two GMPPNP molecules bound to NG heterodimer and a GTP molecule bound to SRβ are shown as red spheres. The positions of these molecules are based on the previous crystal structures (PDB ID: 5L3Q and 2FH5) (*18*, *30*). (**D**) Schematic diagram of SRP and SR (left) and the domain architecture of SRP and SR proteins (right). SRP and SR components are colored as in Fig. 1B. The interaction between NG domains of SRP54 and SRα is depicted in an arrow. In the right panel, regions shown as blank boxes are not modeled in the structure. The N terminus of SRβ, shown as a blank box with a dashed line, depicts the luminal and transmembrane regions and is not included in the purified construct of SR.

At the distal site of SRP RNA, the SRP72 RBD can be observed extending from the SRP RNA to the ribosome surface, where it approaches a hairpin of the ribosomal protein uL3 (fig. S4F). The C-terminal region of SRP72 RBD was proposed to form an alpha helix (C4 helix) and mediate interactions with the 28*S* rRNA (C4 contact) (fig. S11A) (*22*). However, in our map, it appears that the C-terminal region of the SRP72 RBD rather extends toward the SRP·SR NG heterodimer interface and interacts with the two stacking residues in the GTPase active site, G232 of the SRP RNA and Phe^{456} of SRα (Fig. 3, A and B, and figs. S9A and S11B). Thus, a eukaryotic-specific GTPase regulation mechanism may involve a protein component of the SRP (SRP72) in addition to the SRP RNA. To determine whether mutations in SRP72 affect the GTPase activity of the targeting complex, we prepared SRP variants harboring mutation or deletion in the SRP72 C-terminal region containing the C4 helix (Arg^{589}-Gln^{603} based on the human numbering) (fig. S12A). All tested SRP mutants were active in translation arrest and membrane targeting in vitro (fig. S13), indicating that they were not defective in ribosome interactions. GTPase assays of SRP·SRαβΔTM (lacking nonessential luminal and transmembrane regions of SRβ) complexes assembled with the ribosome and signal sequence showed that, unexpectedly, all the SRP variants exhibited enhanced GTPase activity compared with wild-type SRP (Fig. 3C and fig. S14). Enhanced GTPase activity was also observed when we modified SRP RNA by mutating the flipped-out G232 into A (G232A) or by closing the SRP RNA 5f-loop (ΔA231, G232U) (Fig. 3D and figs. S12B and S14), and the highest GTPase activity was observed by combining the mutations on the SRP RNA and SRP72 (Fig. 3D and fig. S14). In addition, an SRX·SRβ–deletion mutant of SR (SRαΔX) also exhibited a two-fold more enhanced GTPase activity than SRαβΔTM (Fig. 3C). Thus, eukaryotic-specific interactions at the SRP RNA distal site delay GTP hydrolysis, in contrast to the bacterial system, where distal site interactions between SRP RNA and NG heterodimer stimulate GTP hydrolysis (*20*, *32*). The delay of GTP hydrolysis by the SRP·SR NG heterodimer at the distal site of the SRP RNA may provide a eukaryotic-specific regulatory role. Interestingly, it was reported that the C-terminal region of SRP72 is cleaved during apoptosis, further supporting its possible role in regulating the targeting process (*22*, *36*). We propose that the current structure represents a "prehandover" complex of SRP·SR at the SRP RNA distal site, which allows docking of the Sec61p translocon onto the exposed ribosomal tunnel exit. Delayed GTP hydrolysis in this state might provide a longer time window for the nascent polypeptide to engage the Sec61p translocon, which was proposed to be slow in eukaryotic systems (*37*), before the detachment of SRP and SR from RNC. Indeed, the mammalian Sec61p translocon reduces the GTPase activity of SRP·SR in the presence of RNC (*38*).

In conclusion, this study reveals the roles of eukaryotic-specific components of the SRP and SR proteins in the mammalian targeting process. This process starts by the recognition of the signal sequence by SRP at the exit of the ribosomal tunnel and formation of the SRP54·SRα NG heterodimer in the presence of GTP (Fig. 4, A and B). The NG heterodimer then relocates to the SRP RNA distal site, where it forms a large complex stabilized by interactions with the eukaryotic-specific components of the SRP and SR together with SRP RNA (Fig. 4C). At this stage, the GTP hydrolysis of the NG heterodimer is delayed to provide a time window for the signal sequence handover from SRP to the Sec61p translocon at the exit of the ribosomal tunnel.

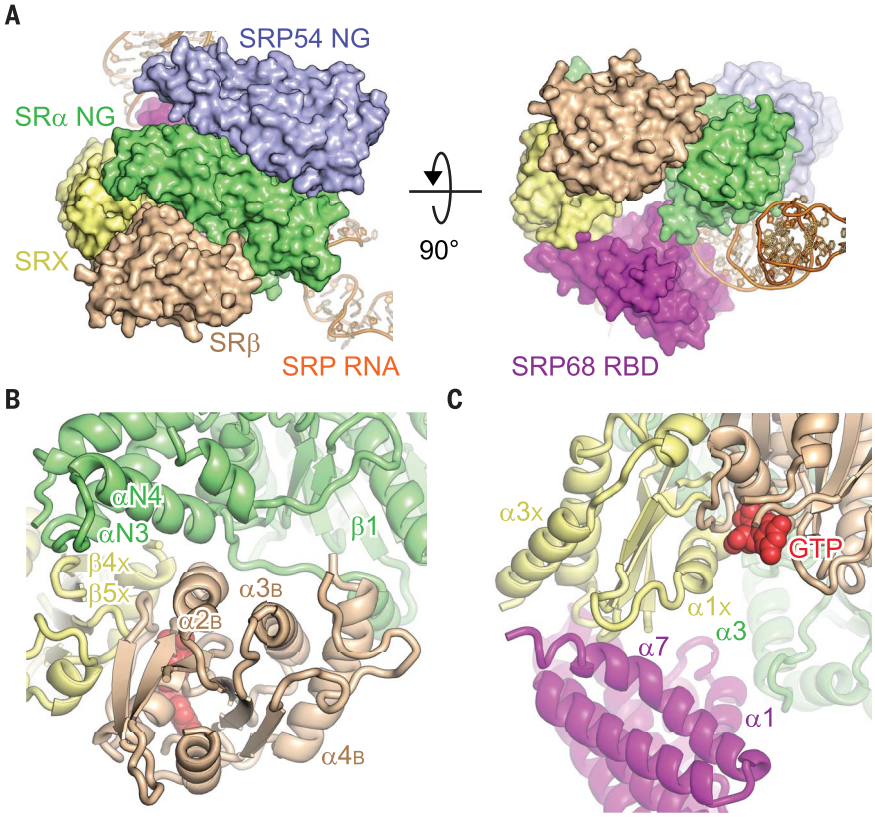


Fig. 2. Assembly of the NG heterodimer, SRP68 RBD, and SRX·SRβ at the SRP RNA distal site. (**A**) Surface representation of the NG heterodimer, SRP68 RBD, SRX·SRβ at the SRP RNA distal site from two perpendicular directions. SRP RNA is shown in the cartoon. Other components are omitted for clarity. Models are colored as in Fig. 1B. (**B** and **C**) The interaction between NG heterodimer, SRP68 RBD, and SRX·SRβ. Proteins are shown in the cartoon and colored as in Fig. 1B. A GTP molecule bound to SRβ is also shown as in Fig. 1C. Secondary structure elements of proteins are labeled.

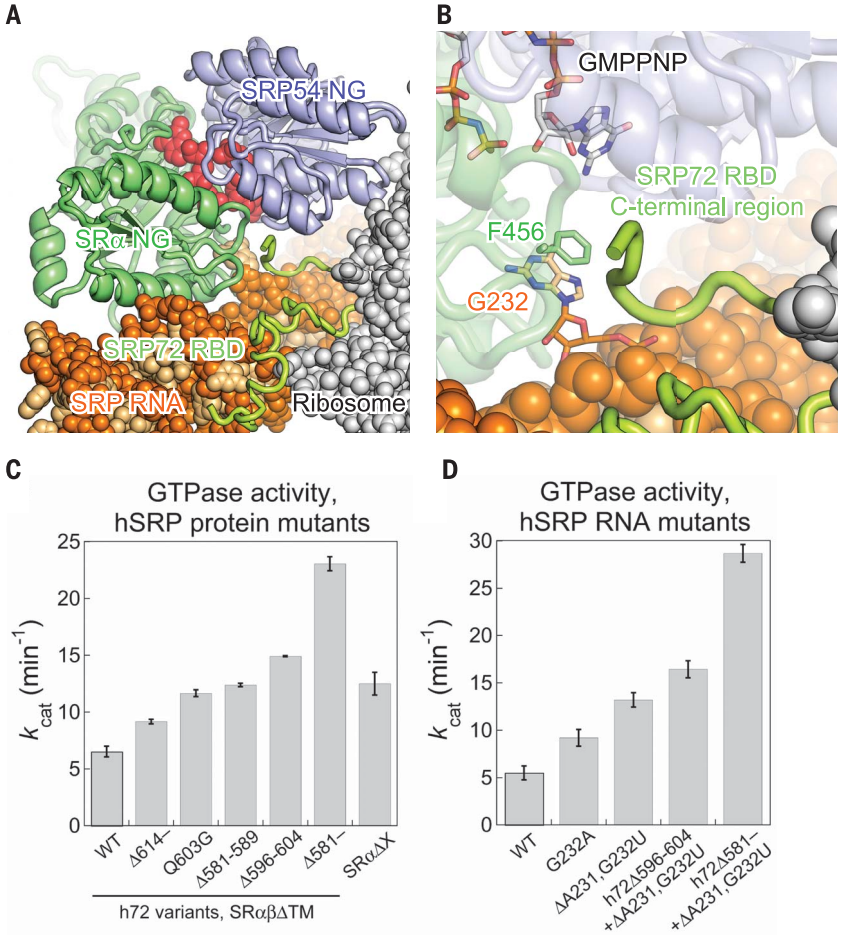


Fig. 3. The GTPase regulation of mammalian SRP·SR on RNC. (**A**) The structure model of the NG heterodimer GTPase active site shown as in Fig. 1C. Components other than the NG heterodimer, SRP RNA, SRP72 RBD, and ribosome are omitted for clarity. (**B**) Close-up view of (A). Two GMPPNP molecules, Phe^{456} of SRα NG, and flipped-out G232 of SRP RNA are shown in the stick model. (**C**) GTPase rate constants of the SRP·SR complex formed by wild-type human SRP (hSRP) with SRαβΔTM, SRPs bearing indicated SRP72 (h72) mutations with SRαβΔTM, and wild-type SRP with SRαΔX. All measurements were carried out with SRP54 fused to a model signal sequence and with the 80S ribosome present. The values of k_{cat} were derived using analysis as shown in fig. S14. See the supplementary materials for details. (**D**) GTPase rate constants of the SRP·SR complex containing wild-type human SRP and SRPs bearing indicated SRP RNA mutations. The last two columns show the GTPase rate constants with SRPs containing mutations in both SRP72 (h72) and SRP RNA. All measurements used SRαβΔTM and were carried out with SRP54 fused to a model signal sequence and with the 80S ribosome present. The values of k_{cat} were derived using analysis as shown in fig. S14. Data were reported as mean ± SD, with n = 3 to 6 in (C) and (D).

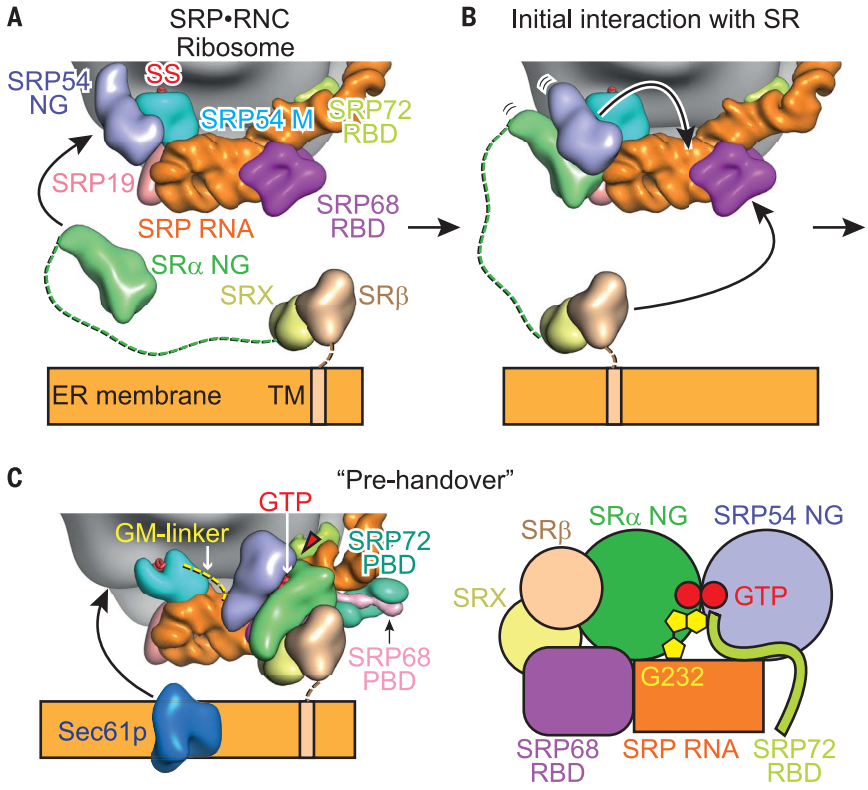


Fig. 4. Schematic diagram of the mammalian targeting system. (**A**) In the SRP·RNC complex, the SRP54 NG domain occupies the ribosomal tunnel exit. SR is anchored to the ER membrane through the transmembrane helix (TM) of SRβ, whose cytosolic domain binds the SRX domain of SRα connected to the NG domain through the long linker. (**B**) The SRP54 NG domain initially binds SRα at the ribosomal tunnel exit, forming the NG heterodimer in the presence of GTP, which then relocates to the distal site of SRP RNA. (**C**) At the distal site of SRP RNA, SRX·SRβ together with SRP68 RBD accommodates the NG heterodimer. The GTPase active site of the NG heterodimer is pointed by the red triangle (left panel). The schematic diagram of the GTPase active site conformation is shown in the right panel. At this stage, the GTP hydrolysis is delayed by SRX·SRβ, SRP72 RBD, and SRP RNA, possibly to keep RNC on the membrane with its tunnel exit exposed for the handover of the signal sequence from SRP to the Sec61p translocon.

REFERENCES AND NOTES

1. P. Walter, G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7112–7116 (1980).
2. R. Gilmore, P. Walter, G. Blobel, *J. Cell Biol.* **95**, 470–477 (1982).
3. D. I. Meyer, E. Krause, B. Dobberstein, *Nature* **297**, 647–650 (1982).
4. J. W. Chartron, K. C. L. Hunt, J. Frydman, *Nature* **536**, 224–228 (2016).
5. P. Walter, I. Ibrahimi, G. Blobel, *J. Cell Biol.* **91**, 545–550 (1981).
6. U. C. Krieg, P. Walter, A. E. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8604–8608 (1986).
7. S. M. Simon, G. Blobel, *Cell* **65**, 371–380 (1991).
8. R. M. Voorhees, R. S. Hegde, *Curr. Opin. Cell Biol.* **41**, 91–99 (2016).
9. Y. Nyathi, B. M. Wilkinson, M. R. Pool, *Biochim. Biophys. Acta* **1833**, 2392–2402 (2013).
10. E. C. Mandon, S. F. Trueman, R. Gilmore, *Cold Spring Harb. Perspect. Biol.* **5**, a013342 (2013).
11. P. Walter, G. Blobel, *Nature* **299**, 691–698 (1982).
12. M. Halic *et al.*, *Nature* **427**, 808–814 (2004).
13. H. Lütcke, S. High, K. Römisch, A. J. Ashford, B. Dobberstein, *EMBO J.* **11**, 1543–1551 (1992).
14. J. D. Miller, S. Tajima, L. Lauffer, P. Walter, *J. Cell Biol.* **128**, 273–282 (1995).
15. S. Tajima, L. Lauffer, V. L. Rath, P. Walter, *J. Cell Biol.* **103**, 1167–1178 (1986).
16. J. C. Young *et al.*, *J. Biol. Chem.* **270**, 15650–15657 (1995).
17. P. J. Rapiejko, R. Gilmore, *Cell* **89**, 703–713 (1997).
18. K. Wild *et al.*, *J. Mol. Biol.* **428**, 2880–2897 (2016).
19. S. F. Ataide *et al.*, *Science* **331**, 881–886 (2011).
20. K. Shen, S. Arslan, D. Akopian, T. Ha, S. O. Shan, *Nature* **492**, 271–275 (2012).
21. M. Halic *et al.*, *Science* **312**, 745–747 (2006).
22. M. M. Becker, K. Lapouge, B. Segnitz, K. Wild, I. Sinning, *Nucleic Acids Res.* **45**, 470–481 (2017).
23. V. Siegel, P. Walter, *Cell* **52**, 39–49 (1988).
24. E. C. Mandon, Y. Jiang, R. Gilmore, *J. Cell Biol.* **162**, 575–585 (2003).
25. G. Bacher, M. Pool, B. Dobberstein, *J. Cell Biol.* **146**, 723–730 (1999).
26. T. A. Fulga, I. Sinning, B. Dobberstein, M. R. Pool, *EMBO J.* **20**, 2338–2347 (2001).
27. R. M. Voorhees, R. S. Hegde, *eLife* **4**, e07975 (2015).
28. A. Jomaa, D. Boehringer, M. Leibundgut, N. Ban, *Nat. Commun.* **7**, 10471 (2016).
29. A. Jomaa *et al.*, *Nat. Commun.* **8**, 15470 (2017).
30. O. Schlenker, A. Hendricks, I. Sinning, K. Wild, *J. Biol. Chem.* **281**, 8898–8906 (2006).
31. J. T. Grotwinkel, K. Wild, B. Segnitz, I. Sinning, *Science* **344**, 101–104 (2014).
32. F. Voigts-Hoffmann *et al.*, *Mol. Cell* **52**, 643–654 (2013).
33. K. R. Legate, D. Falcone, D. W. Andrews, *J. Biol. Chem.* **275**, 27439–27446 (2000).
34. T. Schwartz, G. Blobel, *Cell* **112**, 793–803 (2003).
35. K. Scheffzek *et al.*, *Science* **277**, 333–338 (1997).
36. P. J. Utz *et al.*, *J. Biol. Chem.* **273**, 35362–35370 (1998).
37. Z. Cheng, R. Gilmore, *Nat. Struct. Mol. Biol.* **13**, 930–936 (2006).
38. G. Bacher, H. Lütcke, B. Jungnickel, T. A. Rapoport, B. Dobberstein, *Nature* **381**, 248–251 (1996).

ACKNOWLEDGMENTS

Cryo-EM data was collected at the Scientific Center for Optical and Electron Microscopy at the ETH Zurich (ScopeM). We thank P. Tittmann and A. Scaiola for support with data collection and C. Aylett and M. Itten for support with data processing. We also thank P. Bieri for initial help with microsome preparation and structure model building. We are grateful for M. Leibundgut for the critical comments on our structure model and manuscript. **Funding:** K.K. was supported by Long-Term Fellowships from Toyobo Biotechnology Foundation and European Molecular Biology Organization (EMBO) (ALTF 660-2015). This study was supported by the Swiss National Science Foundation (SNSF) (grant number 310030B_163478), National Center of Excellence in Research (NCCR) RNA & Disease Program of the SNSF (grant number 51NF40_141735) (to N.B.), NIH grant GM107368, Gordon and Betty Moore Foundation grant GBMF2939, and a fellowship from the Weston Havens Foundation (to S.-o.S.). **Author contributions:** K.K., A.J., S.-o.S., and N.B. designed the experimental strategy. K.K. prepared samples and grids for the cryo-EM analysis. K.K., A.J., and D.B. collected cryo-EM data. K.K. and A.J. processed data, built the structure model, and refined it. J.H.L. and S.C. prepared samples for the functional assays and performed them. K.K. and A.J. wrote the initial draft of the manuscript, and all authors contributed to the final version. **Competing interests:** None declared. **Data and materials availability:** The atomic coordinates have been deposited in the Protein Data Bank (PDB ID: 6FRK). The cryo-EM density map has been deposited in Electron Microscopy Data Bank (EMD-4300).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/323/suppl/DC1
Materials and Methods
Figs. S1 to S14
Table S1
References (*39*–*64*)

18 December 2017; accepted 12 March 2018
Published online 22 March 2018
10.1126/science.aar7924

AUTISM GENOMICS

Paternally inherited cis-regulatory structural variants are associated with autism

William M. Brandler,[1,2,3,4]* Danny Antaki,[1,2,3,5]* Madhusudan Gujral,[1,2,3]* Morgan L. Kleiber,[1,2,3] Joe Whitney,[6] Michelle S. Maile,[1,2,3] Oanh Hong,[1,2,3] Timothy R. Chapman,[1,2,3] Shirley Tan,[1,2,3] Prateek Tandon,[1,2,3] Timothy Pang,[1,7] Shih C. Tang,[1,7] Keith K. Vaux,[8] Yan Yang,[9] Eoghan Harrington,[9] Sissel Juul,[9] Daniel J. Turner,[10] Bhooma Thiruvahindrapuram,[6] Gaganjot Kaur,[6] Zhuozhi Wang,[6] Stephen F. Kingsmore,[11] Joseph G. Gleeson,[12] Denis Bisson,[4] Boyko Kakaradov,[4] Amalio Telenti,[4] J. Craig Venter,[4,13] Roser Corominas,[14,15] Claudio Toma,[16,17,18] Bru Cormand,[15,16,19,20] Isabel Rueda,[21] Silvina Guijarro,[22] Karen S. Messer,[23] Caroline M. Nievergelt,[2] Maria J. Arranz,[24] Eric Courchesne,[25] Karen Pierce,[25] Alysson R. Muotri,[3] Lilia M. Iakoucheva,[2] Amaia Hervas,[22] Stephen W. Scherer,[6,26,27] Christina Corsello,[2,7] Jonathan Sebat[1,2,3]†

The genetic basis of autism spectrum disorder (ASD) is known to consist of contributions from de novo mutations in variant-intolerant genes. We hypothesize that rare inherited structural variants in cis-regulatory elements (CRE-SVs) of these genes also contribute to ASD. We investigated this by assessing the evidence for natural selection and transmission distortion of CRE-SVs in whole genomes of 9274 subjects from 2600 families affected by ASD. In a discovery cohort of 829 families, structural variants were depleted within promoters and untranslated regions, and paternally inherited CRE-SVs were preferentially transmitted to affected offspring and not to their unaffected siblings. The association of paternal CRE-SVs was replicated in an independent sample of 1771 families. Our results suggest that rare inherited noncoding variants predispose children to ASD, with differing contributions from each parent.

Microarray and exome sequencing studies over the past decade have demonstrated that de novo protein-altering variants contribute to ~25% of cases of autism spectrum disorder (ASD) (*1*, *2*). Much of the allelic spectrum of ASD genetics has been unexplored, particularly variants that lie outside of protein coding sequences of genes. Recent studies have made great progress in identifying regulatory elements throughout the genome (*3*, *4*). The next challenge is to identify ASD risk variants affecting genetic regulatory elements. However, deleterious cis-regulatory variants are not easily distinguishable from the vast background of neutral variation in the genome. Therefore, initial applications of whole-genome sequencing (WGS) in ASD have so far been underpowered to detect the association of rare cis-regulatory single-nucleotide variants (SNVs) with ASD (*5*–*7*).

Structural variants (SVs), such as deletions, duplications, insertions, and inversions (*8*), are more likely than SNVs to affect gene regulation because of their potential to disrupt or rearrange functional elements in the genome. Recent WGS efforts led by the 1000 Genomes Consortium and our group have revealed thousands of rare SVs in each genome that were previously undetectable with microarray or exome sequencing technologies (*8*, *9*).

Here, we investigate the contribution of cis-regulatory SVs (CRE-SVs) to autism in three stages: (i) selection of target functional categories based on evidence of SV-intolerance; (ii) association tests of cis-regulatory elements in a primary WGS data set; and (iii) preregistered replication in an independent cohort.

Our discovery data set consisted of whole-genome sequencing (mean coverage = 42.6X) of 829 families, comprising 880 affected individuals, 630 unaffected individuals, and their parents (table S1). A majority of the subjects in the discovery sample were selected on the basis that they had previously screened negative for de novo loss of function mutations or large copy number variants from exome sequencing (*2*) and microarray (*10*) studies. The ascertainment of this sample was therefore designed to eliminate the well-established categories of genetic risk and thereby to enrich for novel inherited and noncoding risk variants.

We developed a pipeline for genome-wide analysis of SV that consisted of complementary methods for SV discovery (fig. S1). A key innovation was the development of SV^2, a support-vector machine–based software for accurately estimating genotype likelihoods from short-read WGS data, which enabled accurate genotyping of SVs in families with a detection limit of ≥100 base pairs (bp) (*11*). An average of 3746 SVs were detected per individual, including biallelic deletion, tandem duplications, inversions, four classes of complex SV, and four families of mobile element insertion (summarized in figs. S2 and S3 and table S2). The overall false discovery rate (FDR) was estimated from Illumina 2.5 M single-nucleotide polymorphism (SNP) array data to be 4.2% for deletions and 9.4% for duplications (fig. S4 and table S3). SVs were also validated through nanopore WGS of three individuals at a mean coverage of 7X to 9X (table S3). Private deletions and duplications >100 bp in length displayed low Mendelian error rates and 50% transmission to offspring (fig. S4).

Measures of functional constraint that are based on population data are useful metrics for predicting the pathogenicity of rare variants. For example, genes that display strong negative selection against loss-of-function variants in the general population, as assessed by the Exome Aggregation Consortium (ExAC) (*12*), are highly enriched in de novo mutations in children with ASD (*13*), and the vast majority of known autism genes display loss-of-function intolerance scores (pLI) above the 90th percentile for all genes [odds ratio (OR) = 17.6; Fisher's exact $P = 7.3 \times 10^{-30}$] (table S4 and fig. S5). Furthermore, we show here

[1]Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, CA 92093, USA. [2]Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA. [3]Department of Cellular and Molecular Medicine and Pediatrics, University of California San Diego, La Jolla, CA 92093, USA. [4]Human Longevity, Inc., San Diego, CA 92121, USA. [5]Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093, USA. [6]The Centre for Applied Genomics, Genetics, and Genome Biology, The Hospital for Sick Children, Toronto, Canada. [7]Rady Children's Hospital, San Diego, CA 92123, USA. [8]Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA. [9]Oxford Nanopore Technologies, Inc., NY 10013, USA. [10]Oxford Nanopore Technologies Ltd., Oxford, UK. [11]Rady Children's Institute for Genomic Medicine, Rady Children's Hospital, San Diego, CA 92123, USA. [12]Howard Hughes Medical Institute, Rady Children's Institute of Genomic Medicine, Department of Neurosciences, University of California San Diego, La Jolla, CA 92093, USA. [13]J. Craig Venter Institute, La Jolla, CA 92037, USA. [14]Genetics Unit, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain. [15]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain. [16]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Catalonia, Spain. [17]Neuroscience Research Australia, Sydney, Australia. [18]School of Medical Sciences, University of New South Wales, Sydney, Australia. [19]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Catalonia, Spain. [20]Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues, Catalonia, Spain. [21]Department of Psychiatry, Hospital Sant Joan de Déu, Barcelona, Spain. [22]Child and Adolescent Mental Health Unit, Hospital Universitari Mútua de Terrassa, Barcelona, Spain. [23]Division of Biostatistics and Bioinformatics, Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA 92093, USA. [24]Research Laboratory Unit, Fundacio Docencia I Recerca Mutua Terrassa, Barcelona, Spain. [25]Autism Center of Excellence, Department of Neuroscience, University of California San Diego, La Jolla, CA 92093, USA. [26]Department of Molecular Genetics, University of Toronto, Toronto, Canada. [27]McLaughlin Centre, University of Toronto, Toronto, Canada.
*These authors contributed equally to this work.
†Corresponding author. Email: jsebat@ucsd.edu

that the intolerance of genes to exonic deletions is correlated with the SNP-based pLI measure of functional constraint (Fig. 1, A and B).

We reasoned, therefore, that SV intolerance would be a valid criterion for defining categories of functional elements to be tested for disease association in this study. As our measure of SV intolerance, we tested the observed depletion of SVs within functional elements relative to random distributions of SVs generated by two types of permutation (*14*), one in which SVs were shuffled throughout the genome randomly and a second based on a model in which the correlation of SVs to genome features (GC_content of the DNA sequence, coverage, low-complexity repetitive elements, and segmental duplications) was accounted for (*15*). SV depletion was assessed in functional elements grouped by categories such as exons, untranslated regions (UTRs), promoters, cis-regulatory RNAs, enhancers, and evolutionarily conserved and human accelerated regions (28 categories in total, described in table S5). SVs were each assigned to a single category according to the order listed above; for example a SV that disrupts an exon, a UTR, and an enhancer simultaneously would be classified as "exonic." Genes were also defined in advance as "intolerant," based on an ExAC pLI score > 90th percentile (fig. S5). SV depletion was tested for the 28 categories, and analysis was stratified by SV type (deletion or duplication) and by loss-of-function intolerance (pLI) above or below the 90th percentile, a total of 104 tests.

Functional elements that showed significant evidence of SV depletion among intolerant genes (pLI ≥ 90th percentile; Benjamini-Hochberg FDR $q < 0.01$; OR < 1) were selected as our target categories (Fig. 1B and table S5). Nearly identical results were obtained with both random models in the discovery sample and in an independent cohort from the 1000 Genomes Project (table S5 and fig. S6). Categories that showed depletion of SVs relative to simulations included exons (OR = 0.18; $P < 0.0001$), transcription start sites (TSSs) (OR = 0.45; $P < 0.0001$), 3′UTRs (OR = 0.57; $P < 0.0001$), and promoter annotations derived from fetal brain tissue (fetal brain promoters) from the Roadmap Epigenomics Project (OR = 0.73; $P = 0.0011$), and the depletion of CRE-SVs was restricted to intolerant genes (Fig. 1B and table S5). In total, seven categories were significant (four coding and three noncoding). Functional elements were further collapsed into "cis-regulatory" and "coding and noncoding" categories, respectively, and we included one nondepleted category ("intron") as a control, resulting in a total of 10 target categories (table S5).

Focusing on the target functional categories above, family-based association was tested using a group-wise transmission/disequilibrium test (TDT), applying it to private variants (autosomal parent allele frequency = 0.0003), assuming a dominant model of transmission. We confirmed a 50% parental transmission rate for deletions and duplications overall across a range of sizes (table S6). In variant-intolerant genes (pLI ≥ 90th percentile), protein coding deletions were overtransmitted to cases (54/83; transmission rate = 65.1%; $P = 0.002$), but not to controls (26/57, transmission rate = 45.6%; $P = 0.54$) (Fig. 2 and table S6). Paternally inherited CRE-SVs (fetal-brain promoters, TSSs, or 3′UTRs) of intolerant genes were overtransmitted to cases (39/55; transmission rate = 70.9%; $P = 0.0013$), whereas maternal CRE-SVs were not significantly associated with ASD (21/44; transmission rate = 47.7%). The above associations were significant after correction for 20 tests (10 categories of SVs tested for each parent separately) (table S6). Validation of cis-regulatory and exonic SVs was performed where possible using nanopore sequencing, polymerase chain reaction (PCR), or an in silico SNV-based approach (see supplementary materials). In total, 96% (150/156) of SVs were validated with 100% genotype concordance SV[2] (table S7).

The primary hypothesis to be tested in the replication sample (association of paternally inherited CRE-SVs) was preregistered in the form of a preprint describing the analytic details and results of our primary analysis (*16*). We then replicated the association by applying our pipeline to an independent sample of 6105 genomes from 1771 families (*17*). The association of rare (allele frequency ≤ 0.0003) paternally transmitted CRE-SVs was significant in the replication sample (65/109; transmission rate = 59.6%; $P = 0.027$). Also consistent with our primary results, maternally transmitted CRE-SVs were not associated with ASD and inherited coding variants from both parents were associated with ASD (Fig. 2 and table S6).

In the combined data set of 2600 families, the association of paternal CRE-SVs was significant ($P = 3.7 \times 10^{-4}$) after correction for 20 tests. Consistent with a paternal-origin effect, CRE-SVs in cases were inherited more frequently from fathers (104 paternal, 74 maternal; binomial $P = 0.015$). All private cis-regulatory and exonic variants in intolerant genes are given in table S7. The median lengths of cis-regulatory and exonic SVs were 2920 bp [interquartile range (IQR) = 396 to 8282 bp] and 17,261 bp (IQR = 4390 to 112,251 bp), respectively.

The smaller effect size observed in the replication sample (overtransmission of 59.6%, compared with 70.6% in the discovery sample) could be explained by a combination of factors, including chance or true differences in the genetic architecture between samples. Cohorts did not differ dramatically in the numbers of trios and concordant sibling pair (multiplex) families (table S1); thus, family structure is unlikely to have an influence. As mentioned above, selection of families for a subset of the discovery sample (SSC1) was designed to enrich for novel inherited and noncoding risk variants. Thus, ascertainment could in part explain why the SSC1 had the largest effect size of all individual cohorts (fig. S7).

Recurrent CRE-SVs disrupting intolerant genes were observed in cases, including *CNTN4*, *LEO1*, *RAF1*, and *MEST* (table S7) (permutation $P = 0.0036$). Two de novo loss-of-function variants disrupting *LEO1* (*18*, *19*) have been observed in a combined exome data set of ASD and developmental delay from 20 studies, a higher rate of loss-of-function variants than would be expected by chance (expected $n = 0.1$; $P = 0.0025$) (*14*). Both *LEO1* deletions eliminate an upstream regulatory element that has a chromatin signature associated with an active TSS (Fig. 3A) (*20*). A smaller 8.7-kb deletion polymorphism (parent allele frequency = 0.011) was detected within this region, but this variant does not disrupt any annotated functional elements. The deletions were

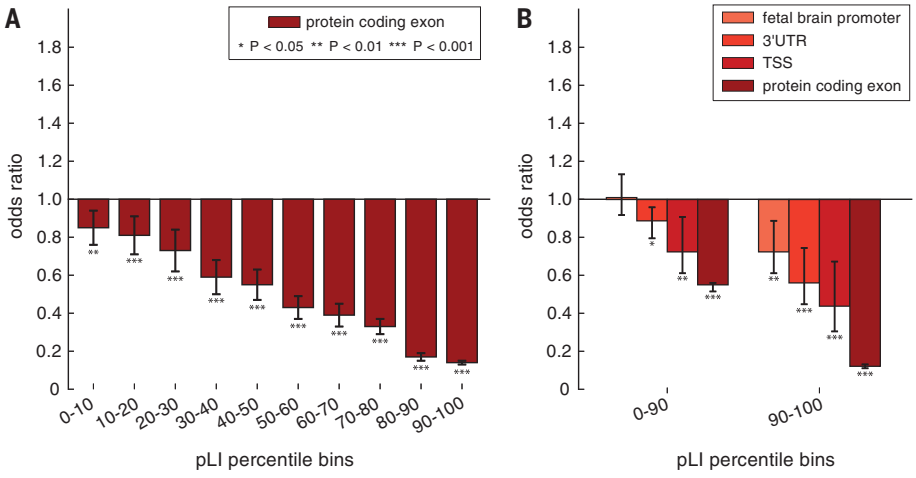


Fig. 1. Selection of target functional categories based on deletion intolerance. Bar charts illustrating functional elements that show depletions in deletions relative to random permutations, stratified by deciles of gene variant intolerance (pLI) as estimated by the ExAC consortium. (**A**) Protein-coding deletions. (**B**) Cis-regulatory elements deletions. Odds ratios calculated based on observed counts versus expected based on permutation. Stars indicate the level of significance in the permutation analysis; whiskers represent 95% confidence intervals. TSS, transcription start site.

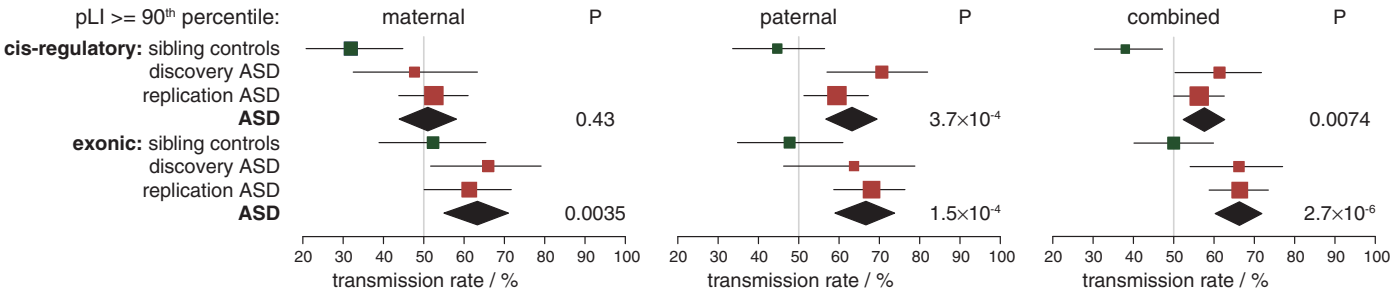


Fig. 2. Parental transmission of private cis-regulatory and exonic SVs to cases and sibling controls. Rate of transmission from parents to offspring was tested for SVs that disrupt cis-regulatory elements or exons of variant-intolerant genes (pLI > 90th percentile). Whiskers represent the 95% confidence intervals. Effect sizes for CRE-SVs in all four cohorts individually is provided in fig. S7).

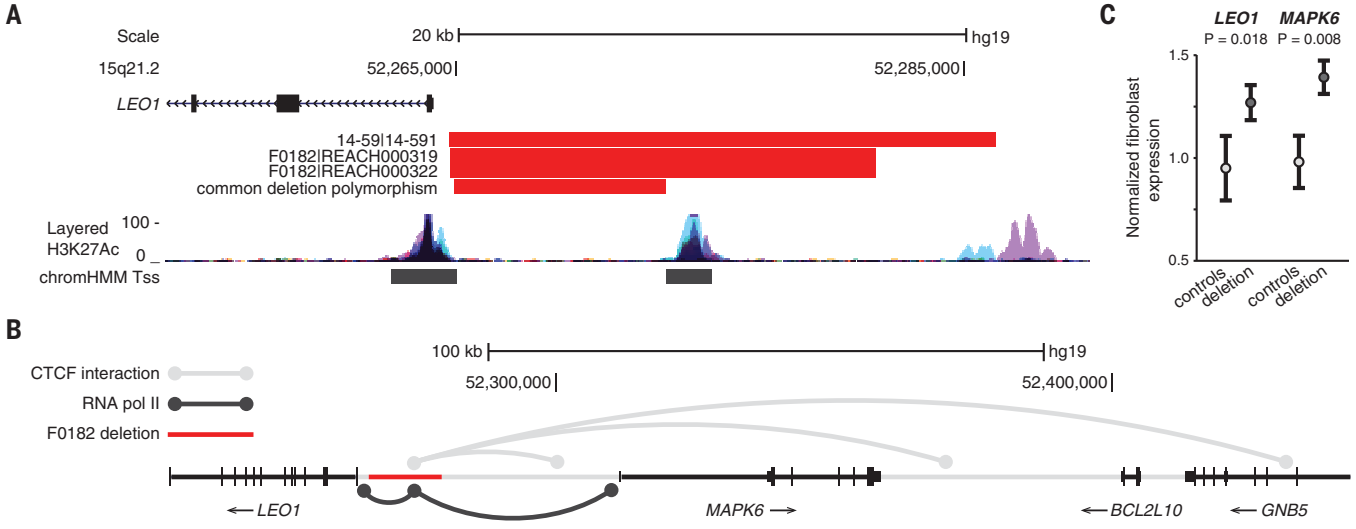


Fig. 3. Recurrent promoter deletions of *LEO1* derepress expression. (**A**) Paternally inherited deletions of the *LEO1* promoter were detected in three affected individuals, one trio (14-59), and one concordant sib pair (F0182). A common deletion polymorphism (parent allele frequency = 0.011) is also present in this locus. (**B**) Chromatin interactions associated with transcription factors RNA polymerase II and CTCF based on ChIA-PET data suggests that the cis-regulatory element upstream of *LEO1* disrupted by both rare deletions (F0182 deletion shown here) serves as a focal point for the spatially organized transcription of *LEO1* and *MAPK6*. (**C**) mRNA expression of *LEO1* and *MAPK6* in fibroblast lines derived from two deletion carriers (REACH00319 and REACH000322), compared with three control lines. Whiskers represent 95% CIs. Layered H3K27Ac, Histone 3 lysine 27 acetylation (an active promoter associated mark) in seven cell types from the Encyclopedia of DNA Elements (ENCODE). ChromHMM Tss is the predicted transcription start site based on chromatin signatures in multiple cell types from the Roadmap Epigenomics Project (*20*).

fine-mapped by nanopore single-molecule sequencing of long PCR products (fig. S8). Published chromatin interactions associated with transcription factors CTCF and RNA polymerase II mapped by chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) (*21*, *22*) revealed this upstream cis-regulatory element to be a focal point for long-range chromatin interactions associated with transcription (Fig. 3B). Expression of *LEO1* and the neighboring *MAPK6* was higher in fibroblast cell lines from two deletion carriers compared with lines from three noncarrier controls (*LEO1 t* test $P = 0.018$; *MAPK6* $P = 0.008$) (Fig. 3C and table S8).

As follow-up to our previous studies of de novo SVs (*9*), we detected de novo mutations in the discovery sample, including 104 deletions, 19 duplications, 2 inversions, 8 complex SVs, and 32 mobile element insertions (MEIs) (fig. S9 and table S9). The majority (68%) of phased de novo SVs originated from the father (binomial test $P = 0.038$) (table S9), comparable to the bias observed for SNVs and indels (*23*). We also confirm that de novo SNVs and indels cluster in proximity to de novo SV break points (permutation $P = 0.0029$) (table S10 and fig. S10) (*9*). ASD cases did not display higher SV mutation rates than sibling controls (fig. S11) (*9*). Considering only the subset of the discovery sample that had not been characterized previously [Relating Genes to Adolescent and Child Health (REACH)], gene disrupting de novo variants were significantly enriched in cases (7.2% in ASD versus 2.1% in controls; permutation $P = 9.2 \times 10^{-5}$; an excess of 5.1% in cases).

Based on this study, we estimate that rare inherited cis-regulatory and coding SVs contribute in 0.77% [95% confidence interval (CI), 0.39 to 1.13] and 1.21% (95% CI, 0.76 to 1.62) of cases, respectively, and inherited known pathogenic SVs not accounted for above (table S11) contribute in another 1.9% of cases. As expected, the contribution of de novo coding SVs is substantial (5.1%); however, no de novo CRE-SVs were detected in cases in the discovery sample (table S9).

Here, we demonstrate that rare SVs that disrupt CREs confer risk for ASD, and this association is concentrated among genes that are highly dosage sensitive. The contribution of CRE-SVs that we observe consists exclusively of inherited variants. This result is consistent with noncoding variants having moderate effects on gene function and disease risk. We find no evidence for a contribution of de novo CRE-SVs, in contrast to anecdotal findings from previous studies (*5*, *7*). We cannot exclude the possibility that de novo CRE-SVs contribute to ASD; however, we can conclude that they are extremely rare.

CRE-SVs exhibited a significant paternal-origin effect. This result was unexpected and

contrasts with a simpler genetic model (*24*) in which inherited genetic risk is transmitted predominantly from mothers due to the reduced vulnerability of females to ASD. Previous studies have shown a maternal bias for inherited truncating variants in genes that were previously implicated from studies of de novo mutation (*25*–*27*). In our study, the contribution of exonic variants to risk was similar for paternal and maternal SVs, suggesting that a maternal origin bias might be restricted to genes that have the most extreme dosage sensitivity. Taken together, our findings indicate that parent-of-origin effects on genetic risk for ASD are more complex than we previously thought, and the allelic spectrum of variants differs between the maternal and paternal genomes.

We propose three possible mechanisms to explain the observed paternal-origin effect of CRE-SVs. The first is a "bilineal two-hit model," in which inherited risk is attributable to a combination of two risk variants: a maternally inherited coding variant of large effect and a paternally inherited CRE variant of moderate effect. This bilineal model predicts that a paternal bias might also be evident for other variants of moderate effect, including hypomorphic missense alleles or loss-of-function variants in genes with a moderate degree of intolerance. While this paper was under review, a genetic study of common variation reported evidence that, for multiplex families, an excess of paternally inherited variants were shared among unrelated children with ASD (*28*), a result that lends support to a bilineal model.

An alternative explanation for a paternal-origin effect is an epigenetic mechanism. For example, deletion of CREs can lead to derepression of imprinted genes (*29*). However, an epigenetic mechanism could only explain our results if noncanonical imprinting of regulatory elements is widespread. Such a phenomenon has not been described, but we cannot rule out this possibility. A third potential mechanism to explain parent-of-origin effects could be a type of "meiotic drive," in which allele-specific selection occurs differently in paternal and maternal germ cells. However, this mechanism is also unlikely given that there are few known examples of gene drive in humans and their effects appear to be quite weak at the population level (*30*).

Due to the greater potential of SVs to affect gene function and regulation relative to SNVs and indels, this class of genetic variation has historically proven effective for illuminating new components of the genetic architecture of disease. Our findings provide a further demonstration of the utility of SV analysis for characterizing the genetic regulatory elements that influence risk for ASD.

REFERENCES AND NOTES

1. J. Sebat *et al*., *Science* **316**, 445–449 (2007).
2. I. Iossifov *et al*., *Nature* **515**, 216–221 (2014).
3. C. M. Rands, S. Meader, C. P. Ponting, G. Lunter, *PLOS Genet.* **10**, e1004525 (2014).
4. M. Kellis *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6131–6138 (2014).
5. R. N. Doan *et al*., *Cell* **167**, 341–354.e12 (2016).
6. T. N. Turner *et al*., *Am. J. Hum. Genet.* **98**, 58–74 (2016).
7. T. N. Turner *et al*., *Cell* **171**, 710–722.e12 (2017).
8. P. H. Sudmant *et al*., *Nature* **526**, 75–81 (2015).
9. W. M. Brandler *et al*., *Am. J. Hum. Genet.* **98**, 667–679 (2016).
10. S. J. Sanders *et al*., *Neuron* **87**, 1215–1233 (2015).
11. D. Antaki, W. M. Brandler, J. Sebat, *Bioinformatics* **2018**, btx813 (2018).
12. M. Lek *et al*., *Nature* **536**, 285–291 (2016).
13. E. B. Robinson *et al*., *Nat. Genet.* **48**, 552–555 (2016).
14. Materials and methods are available as supplementary materials.
15. D. M. Ruderfer *et al*., *Nat. Genet.* **48**, 1107–1111 (2016).
16. W. M. Brandler *et al*., bioRxiv 102327 [Preprint]. 29 March 2017.
17. R. K. C. Yuen *et al*., *Nat. Neurosci.* **20**, 602–611 (2017).
18. S. De Rubeis *et al*., *Nature* **515**, 209–215 (2014).
19. Deciphering Developmental Disorders Study, *Nature* **542**, 433–438 (2017).
20. A. Kundaje *et al*., *Nature* **518**, 317–330 (2015).
21. G. Li *et al*., *Cell* **148**, 84–98 (2012).
22. Z. Tang *et al*., *Cell* **163**, 1611–1627 (2015).
23. J. J. Michaelson *et al*., *Cell* **151**, 1431–1442 (2012).
24. X. Zhao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12831–12836 (2007).
25. I. Iossifov *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5600–E5607 (2015).
26. N. Krumm *et al*., *Nat. Genet.* **47**, 582–588 (2015).
27. B. Wang *et al*., *Sci. Rep.* **6**, 25954 (2016).
28. K. Ye *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7073–7076 (2017).
29. J. L. Thorvaldsen, K. L. Duran, M. S. Bartolomei, *Genes Dev.* **12**, 3693–3702 (1998).
30. A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267–271 (2002).

ACKNOWLEDGMENTS

We thank the families who volunteered for the REACH study. We also thank W. Pfeiffer, M. Tatineni, A. Majumdar, S. Strande, R. Hawkins, the San Diego Supercomputer Center, and Amazon Web Services for hosting the computing infrastructure necessary for completing this project. We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, and E. Wijsman). **Funding:** This study was supported by grants to J.S. from NIH (MH076431 and MH113715) and the Simons Foundation Autism Research Initiative (SFARI) (275724) and by a gift to J.S. from the Beyster Family Foundation. Support to J.S. and K.K.V. was also provided from the ASD Enlight Foundation. Support for the generation of nanopore sequence data was provided by Oxford Nanopore Technologies. Funding for K.P. is from the National Institute of Mental Health (NIMH) (R01MH110558). Funding for E.C. is from NIMH (R01MH110558 and I-P50-MH081755) and a Simons Foundation Grant. L.M.I. was supported by NIH (R21 MH104766, R01 MH105524, and R01 MH109885) and in part by Simons Foundation grant 345469. S.W.S. holds the GlaxoSmithKline–Canadian Institutes of Health Research Chair in Genome Sciences at the University of Toronto and the Hospital for Sick Children. Funding for B.C. is from Ministerio de Economía, Industria y Competitividad (SAF2015-68341-R), Agència de Gestió d'Ajuts Universitaris i de Recerca (2014-SGR-0932), La Marató de TV3 (092020), and the European Commission H2020 Programme MiND (643051). A.H. and M.J.A. received grant support from the Institute Carlos III (FIS PI11/00620) and Mutua Terrassa (FMT grant BE062). Support for the generation of nanopore sequence data was provided by Oxford Nanopore Technologies. Funding for C.T. was provided by La Marató de TV3 (092010). Postdoctoral fellowships were provided to W.B. from the Autism Science Foundation and to M.L.K. from the Canadian Institutes of Health Research. A T32 training grant was provided to D.A. from NIH (GM008666). Funding for collection of fibroblast cell lines was provided by a grant (IT1-06611) to J.G.G. from the California Institute for Regenerative Medicine. A.R.M. is supported by a grant from the National Alliance for Research on Schizophrenia and Depression and NIH grants R01MH108528 and R01MH109885. S.F.K. is supported by NIH grant HD0077693. A.H. is supported by a grant from the Spanish Ministry of Health (FIS PI/15/01295 IP). We thank Genome Canada and the Centre for Applied Genomics (TCAG) for contributing the replication data set from the Autism Speaks MSSNG cohort (www.mss.ng). **Author contributions:** Conceptualization, J.S., W.M.B.; Methodology, W.M.B., D.A., M.G., J.S.; Software, D.A., W.M.B., M.G.; Validation, M.S.M., T.R.C., S.T., M.L.K., Y.Y., E.H.; Formal Analysis, W.M.B., D.A., M.G., M.L.K., J.W., P.T., K.S.M.; Writing–Original Draft, W.M.B., J.S.; Writing–Review and Editing, L.M.I., D.J.T., A.R.M., C.M.N., K.S.M.; Resources, K.K.V., T.P., S.C.T., D.B., B.K., A.T., J.C.V., C.C., A.R.M., R.C., B.C., L.M.I., S.G., A.H., M.J.A., I.R., S.J., D.J.T., S.F.K., J.G.G., E.C., K.P., S.W.S., B.T., G.K.; Visualization, W.M.B., D.A.; Supervision, J.S.; Project Administration, O.H.; Funding Acquisition, J.S., W.M.B., D.A., M.K. **Competing interests:** J.S. declares that a patent has been issued to the Cold Spring Harbor Laboratory by the U.S. Patent and Trademark Office on genetic methods for the diagnosis of autism (patent number 8554488). W.M.B., B.K., A.T., and J.C.V. are employed by Human Longevity, Inc. Y.Y., E.H., S.J., and D.J.T. work for Oxford Nanopore Technologies, Inc. **Data and materials availability:** The data reported in this paper are archived at the National Database for Autism Research (DOI:10.15154/1340302), including the structural variant callset and raw sequence (FASTQ), alignment (BAM), and variant call (VCF) files from the REACH cohort. We appreciate obtaining access to Simons Simplex Collection genomic and phenotypic data on SFARI Base. Approved researchers can obtain the SSC population data set described in this study (https://sfari.org/resources/autism-cohorts/simons-simplex-collection) by applying at https://base.sfari.org.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/327/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S11
References (*31*–*34*)

14 March 2017; resubmitted 7 August 2017
Accepted 27 February 2018
10.1126/science.aan2261

CANCER GENOMICS

Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq

Mariella G. Filbin,[1,2,3,4,5]* Itay Tirosh,[3,4,6]* Volker Hovestadt,[1,3,4]* McKenzie L. Shaw,[1,3,4] Leah E. Escalante,[1,3,4] Nathan D. Mathewson,[7] Cyril Neftel,[1,3,4,8] Nelli Frank,[9] Kristine Pelton,[10] Christine M. Hebert,[1,3,4] Christine Haberler,[11] Keren Yizhak,[4] Johannes Gojo,[5] Kristof Egervari,[1] Christopher Mount,[12] Peter van Galen,[1,3,4] Dennis M. Bonal,[13] Quang-De Nguyen,[13] Alexander Beck,[1] Claire Sinai,[2,10] Thomas Czech,[14] Christian Dorfer,[14] Liliana Goumnerova,[2] Cinzia Lavarino,[15] Angel M. Carcaboso,[15] Jaume Mora,[15] Ravindra Mylvaganam,[1] Christina C. Luo,[1] Andreas Peyrl,[5] Mara Popović,[16] Amedeo Azizi,[5] Tracy T. Batchelor,[17] Matthew P. Frosch,[1] Maria Martinez-Lage,[1] Mark W. Kieran,[2] Pratiti Bandopadhayay,[2,4] Rameen Beroukhim,[4,18] Gerhard Fritsch,[9] Gad Getz,[1,4] Orit Rozenblatt-Rosen,[3,4] Kai W. Wucherpfennig,[7] David N. Louis,[1] Michelle Monje,[12] Irene Slavc,[5] Keith L. Ligon,[2,4,10] Todd R. Golub,[2,4] Aviv Regev,[3,4,19]†‡ Bradley E. Bernstein,[1,3,4]†‡ Mario L. Suvà[1,3,4]†‡

Gliomas with histone H3 lysine27-to-methionine mutations (H3K27M-glioma) arise primarily in the midline of the central nervous system of young children, suggesting a cooperation between genetics and cellular context in tumorigenesis. Although the genetics of H3K27M-glioma are well characterized, their cellular architecture remains uncharted. We performed single-cell RNA sequencing in 3321 cells from six primary H3K27M-glioma and matched models. We found that H3K27M-glioma primarily contain cells that resemble oligodendrocyte precursor cells (OPC-like), whereas more differentiated malignant cells are a minority. OPC-like cells exhibit greater proliferation and tumor-propagating potential than their more differentiated counterparts and are at least in part sustained by *PDGFRA* signaling. Our study characterizes oncogenic and developmental programs in H3K27M-glioma at single-cell resolution and across genetic subclones, suggesting potential therapeutic targets in this disease.

Diffuse midline gliomas with histone H3 lysine27-to-methionine mutations (H3K27M-glioma) are uniformly fatal malignancies (*1*). They are both spatially and temporally restricted, occurring in midline structures of the brain with peak incidence at 6 to 9 years of age. These patterns suggest that a particular cell type, potentially undergoing rapid expansion at this stage, is susceptible to transformation by H3K27M. Experimental models suggest that neural precursor cells (NPCs) can be transformed in vitro by H3K27M, in combination with *TP53* mutation and *PDGFRA* overexpression (*2*). H3K27M suppresses EZH2, the catalytic subunit of Polycomb Repressive Complex 2 (PRC2), compromising epigenetic repression and potentially affecting cellular differentiation (*1*–*3*). In patient samples, little is known about the developmental cell states present or how they cooperate with H3K27M for tumorigenesis. Single-cell RNA sequencing (scRNA-seq) can help address such questions by characterizing cancer cell states, their proliferative signatures, and their similarity to normal or other malignant cell types. scRNA-seq also helps relate single cell states to genetics through inferred chromosomal copy number variations (CNVs) or mutation detection in expressed transcripts (*4*–*6*).

We obtained fresh tissue from diagnostic biopsies of six H3K27M-glioma (table S1 and fig. S1). Each sample was dissociated into single cells, flow sorted, and profiled by means of full-length scRNA-seq (fig. S2) (*7*, *8*). We retained 2458 cells that passed quality controls (*8*) for downstream analyses (table S2). The cells' profiles group primarily by tumor of origin (Fig. 1A and fig. S3), but two clusters contain cells from multiple tumors, expressing markers of either microglia (such as *CD14*, *CX3CR1*, and *AIF1*) or oligodendrocytes (such as *MBP* and *PLP1*) (Fig. 1, A and B), suggesting that they are nonmalignant cells. Accordingly, in presumed malignant cells only, we detected evidence for cancer-specific aberrations—point mutations and/or CNVs (*4*–*6*)—in 93.7% of cells in five of the six tumors (Fig. 1, C and D, and figs. S4 to S7). Thirty-four percent (833 of 2458) of the cells had H3K27M mutations in scRNA-seq reads mapping to *H3F3A* or *HIST1H3B* (Fig. 1D and fig. S7) only detected in presumed malignant cells; other point mutations from whole-genome sequencing (WGS)/whole-exome sequencing (WES) (tables S1 and S2) were also detected only in presumed malignant cells (Fig. 1D and fig. S7) (*8*). Analyzing a seventh tumor (MUV17) with *H3F3A*-specific primers in the scRNA-seq protocol confirmed H3K27M in 97% of the presumed malignant cells (fig. S8) (*8*).

Next, we compared malignant cell transcriptomes across different glioma types, including H3K27M-glioma, isocitrate dehydrogenase (IDH)–mutant oligodendroglioma (IDH-O), IDH-mutant astrocytoma (IDH-A), and IDH–wild-type glioblastoma (GBM) (fig. S9A and tables S3 and S4) (*5*, *6*, *9*). Many genes were up-regulated in H3K27M-glioma versus other tumors (n = 182 genes), but few were down-regulated (n = 12 genes), which is consistent with H3K27M blocking repression by PRC2 (fig. S9, B to D). Accordingly, H3K27M–up-regulated genes are enriched for PRC2 target genes (P < 0.0001, hypergeometric test) (fig. S9E) (*10*, *11*). The PRC1 subunit *BMI1* was up-regulated in H3K27M-glioma, possibly representing a compensatory mechanism for PRC2 suppression. Suppression of *BMI1* by means of CRISPR knockout or its pharmacological inhibition reduced viability of H3K27M glioma cells, relative to treatment controls and non-H3K27M glioma lines (fig. S10) (*12*).

We next assessed patterns of intratumor heterogeneity and distinguished subpopulations of malignant cells within H3K27M-glioma. We identified four programs (P1 to P4) that were consistently observed as variable within tumors and across computational methods (*8*): cell cycle (such as *PCNA* and *CDK1*), astrocytic

[1]Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. [2]Department of Pediatric Oncology, Dana-Farber Boston Children's Cancer and Blood Disorders Center, Boston, MA 02215, USA. [3]Klarman Cell Observatory, Broad Institute of Harvard and Massachussetts Institute of Technology (MIT), Cambridge, MA 02142, USA. [4]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. [5]Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria. [6]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 7610001, Israel. [7]Department of Cancer Immunology and Virology, Department of Microbiology and Immunobiology, Department of Neurology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA. [8]Institute of Pathology, Faculty of Biology and Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. [9]Children's Cancer Research Institute (CCRI), St. Anna Kinderspital, Medical University of Vienna, Vienna, Austria. [10]Department of Oncologic Pathology, Brigham and Women's Hospital, Boston Children's Hospital, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [11]Institute of Neurology, Medical University of Vienna, Vienna, Austria. [12]Departments of Neurology, Neurosurgery, Pediatrics, and Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA. [13]Center for Biomedical Imaging in Oncology, Lurie Family Imaging Center, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [14]Department of Neurosurgery, Medical University of Vienna, Vienna, Austria. [15]Developmental Tumor Biology Laboratory, Hospital Sant Joan de Déu, Esplugues de Llobregat, 08950 Barcelona, Spain. [16]Institute of Pathology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia. [17]Departments of Neurology and Radiation Oncology, Division of Hematology/Oncology, Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, USA. [18]Departments of Cancer Biology and Medical Oncology, Dana-Farber Cancer Institute, and Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215, USA. [19]Department of Biology, Koch Institute for Integrative Cancer Research, Howard Hughes Medical Institute, MIT, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: suva.mario@mgh.harvard.edu (M.L.S.); bernstein.bradley@mgh.harvard.edu (B.E.B.); aregev@broadinstitute.org (A.R.) ‡These authors contributed equally to this work.

differentiation (AC-like) (such as *GFAP* and *APOE*), oligodendrocytic differentiation (OC-like) (such as *MBP* and *PLP1*), and OPC-like programs (such as *PDGFRA* and *CSPG4*) (Fig. 2, A and B; figs. S11 and S12; and table S5) (*13*, *14*).

Scoring each cell for these expression programs highlighted a putative developmental hierarchy, in which proliferation is largely restricted to OPC-like cells that presumably both self-renew and give rise to the OC-like and AC-like cells (Fig. 2C). The relative fraction of cells in each compartment (OPC-like, AC-like, and OC-like) varied between tumors, but OPC-like cells were consistently the most prevalent (fig. S13). We validated this cellular composition with RNA in situ hybridization (ISH) of seven tumor specimens (table S4), using markers for OPC-like, AC-like, and cycling cells (*PDGFRA*, *APOE*, and *Ki-67*, respectively) (Fig. 2D and fig. S13). Analysis of normal brain cell types indicated that PRC2 targets are highly expressed by OPCs and lowly expressed by OCs ($P < 0.05$, Student's t test) (fig. S14 and table S6). Thus, OPC differentiation into OC may require the repressive activity of PRC2 and may be hindered by PRC2 inactivation through H3K27M, which may result in accumulation of self-renewing OPC-like cells. Accordingly, some tumors had little evidence of lineage differentiation (fig. S13). Furthermore, knockout of the OPC lineage factor *PDGFRA* by use of CRISPR/Cas9 reduced viability in two models of H3K27M-glioma (fig. S15). Combined targeting of *PDGFRA* and the PRC1 subunit *BMI1* further reduced viability of these models (fig. S15). Thus, OPC-like cells are the predominant subpopulation of H3K27M tumors, propagate the disease in patients, and may be susceptible to therapeutic strategies that concurrently target lineage-defined (*PDGFRA*) and somatically altered (*BMI1*) cellular programs.

The H3K27M-glioma hierarchy resembles that in IDH-mutant gliomas; both contain a cycling stem-like subpopulation and differentiated subpopulations of OC-like and AC-like cells. However, they differ in the proportions of subpopulations and in the exact set of genes associated with each program (*5*, *6*). We scored malignant cells from both cohorts by the signatures for H3K27M and IDH-mutant gliomas (*5*, *6*). Astrocytic programs were highly similar (Fig. 3, A and B), whereas oligodendrocytic and stem-like programs are largely distinct, with only a small set of shared genes (Fig. 3, C to F). Indeed, H3K27M-specific

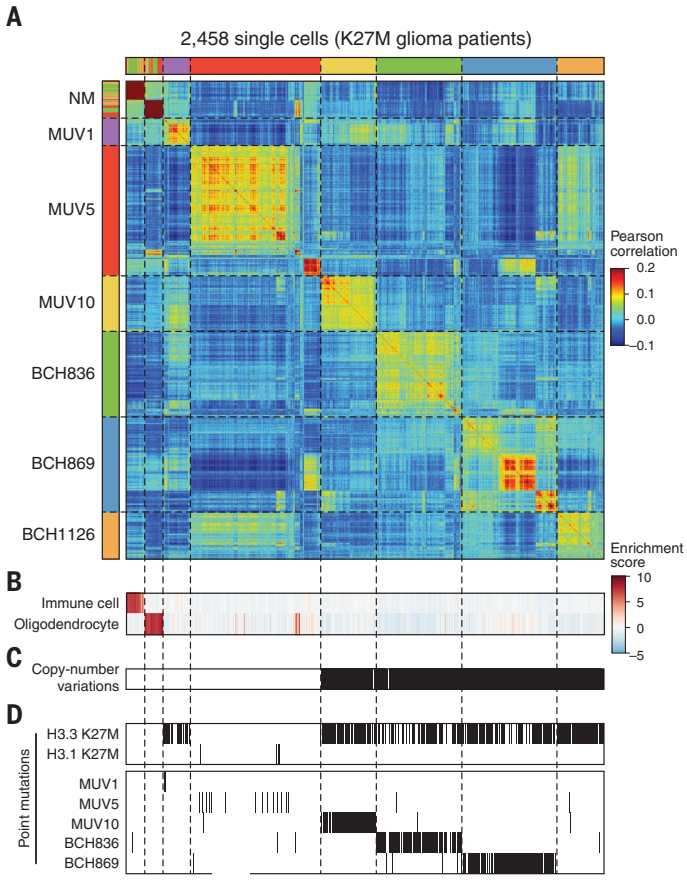


Fig. 1. Characterization of H3K27M-glioma by means of scRNA-seq. (**A**) Pairwise correlations between the expression profiles of 2458 single cells (rows, column) from six H3K27M-glioma samples (color bar). Two clusters of nonmalignant cells are marked as "NM." (**B**) Enrichment score of microglia and oligodendrocyte signatures. (**C**) Inferred CNV profiles. Black indicates CNV present (fig. S4). (**D**) Gene mutations. (Top) H3K27M.; (Bottom) All other mutations identified per sample by means of WGS/WES. Black line indicates at least one mutated gene identified (fig. S6).

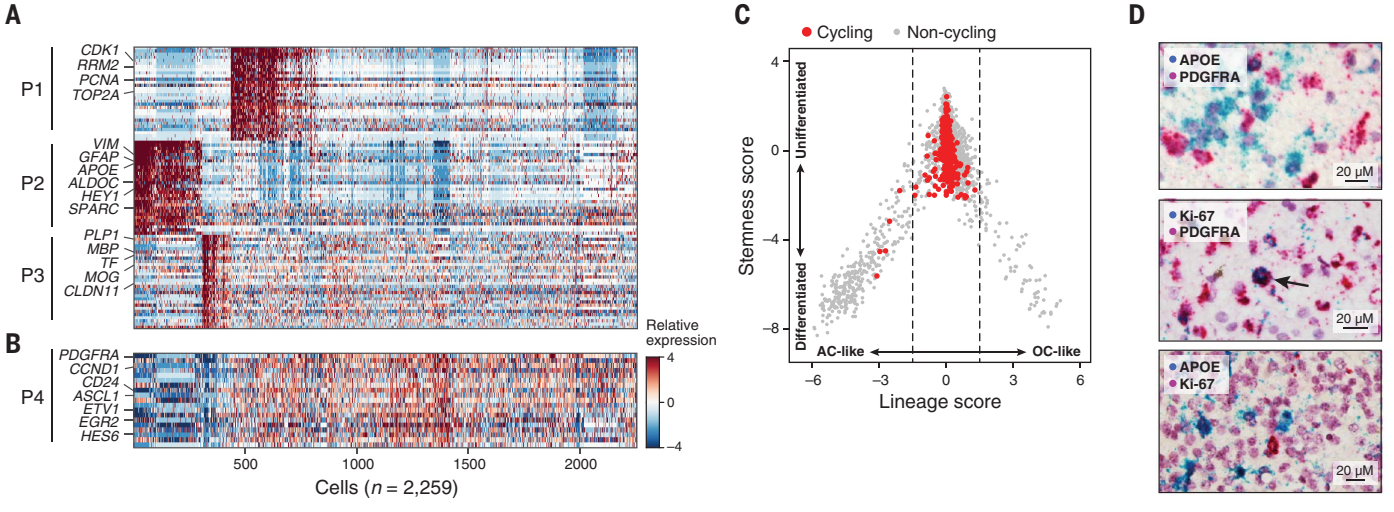


Fig. 2. Intratumor heterogeneity in H3K27M-glioma. (**A** and **B**) Relative expression (color bar) across 2259 malignant cells (columns) of the top 30 genes for each of the combined expression programs P1, P2, and P3 [(A), rows] or 19 genes [(B), P4] that are preferentially expressed in cells with low expression of P1, P2, and P3 (*8*). (**C**) Plot of the lineage (x axis) and stemness (y axis) scores for each of 2259 malignant cells (dots). Red dots indicate positive score for the cell cycle program. (**D**) In situ RNA hybridization of H3K27M glioma for astrocytic-like (*APOE*), OPC-like (*PDGFRA*), and proliferation (*Ki-67*) markers. Arrow highlights cell coexpressing *PDGFRA* and *Ki-67*.

stem cell genes were expressed highly by OPCs (such as *PDGFRA*), whereas IDH-specific stem cell genes were expressed by NPCs (such as *SOX11*) (Fig. 3, F and G) (*5*, *6*). Additionally, H3K27M-glioma harbored more cycling and undifferentiated cells than IDH-mutant gliomas (Fig. 3H), potentially accounting for their more aggressive behavior.

We next considered how this cellular architecture relates to genetic heterogeneity and tumor evolution. CNV analysis and inference of haplotypes uncovered distinct genetic subclones in two tumors (Fig. 3, I and J, and figs. S16 and S17) (*8*). Although all cells in BCH869 have lost one copy of chromosome 14, the subclones had

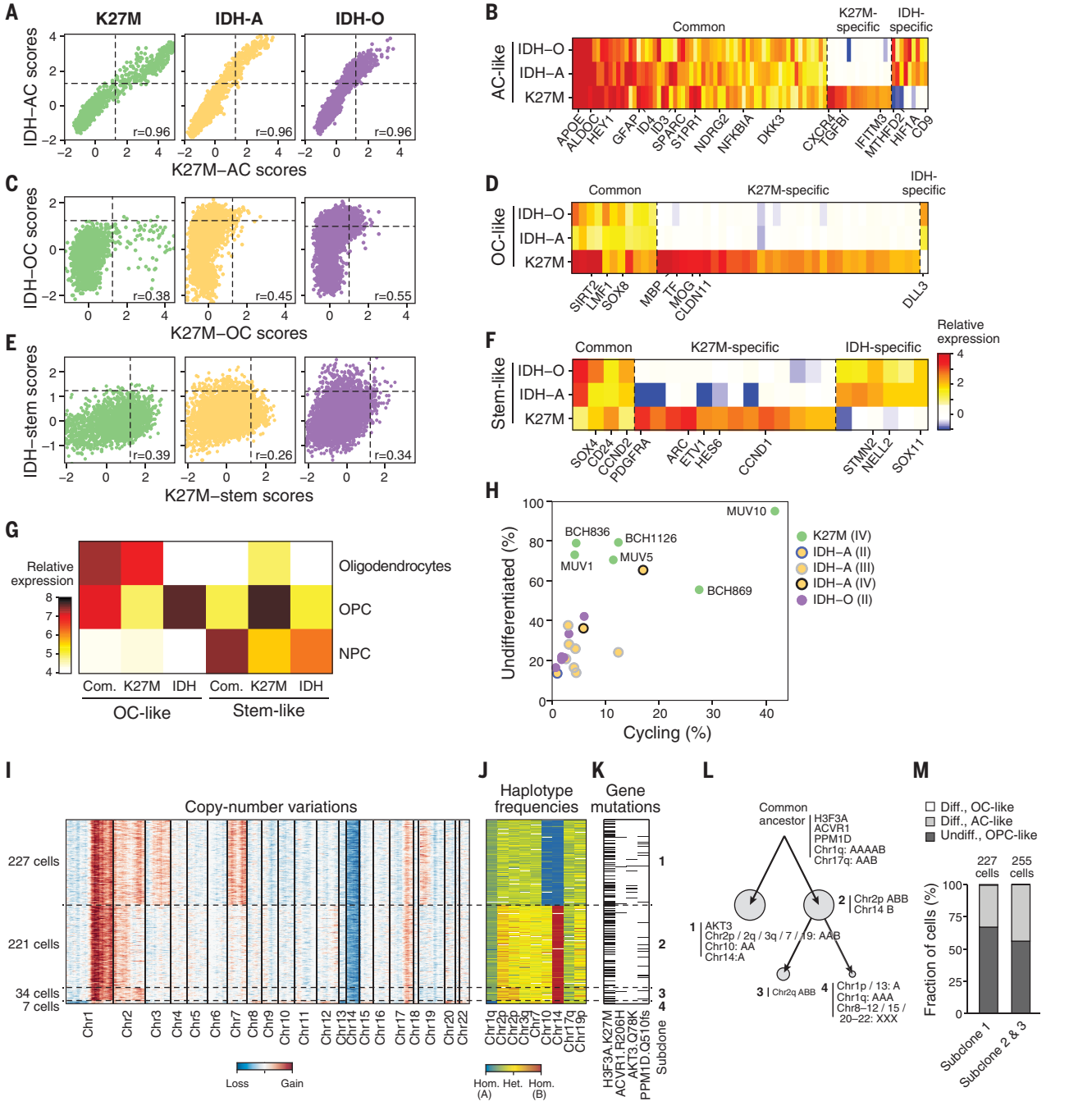


Fig. 3. Cellular hierarchies of H3K27M-glioma and IDH-mutant gliomas. (**A**, **C**, and **E**) Malignant cells (dots) from H3K27M, IDH-A, and IDH-O scored for the (A) AC-like, (C) OC-like, and (E) stem-like signatures of H3K27M-glioma (*x* axis) and of IDH-mutant gliomas (*y* axis). Correlation values are in the bottom right quadrant. (**B**, **D**, and **F**) Relative expression in (B) AC-like, (D) OC-like, and (F) stem-like cells in each glioma class (rows) of genes with preferential expression in the respective cell subset (*8*), with genes ordered into those common to H3K27M and IDH-mutant gliomas, or specific to either tumor type. (**G**) Relative expression (color bar) of OC-like and stem-like genes shared between (common) or specific to H3K27M and IDH-mutant gliomas in nonmalignant oligodendrocytes, OPCs, and NPCs (*5*). (**H**) Percentage of cycling cells (*x* axis) and undifferentiated cells (*y* axis) in each glioma sample, marked by type and grade. (**I**) CNVs, (**J**) haplotype frequencies, and (**K**) point mutations in selected genes identified with WGS (columns) inferred for individual malignant cells (rows) from BCH869. Dashed lines indicate four subclones based on CNV and haplotype profiles. (**L**) Inferred phylogenetic tree (*8*) of individual subclones detected for BCH869. Circle sizes indicate relative number of cells in subclone. Genetic events are indicated at the inferred point of their first detection. (**M**) Relative number of malignant cells classified into OPC-, AC-, or OC-like states for BCH869 subclone 1 or the combination of subclones 2 and 3.

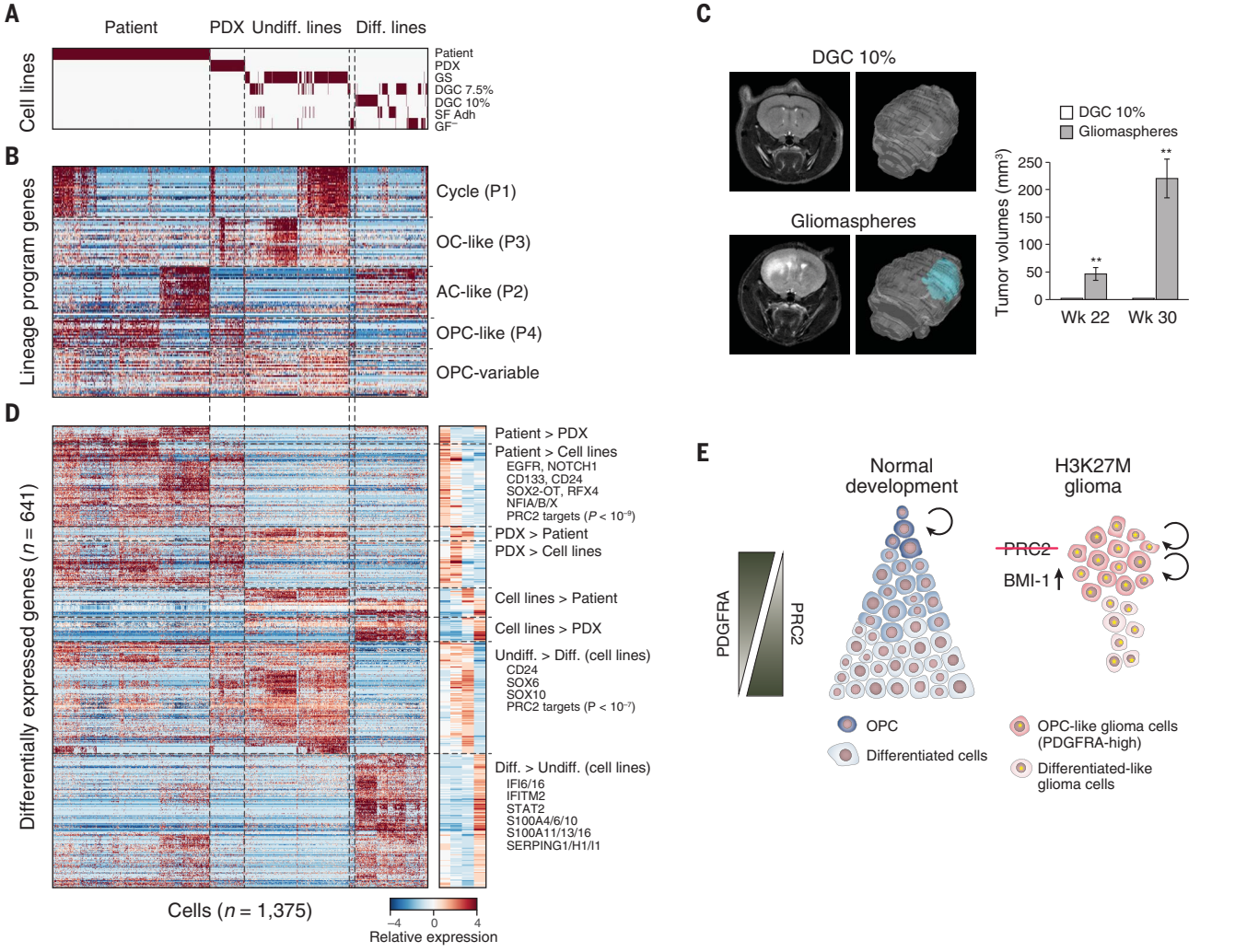


Fig. 4. Single-cell comparisons of matched H3K27M-glioma patient sample, PDX, and culture models for tumor BCH869. (**A**) Cells ordered by sample type and within each sample by means of hierarchical clustering (fig. S19). (**B**) Heatmap shows expression of the top 30 genes of the cell cycle and lineage programs (P1 to P4) described in Fig. 2 and in (*8*), for cells ordered as in (A). (**C**) (Left) Mouse brain magnetic resonance images (MRIs) with three-dimensional reconstruction at 22 weeks after injection of 200,000 BCH869 cells. (Right) MRI tumor volume (*8*). $^{**}P < 0.01$ by paired, two-tailed Student's *t* test. Error bars indicate SEM. (**D**) Heatmap shows expression of differentially expressed genes between sample types, for each pairwise comparison. Cells are ordered as in (A). (Right) Average expression in each sample type. (**E**) Model of H3K27M-glioma cellular architecture (right) compared with normal development (left).

different haplotypes, indicating that two distinct events led to loss of alternate chromosome 14 alleles (Fig. 3J). Additionally, certain somatic gene mutations could be assigned to just one of the subclones (Fig. 3K). Use of both CNVs and haplotype frequencies enabled inference of phylogenetic trees (Fig. 3L, fig. S17D, and table S7) (*8*). Each genetic subclone contained cells spanning a similar diversity of cellular states, although with some variation in their relative proportions (Fig. 3M and fig. S17E). This suggests that distinct genetic subclones share similar developmental hierarchies in H3K27M-glioma.

Next, we profiled 863 single cells from several models derived from BCH869—including patient-derived xenograft (PDX), gliomaspheres (GS), and differentiated glioma cells (DGC) (fig. S18) (*15*)—and compared them with cells from the corresponding tumor. Cells from the PDX most closely approximated malignant cell states in the primary tumor, whereas each in vitro model recapitulated some, but not all, of these states (Fig. 4, A and B, and fig. S19A). Tumor-initiation capacity was exclusive to cells grown in GS conditions (n = 8 of 8 mice) that partially recapitulated the OPC-like state, and was abolished in DGCs that mirrored the AC-like state (n = 0 of 8 mice) (Fig. 4C and fig. S19, B and C), supporting the functional relevance of the inferred hierarchies. Last, we identified differentially expressed genes between the primary tumor, PDX, and culture models (Fig. 4D and table S8). A large number of genes were down-regulated in culture models compared with the primary tumor, including glioma-related oncogenes (such as epidermal growth factor receptor), putative stemness genes (such as SOX2, RFX4, and CD133), and PRC2-targets ($P < 10^{-9}$) (*10*). PRC2 targets were further down-regulated in DGC (compared with GS), along with neurodevelopmental regulators such as SOX6 and SOX10 (Fig. 4D). This highlights the specificities and limitations of each model.

scRNA-seq of primary H3K27M-glioma defines a putative developmental hierarchy and contrasts the underlying stem cell and differentiation programs with other classes of glioma. The findings are consistent with an emerging cancer stem cell model for gliomas in which (i) genetically defined glioma classes, such as IDH-mutant gliomas and H3K27M-glioma, contain different types of stem-like cells; (ii) the fraction of stemlike cells can vary substantially between glioma types—this extends the traditional cancer stem cell model, which posits that stem cells represent a minority of

malignant cells; (iii) differentiation hierarchies play a critical role in the functional properties of glioma cells because self-renewal and tumor-propagating potential are exclusive to the most primitive cells; and (iv) genetic subclones in tumors tend to share similar cellular architecture. Our study also highlights opportunities for therapeutic intervention. We show that OPC-like cells drive H3K27M-glioma, suggesting that the OPC marker PDGFRA could be a lineage-defined therapeutic target, relevant even in the absence of genetic amplification or mutation. H3K27M-glioma also overexpress the PRC1 subunit BMI1 and are sensitive to its inhibition, either alone or in combination with PDGFRA inhibition, hinting at a potential compensatory mechanism for PRC2 dysfunction (Fig. 4E). Thus, lineage-defined and somatically altered cellular programs in H3K27M-glioma suggest complementary opportunities for therapeutic intervention in these incurable malignancies.

REFERENCES AND NOTES

1. D. Sturm *et al*., *Nat. Rev. Cancer* **14**, 92–107 (2014).
2. K. Funato, T. Major, P. W. Lewis, C. D. Allis, V. Tabar, *Science* **346**, 1529–1533 (2014).
3. P. W. Lewis *et al*., *Science* **340**, 857–861 (2013).
4. A. P. Patel *et al*., *Science* **344**, 1396–1401 (2014).
5. I. Tirosh *et al*., *Nature* **539**, 309–313 (2016).
6. A. S. Venteicher *et al*., *Science* **355**, eaai8478 (2017).
7. S. Picelli *et al*., *Nat Protoc.* **9**, 171–181 (2014).
8. Material and methods are available as supplementary materials.
9. B. B. Liau *et al*., *Cell Stem Cell* **20**, 233–246.e7 (2017).
10. I. Ben-Porath *et al*., *Nat. Genet.* **40**, 499–507 (2008).
11. S. Darmanis *et al*., *Cell Reports* **21**, 1399–1410 (2017).
12. A. Kreso *et al*., *Nat. Med.* **20**, 29–36 (2014).
13. G. La Manno *et al*., *Cell* **167**, 566–580.e19 (2016).
14. S. Marques *et al*., *Science* **352**, 1326–1329 (2016).
15. M. L. Suvà *et al*., *Cell* **157**, 580–594 (2014).

ACKNOWLEDGMENTS

We thank L. Gaffney for assistance with figures. **Funding:** This work was supported by grants from the Alex Lemonade Stand (M.M. and M.L.Su.), the Wang Family Fund (M.L.Su., A.R., B.E.B., and M.M.), the Smith Family Foundation (M.L.Su.), the Broad Institute Broad*next10* (M.L.Su. and O.R.-R.), the V Foundation for Cancer Research (M.L.Su.), the Howard Goodman Fellowship at Massachussetts General Hospital (MGH) (M.L.Su.), the Merkin Institute Fellowship at the Broad Institute of MIT and Harvard (M.L.Su.), the Rachel Molly Markoff Foundation (M.L.Su. and B.E.B.), NIH–National Cancer Institute (NCI) brain cancer SPORE P50CA165962 (M.G.F., T.T.B., and M.L.Su.), start-up funds from the MGH department of Pathology, The Cure Starts Now Foundation, Hope for Caroline Foundation, Julian Boivin Courage for Cures Foundation, Abbie's Army, Michael Mosier Defeat DIPG Foundation, Reflections of Grace Foundation, The Cure Starts Now Australia, Brooke Healey Foundation, Soar With Grace Foundation, Jeffrey Thomas Hayden Foundation, Cure Brain Cancer Foundation, The Jones Family Foundation, Musella Foundation, Pray, Hope Believe Foundation, Smiles for Sophie Foundation, Benny's World, Love Chloe Foundation, Aiden's Avengers, A Cure from Caleb Society, The Operation Grace White Foundation, Ryan's Hope, Wayland Villars DIPG Foundation, American Childhood Cancer Organization, Juliana Rose Donnelly Trust, Sheila Jones & Friends, The Ellie Kavalieros DIPG Research Fund, Voices Against Brain Cancer, The DIPG Collaborative, Zach Carson Foundation, the Micky Czech Foundation, The Guglietti Family Trust, Prayers From Maria Foundation, Ryan Harvey Foundation, Markoff Art in Giving, and the Brock Fleming Foundation. M.G.F. holds a Career Award for Medical Scientist from Burroughs Wellcome Fund and K12 Paul Calabresi Career Award for Clinical Oncology–Training Program in Nervous System Tumors (K12CA090354). I.T. was supported by the Zuckerman STEM Leadership Program and the Benoziyo Endowment Fund for the Advancement of Science. V.H. was supported by a European Molecular Biology Organization long-term fellowship and a Human Frontier Science Program fellowship. C.N. is supported by the Placide Nicod Foundation. K.L.L., L.G., and M.W.K. received support from NCI grant P01 CA142536. K.E. is supported by the Fondation Nuovo-Soldati. P.B. and R.B were supported by the St. Baldrick's Foundation, the Pediatric Brain Tumor Foundation, Gray Matters Brain Cancer Foundation Olivia Caldwell Foundation, and NCI (grants R01CA219943 and R01CA188228). A.R. was supported by funds from the Howard Hughes Medical Institute, the Klarman Cell Observatory, STARR cancer consortium, NCI grant 1U24CA180922, NCI grant R33CA202820, by the Koch Institute Support (core) grant P30-CA14051 from NCI, the Ludwig Center, and the Broad Institute. B.E.B. was supported by the NIH Common Fund and National Cancer Institute (grant DP1CA216873), the American Cancer Society, the Ludwig Center at Harvard Medical School, and the Bernard and Mildred Kayden MGH Research Institute Chair. Flow cytometry and sorting services were supported by shared instrumentation grant 1S10RR023440-01A1. M.M. was supported by the California Institute of Regenerative Medicine (CIRM) grants RB4-06093 and RN3-06510, Lyla Nsouli Foundation, and the Virginia and D. K. Ludwig Fund for Cancer Research. C.L., A.M.C., and J.M. are supported by The Alicia Pueyo Foundation. **Author contributions:** M.G.F., I.T., V.H., B.E.B., A.R., and M.L.Su. conceived the project and designed the study. M.G.F. collected and processed tumor samples, performed all cell culture assays, and planned mouse studies. I.T. and V.H. performed all computational analyses. M.L.Sh., L.E.E., N.D.M., C.N., N.F., K.P., C.M.H., C.H., J.G., K.E., A.B., C.S., T.C., C.D., L.G., C.L., A.M.C., J.M., A.P., M.P., A.A., T.T.B., M.W.K., G.F., and I.S. helped with collecting tumors and generating single-cell sequencing data. N.D.M. helped with CRISPR experiments. R.M. and C.C.L. provided flow cytometry expertise. M.G.F., V.H., and P.V.G. designed and performed targeted sequencing experiments. K.Y., G.G., P.B., and R.B. provided support for genomic and genetic analyses. M.L.Sh. performed in situ hybridization experiments. M.P.F. and M.M.-L. provided samples for in situ hybridization experiments. C.M. and M.M. developed normal human cell cultures used in the study. D.M.B. and Q.-D.N. performed in vivo mouse studies. K.W.W., O.R.R., D.N.L., I.S., K.L.L., and T.R.G. provided experimental and analytical support. M.G.F., I.T, V.H., A.R., B.E.B., and M.L.Su. interpreted results and wrote and edited the manuscript with feedback from all authors. **Competing interests:** B.E.B. is an advisor and equity holder for Fulcrum Therapeutics, 1CellBio, and HiFiBio. A.R. is a scientific advisory board member for ThermoFisher Scientific, Syros Pharmaceuticals, and Driver Group. M.G.F., I.T., V.H., A.R., B.B., and M.L.Su. are co-inventors on a U.S. provisional patent application (U.S. 62/585,468) relating to advances described in this manuscript filed by the Broad Institute, MIT, and MGH. G.G. has patent applications on tools for cancer genome analysis. All other authors have no competing interests. **Data and materials availability:** Data generated for this study are available through the Gene Expression Omnibus (GEO, accession no. GSE102130) or the Broad Institute Single-Cell Portal (https://portals.broadinstitute.org/single_cell/study/single-cell-analysis-in-pediatric-midline-gliomas-with-histone-h3k27m-mutation). All other data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/331/suppl/DC1
Material and Methods
Figs. S1 to S19
Tables S1 to S8
References (*16*–*27*)

7 August 2017; accepted 26 February 2018
10.1126/science.aao4750

DRUG DEVELOPMENT

MFN2 agonists reverse mitochondrial defects in preclinical models of Charcot-Marie-Tooth disease type 2A

Agostinho G. Rocha,[1]* Antonietta Franco,[1]* Andrzej M. Krezel,[2] Jeanne M. Rumsey,[1] Justin M. Alberti,[1] William C. Knight,[1] Nikolaos Biris,[3] Emmanouil Zacharioudakis,[3] James W. Janetka,[2] Robert H. Baloh,[4] Richard N. Kitsis,[5] Daria Mochly-Rosen,[6] R. Reid Townsend,[1] Evripidis Gavathiotis,[3] Gerald W. Dorn II[1]†

Mitofusins (MFNs) promote fusion-mediated mitochondrial content exchange and subcellular trafficking. Mutations in *Mfn2* cause neurodegenerative Charcot-Marie-Tooth disease type 2A (CMT2A). We showed that MFN2 activity can be determined by Met^{376} and His^{380} interactions with Asp^{725} and Leu^{727} and controlled by PINK1 kinase–mediated phosphorylation of adjacent MFN2 Ser^{378}. Small-molecule mimics of the peptide-peptide interface of MFN2 disrupted this interaction, allosterically activating MFN2 and promoting mitochondrial fusion. These first-in-class mitofusin agonists overcame dominant mitochondrial defects provoked in cultured neurons by CMT2A mutants MFN2 Arg^{94}→Gln^{94} and MFN2 Thr^{105}→Met^{105}, as demonstrated by amelioration of mitochondrial dysmotility, fragmentation, depolarization, and clumping. A mitofusin agonist normalized axonal mitochondrial trafficking within sciatic nerves of MFN2 Thr^{105}→Met^{105} mice, promising a therapeutic approach for CMT2A and other untreatable diseases of impaired neuronal mitochondrial dynamism and/or trafficking.

Mitochondria are organelles that generate a rich energy source for cells, which requires their continuous subcellular redistribution via mitochondrial trafficking and mutual repair via mitochondrial fusion (*1*). Mitochondrial fusion and subcellular trafficking are mediated in part by mitofusin 1 (MFN1) and MFN2. Genetic mutations in *Mfn2* that suppress mitochondrial fusion and motility cause Charcot-Marie-Tooth disease type 2A (CMT2A), the most common inheritable axonal neuropathy (*2*). Because no therapeutics exist that directly enhance mitochondrial fusion or trafficking, this disease is unrelenting and irreversible.

Computational modeling based on the closed structure of bacterial dynamin–related protein (BDRP) and the more open structure of optic atrophy-1 suggested that MFN2 can change conformation according to how closely the first and second heptad repeat (HR) domains interact (fig. S1). A closed conformation is fusion incompetent, whereas an open conformation favoring mitochondrial fusion can be induced by a competing peptide analogous to amino acids 367 to 384 within the MFN2 HR1 domain (*3*). We identified amino acids controlling these events, first by truncation analysis to define the smallest fusion-promoting minipeptide (residues 374 to 384) (Fig. 1, A and B) and then through functional investigation of this minimal peptide by alanine (Ala) scanning. Substitution of Ala for Met^{376}, Ser^{378}, His^{380}, and Met^{381}, which are highly conserved across vertebrate species (figs. S2 and S3), impaired minipeptide-stimulated mitochondrial fusion, as measured by an increase in the mitochondrial length/width ratio (aspect ratio) (Fig. 1C). The structural model of human MFN2 in a closed conformation on the basis of homology with BDRP predicted a helical interaction between HR1 and HR2 domains, with alignment of Met^{376} and His^{380} side chains in the HR1 domain to Leu^{727} and Asp^{725} in the HR2 domain (fig. S1). This arrangement suggested that Met^{376} and His^{380} stabilize the MFN2 HR1-HR2 interaction, potentially explaining their critical function as defined by minipeptide Ala scanning. By contrast, Ser^{378} was modeled as extending from the noninteracting surface of the HR1 α helix (fig. S1), implying a different mechanism for its involvement in mitochondrial fusion.

To address whether Ser^{378} might be phosphorylated, we replaced Ser^{378} (with Ala, Cys, Asn, or Gly) in the minipeptide and found that phosphorylation and fusion activity were abrogated. Functionality was restored by substituting Asp to mimic phosphorylated Ser or by inserting phospho-Ser itself (Fig. 1D and fig. S4). Moreover, in an in vitro binding assay devoid of cellular kinases, the Asp^{378}-containing minipeptide bound to its putative HR2 interacting domain whereas Ser^{378} and Ala^{378} minipeptides did not (Fig. 1E). Elimination of minipeptide binding by replacement of HR2 Leu^{724}, Asp^{725}, and Leu^{727} with Ala confirmed the HR1-HR2 interaction model (Fig. 1F).

Nuclear magnetic resonance spectrometry of the minipeptides showed low conformational stability with a propensity to form helical structures. Ser^{378} phosphorylation reduced the peptide dynamics most visibly for residues Leu^{379} to Met^{381}, potentially changing amino acid side chains presented to HR2 (fig. S5, data S1, and movie S1). Recombinant MFN2 mutations that replaced Ser^{378} with Asp (mimicking MFN2 Ser^{378} phosphorylation) or substituted Ala for Met^{376} or His^{380} (disrupting the putative HR1-HR2 interaction controlled by Ser^{378} phosphorylation) impaired MFN2-stimulated mitochondrial fusion (fig. S6). By contrast, replacing MFN2 Ser^{378} with Ala (to prevent phosphorylation) or substituting Ala for neighboring Val^{372}, which is not critical for HR1-HR2 interactions, did not depress MFN2-mediated fusion (fig. S6).

MFN2 can be phosphorylated by mitochondrial PTEN-induced putative kinase 1 (PINK1) (*4*, *5*). Targeted mass spectrometry demonstrated phosphorylation of MFN2 Ser^{378}, as well as MFN2 Thr^{111} and Ser^{442}, which were previously reported (*4*, *5*), by PINK1 kinase (Fig. 1G, figs. S7 and S8, and table S1) but not by software-nominated G-protein receptor kinase 2 (fig. S9). We expressed MFN2 Ser^{378} mutants with or without PINK1 in MFN1 and MFN2 doubly deficient cells (designated MFN1$^{-/-}$ MFN2$^{-/-}$). Fusion-defective mitochondria in these cells were abnormally short at baseline, but forced expression of wild-type (WT) Ser^{378} MFN2 resulted in elongation from restoration of fusion (Fig. 1H and fig. S10). Coexpression of PINK1 with MFN2 or mutational replacement of MFN2 Ser^{378} with Asp (which mimics PINK1-mediated Ser^{378} phosphorylation) restrained MFN2-stimulated elongation (Fig. 1H and fig. S10). By contrast, MFN2 Ala^{378} (which cannot be phosphorylated) promoted mitochondrial fusion resistant to PINK1 suppression (Fig. 1H and fig. S10). The effects of MFN2 Ser^{378} mutants were recapitulated in in vitro assays of fusion-mediated mitochondrial content exchange (fig. S11).

We assessed fusogenic activity of commercially available small-molecule candidate pharmacophores (data S2 and supplementary materials and methods), focusing on those having structures that mimicked Ser^{378}-phosphorylated (class A) and -unphosphorylated (class B) minipeptide amino acid side chains (fig. S12 and data S3). We reasoned that simultaneous application of class A and B agonists could enhance

[1]Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO, USA. [2]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA. [3]Departments of Biochemistry and Medicine, Wilf Family Cardiovascular Research Institute, Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, NY, USA. [4]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [5]Departments of Medicine and Cell Biology, Wilf Family Cardiovascular Research Institute, Albert Einstein Cancer Center, and Einstein-Mount Sinai Diabetes Research Center, Albert Einstein College of Medicine, Bronx, NY, USA. [6]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: gdorn@wustl.edu

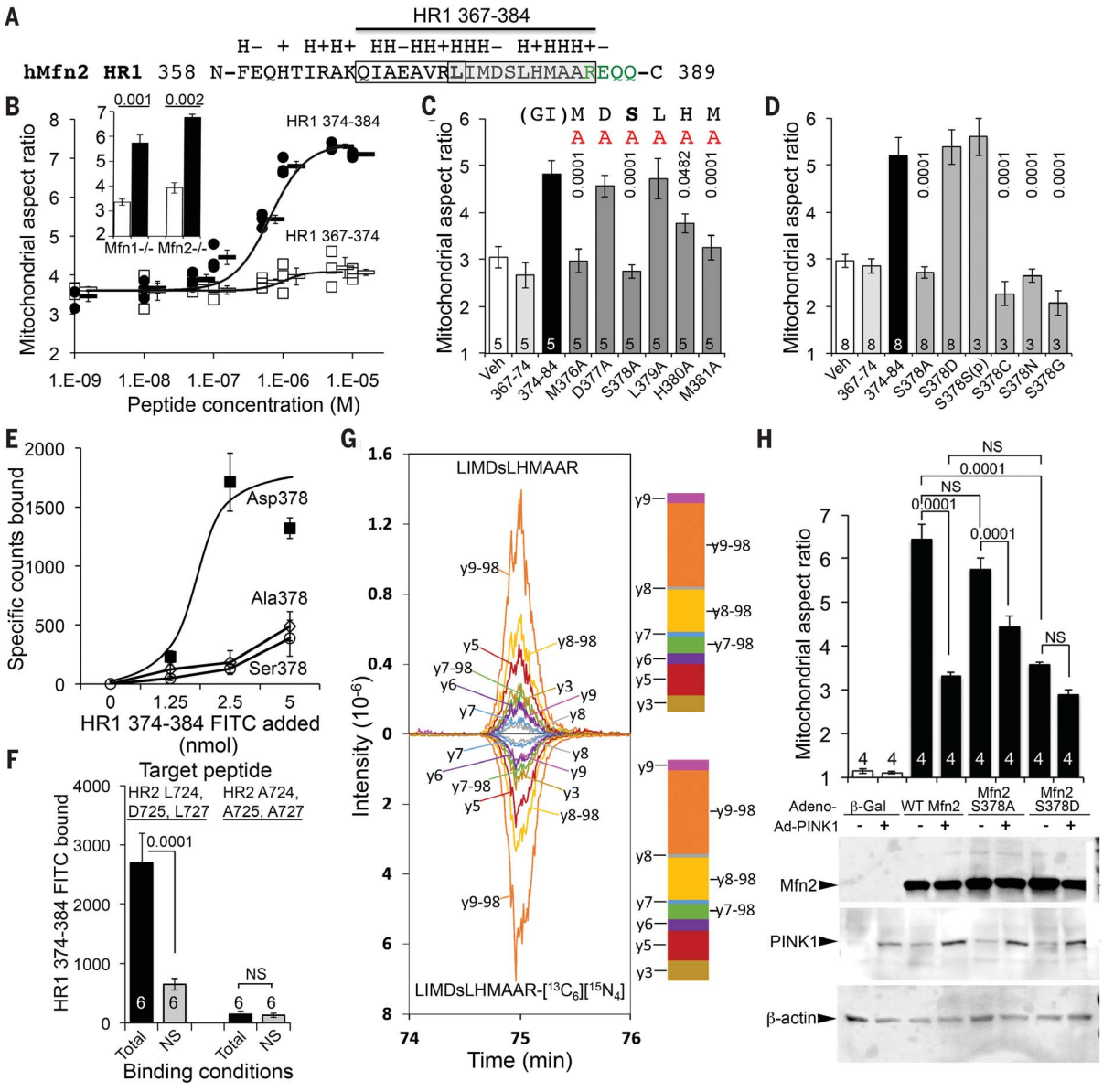


Fig. 1. MFN2 Ser^{378} phosphorylation by PINK1 regulates mitochondrial fusion. (**A**) Amino acid sequence surrounding the fusion-promoting MFN2 peptide. Side chain characteristics (H, hydrophobic; +, basic; −, acidic) are indicated above. HR1 hinge region amino acids are green. The open box encloses N-terminal residues 367 to 374, and the shaded box encloses the C-terminal minipeptide comprising residues 374 to 384 (minipeptide 374–384). Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. hMfn2, human MFN2. (**B**) Mitochondrial fusion stimulated by N- and C-terminal minipeptides. The aspect ratio is the ratio of the mitochondrial long axis to the short axis. (Inset) Fusion in MFN1- or MFN2-null MEFs. White elements correspond to treatment with HR1 minipeptide 367–374, and black elements correspond to treatment with HR1 minipeptide 374–384. Data are means ± SEM of results from three independent experiments; *P* values (shown above the bars in the inset) were determined by Student's *t* test. (**C**) Alanine (A) scanning of minipeptide 374–384 fusion activity. Veh, vehicle. (**D**) Ser^{378} substitution analysis of minipeptide 374–384 fusion activity in MFN2-null MEFs. Data in (C) and (D) are means ± SEM of results from three, five, or eight independent experiments, as indicated. *P* values in (C) and (D) (shown above the bars) are versus the parent minipeptide 374–384 [analysis of variance (ANOVA)]. (**E**) Binding of minipeptides with Ser^{378} substitutions to the HR2 target sequence. Data are means ± SEM of result from six experimental replicates. (**F**) Binding of the Asp^{378} minipeptide to the HR2 target sequence before (left) and after (right) Ala substitution for putative interacting amino acids. Data are means ± SEM of results from six experimental replicates. *P* values were determined by Student's *t* test. NS (above bars), not significant; NS (below bars), nonspecific binding; FITC, fluorescein isothiocyanate. (**G**) Ion chromatograms from assigned MFN2 Ser^{378} phosphopeptide fragment ions (identified by "y" designations) after incubation with PINK1 kinase (top) and a stable isotope-labeled synthetic counterpart (bottom). Proportional intensities are in adjacent stack plots. The graph shows coelution of the synthetic phosphopeptide with the PINK1 product, as well as the positive identification of the product ions from tandem mass spectrometry. The lowercase "s" in the sequence indicates the phosphorylated residue. (**H**) Mitochondrial fusion promoted by MFN2 Ser^{378} mutants with and without PINK1 kinase. An immunoblot of protein expression is shown at the bottom. Data are means ± SEM of results from four independent experiments. *P* values were determined by ANOVA. β-Gal, β-galactosidase.

mitofusin function by acting on both MFN2 Ser^{378} phosphorylation states. Lead compound A (Cpd A) and Cpd B acted synergistically to promote mitochondrial fusion (fig. S13; compare aspect ratios with those in fig. S12C). Therefore, we assimilated Cpd A and Cpd B functionality into a single molecule by creating Cpd A–Cpd B chimeras (Fig. 2A, fig. S14, and data S4). Chimera B-A/long (B-A/l) potently stimulated mitochondrial fusion in MFN2-deficient cells (Fig. 2B), competed for minipeptide binding at the MFN2 HR2 interaction site (Fig. 2C), and was as effective as the combination of Cpd A and Cpd B in reversing mitochondrial dysmorphology provoked by the fusion-defective CMT2A mutant MFN2 Thr^{105}→

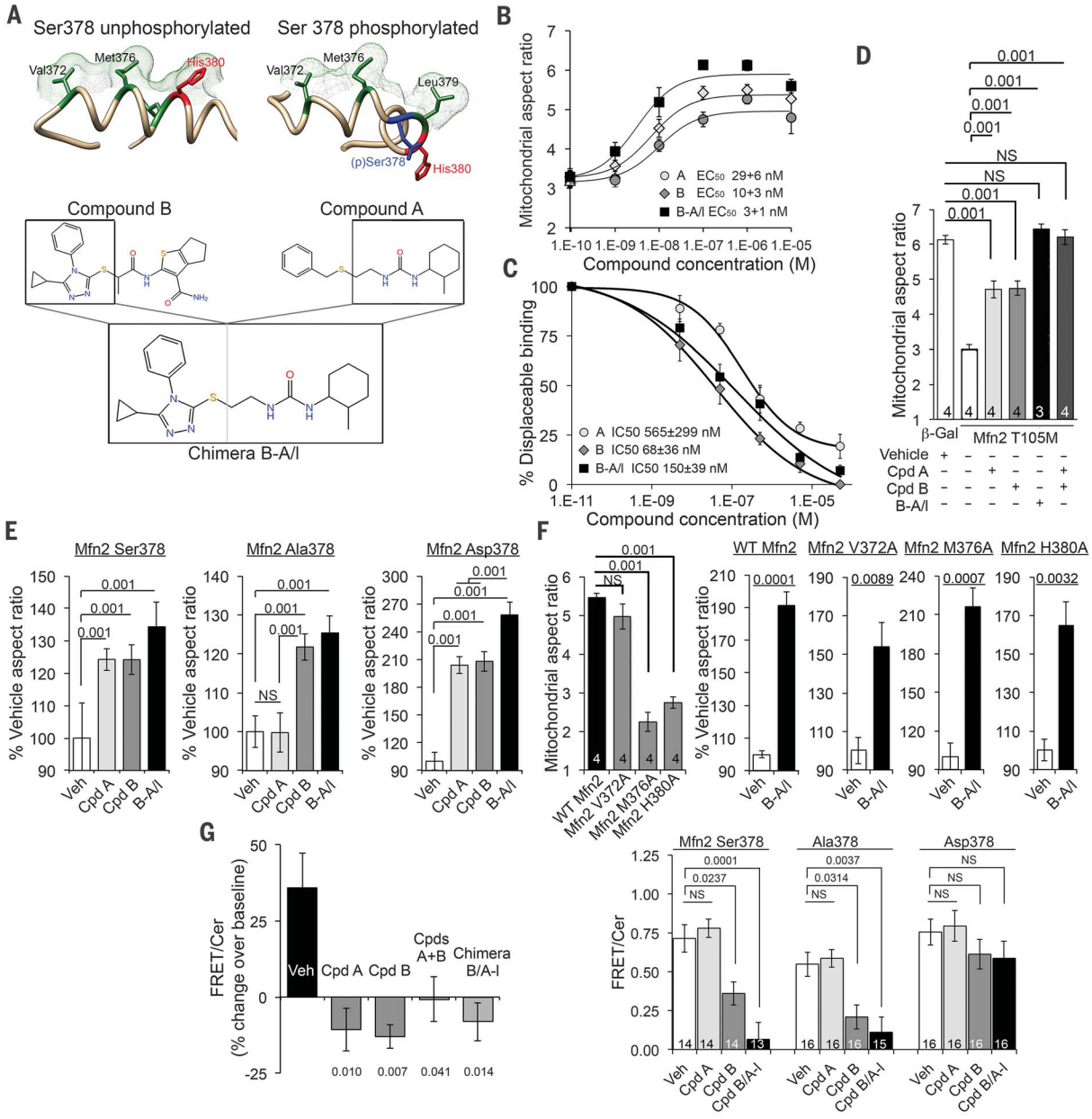


Fig. 2. Small-molecule mimetics of MFN2 HR1 amino acid side chains that interact with HR2 are mitofusin agonists. (**A**) (Top) Three-dimensional representations of hypothetical minipeptide conformations driven by Ser^{378} phosphorylation, and (bottom) their respective small-molecule mimetics. (p)Ser378, phospho-Ser^{378}. (**B**) Dose-dependent mitofusin agonism by small-molecule agonists. Data are means ± SEM of results from six independent experiments. EC_{50}, mean effective concentration. (**C**) Displacement of minipeptide 374–384 from its HR2 binding site by mitofusin agonists. Data are means ± SEM of results from three independent experiments. IC_{50}, mean inhibitory concentration. (**D**) Restoration of MFN2 T105M-impaired mitochondrial fusion in MFN2$^{-/-}$ MEFs by mitofusin agonists. Data are means ± SEM of results from three or four independent experiments, as indicated. *P* values were measured by ANOVA. (**E**) Selectivity of a class A but not a class B mitofusin agonist for Ser^{378}-phosphorylated MFN2. Data are means ± SEM of results from four independent experiments. *P* values were measured by ANOVA. (**F**) Impaired basal function but normal proportional agonist responsiveness of MFN2 with mutations altering HR1-HR2–interacting amino acids. The absolute fusogenicity of these MFN2 mutants is depicted in fig. S6. Data are means ± SEM of results from four independent experiments. *P* values were determined by ANOVA (left) or Student's *t* test (right). (**G**) Change in FRET evoked by mitofusin agonists in isolated mitochondria (left) and intact cells (right). Decreased FRET reflects conformational opening. Data are means ± SEM of results from 3 independent (left) and 14 to 16 replicate (right) experiments. *P* values versus the vehicle were measured by ANOVA. Cer, Cerulean.

Met^{105} (MFN2 T105M) (Fig. 2D). Fusogenic effects of Cpd A were specific for the Asp^{378} mutant of MFN2 that mimicked Ser^{378} phosphorylation, whereas Cpd B and chimera B-A/l were nonselective for the phosphomimic Asp^{378} and nonphosphorylatable Ala^{378} mutants (Fig. 2E). Because they mimic the WT MFN2 HR1 sequence and interact with HR2, mitofusin agonists evoked fusion to proportionally similar degrees in mitochondria expressing mutants of HR1 that are fusion deficient (Fig. 2F; compare with Fig. 1H and fig. S6). Small-molecule mitofusin agonists required endogenous MFN1 or MFN2 to promote mitochondrial fusion, exhibited no detectable promiscuous activity for structurally related dynamin, and did not compromise cell viability (fig. S15). On the basis of fluorescence resonance energy transfer (FRET) analysis of MFN2 labeled at the N and C termini (supplementary materials and methods), mitofusin agonists promoted an open MFN2 conformation favoring mitochondrial fusion (*3*), with a rank order paralleling that for HR2 binding and mitochondrial fusion (Fig. 2G; compare with Fig. 2, B and C), supporting allosteric activation.

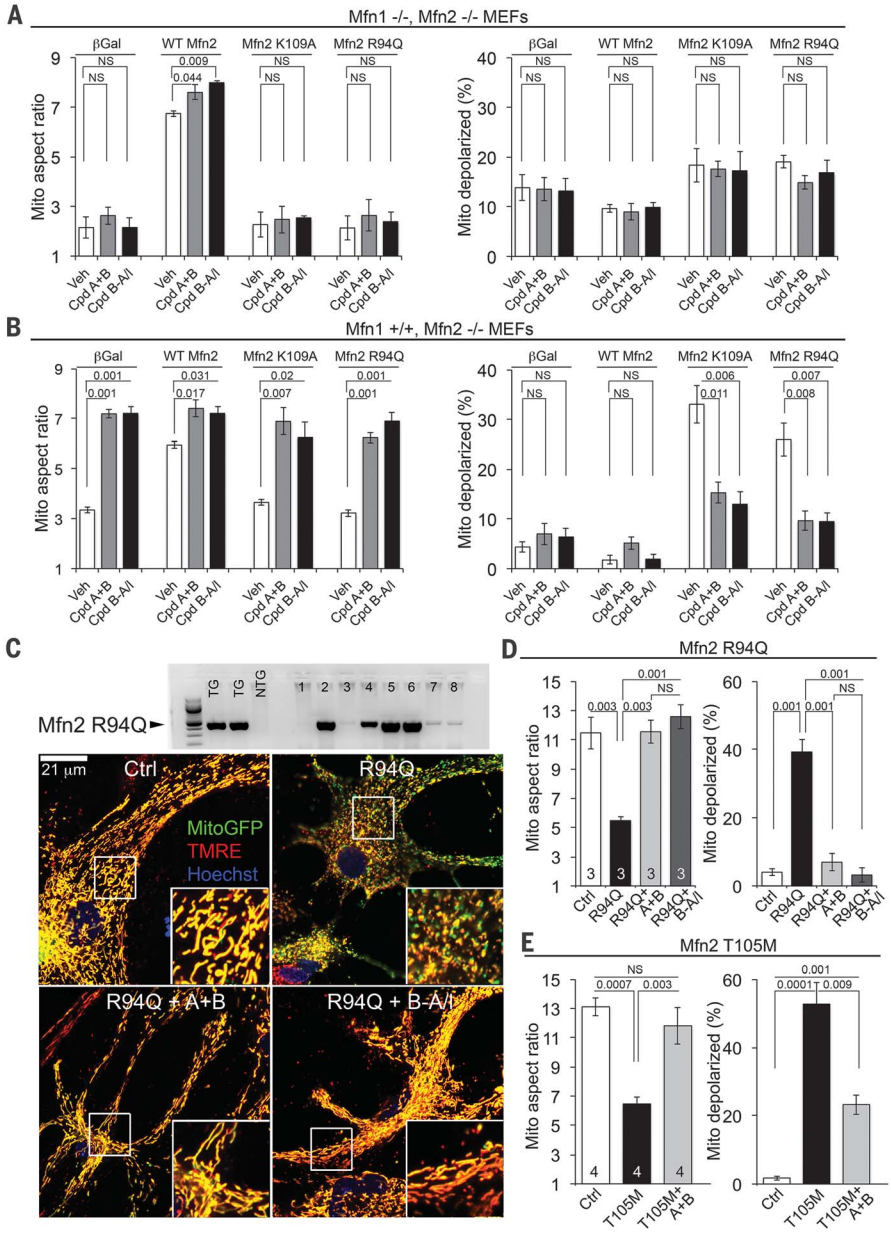


Fig. 3. Mitofusin agonists correct mitochondrial damage induced by nonfunctioning MFN2 mutants by activating endogenous mitofusins. (**A** and **B**) Effects of mitofusin agonists in MFN1$^{-/-}$ MFN2$^{-/-}$ cells (A) and in MFN1$^{+/+}$ MFN2$^{-/-}$ cells (B) expressing recombinant WT or mutant MFN2. Data are means ± SEM of results from three independent experiments. *P* values were measured by ANOVA. Mito, mitochondria. (**C**) (Top) Genotyping of MFN2 R94Q individual mouse pups from which primary cultured neurons were derived. TG, transgenic MFN2 R94Q; NTG, nontransgenic. (Bottom) Representative confocal images of mitochondrial pathology in cultured neonatal mouse neurons expressing MFN2 R94Q, and correction by mitofusin agonists. Magnified views are from white squares. Ctrl, control; MitoGFP, mitochondrion-targeted green fluorescent protein; TMRE, tetramethylrhodamine ethyl ester. (**D**) Group data for mitochondrial aspect ratio and polarization studies in cultured neonatal mouse neurons expressing MFN2 R94Q, as depicted in (C). Data are means ± SEM of results from three independent experiments. *P* values were measured by ANOVA. (**E**) Results of mitochondrial aspect ratio and polarization studies in cultured neonatal mouse neurons expressing MFN2 T105M. Data are means ± SEM of results from four independent experiments. *P* values were measured by ANOVA.

In CMT2A, MFN2 mutants produce mitochondrial "fragmentation" (defined as decreased aspect ratio) and loss of normal membrane polarization through dominant inhibition of normal mitofusins. Experiments using MFN1$^{-/-}$ MFN2$^{-/-}$ murine embryonic fibroblasts (MEFs) showed that in the absence of normal mitofusins, small-molecule mitofusin agonists did not improve mitochondria of cells expressing the guanosine triphosphatase (GTPase)–crippled MFN2 Arg^{94}→Gln^{94} (R94Q) or Lys^{109}→Ala^{109} (K109A) mutant (Fig. 3A). However, mitofusin agonists corrected the mitochondrial dysmorphology and reversed the mitochondrial hypopolarization induced by these MFN2 mutants when endogenous MFN1 was present (Fig. 3B). Mitofusin agonists also reversed mitochondrial fragmentation and hypopolarization in cultured mouse neurons expressing (in addition to endogenous mitofusins) CMT2A mutant MFN2 R94Q (Fig. 3, C and D) or CMT2A mutant MFN2 T105M (Fig. 3E). Thus, mitofusin agonists do not restore function of CMT2A MFN2 GTPase domain mutants. Rather, by stabilizing the fusion-permissive open conformation of endogenous normal MFN1 or MFN2, mitofusin agonists can overcome dominant suppression of mitochondrial fusion by these disease-causing dysfunctional proteins.

Clinical CMT2A classically affects long nerves innervating the lower and upper limbs (*6*, *7*). It is unclear how a principal defect in mitochondrial fusion would cause length-dependent neuronal disease. Conversely, disruption of axonal mitochondrial trafficking (*8*) would be predicted to preferentially affect cells requiring mitochondrial transport over the greatest physical distance, such as the sciatic nerves originating in the spine and terminating in the foot. MFN2 interacts with the Miro/Milton complex to promote mitochondrial motility in neurons (*9*), so we tested the effects of mitofusin agonism on murine neuronal mitochondrial trafficking. Chimera B-A/l reversed mitochondrial "clumping" (formation of static mitochondrial aggregates) and restored mitochondrial motility in cultured mouse neurons expressing the CMT2A mutant MFN2 T105M (Fig. 4A and movie S2). Mitochondrial hypopolarization and increased autophagy (Fig. 4B and fig. S16) and mitochondrial dysmorphology (Fig. 4C and fig. S16) were concomitantly ameliorated. Thus,

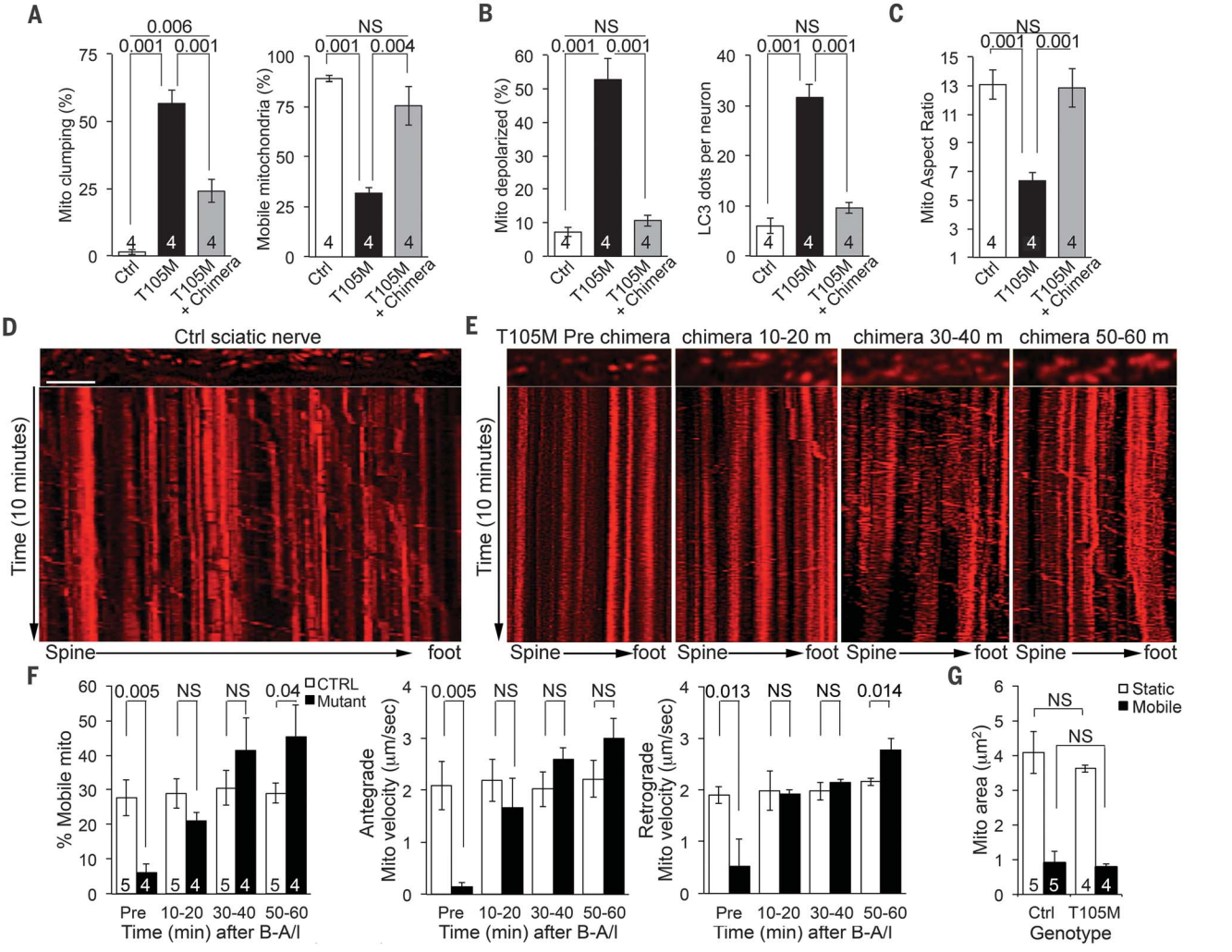


Fig. 4. A mitofusin agonist restores axonal mitochondrial trafficking suppressed by CMT2A mutant MFN2 T105M. (**A** to **C**) Effects of chimera B-A/I on mitochondrial mobility (A), function (B), and morphology (C) in cultured CMT2A MFN2 T105M–expressing mouse neurons. Data are means ± SEM of results from four independent experiments. *P* values were measured by ANOVA. (**D**) Representative kymograph of mitochondrial trafficking in a Ctrl mouse sciatic nerve. Scale bar, 10 μm. (**E**) Representative serial kymographs of mitochondria in an MFN2 T105M mouse sciatic nerve before and after treatment with chimera B-A/I. m, minutes. (**F**) Quantitative data for sciatic nerve mitochondrial motility studies. (**G**) Sizes of motile and static mitochondria in Ctrl and B-A/I–treated (60 min) sciatic nerves. Data in (F) and (G) are means ± SEM of results from four or five independent experiments. *P* values were measured by ANOVA.

a small-molecule mitofusin agonist enhanced organelle and cell fitness in CMT2A neurons by promoting mitochondrial fusion and subcellular transport.

We evaluated the concept of activating mitofusins to stimulate in vivo axonal mitochondrial trafficking in sciatic nerves of mice expressing the CMT2A mutant MFN2 T105M. In normal sciatic nerves, ~30% of axonal mitochondria exhibited robust bidirectional transport (Fig. 4D and movie S3). Mitochondria of MFN2 T105M sciatic nerves were severely hypomotile (Fig. 4E and movie S4), but application of chimera B-A/I to MFN2 T105M sciatic nerves restored mitochondrial motility to within normal levels (Fig. 4F and movie S4). Mobile mitochondria in WT and B-A/I–treated MFN2 T105M axons were smaller than static mitochondria (Fig. 4G), supporting in vitro observations discriminating between MFN2-mediated mitochondrial dysmotility and defective fusion in CMT2A (*10*).

In this study, we found that PINK1 phosphorylation of MFN2 at Ser^{378} can alter the positions of Met^{376} and His^{380} (in the HR1 domain), which normally interact with HR2 domain amino acids to orchestrate MFN2 toggling between conformations that modulate mitochondrial fusion. These findings establish a mechanistic basis for clinical observations that MFN2 Met^{376} mutations to Ile, Thr, and Val can cause CMT2A (*7*, *11*, *12*).

By combining in silico pharmacophore modeling with structural and functional interrogation of MFN2 HR1 domain–derived minipeptides, we developed a novel small-molecule mitofusin agonist that reversed the neuronal mitochondrial dysmorphometry and impaired mobility evoked by two CMT2A *Mfn2* mutants. CMT2A is the prototypical clinical disorder of defective mitochondrial fusion, but impaired mitochondrial trafficking may play as great a role as mitochondrial fragmentation in CMT2A axonal degeneration (*8*–*10*). Individuals with CMT2A express one mutant MFN2 allele in combination with one normal MFN2 allele and harbor two normal MFN1 alleles (*13*). It is therefore possible that a therapeutic substrate exists for mitofusin agonists to "supercharge" normal mitofusins and overcome dominant inhibition by MFN2 mutants. Our observation that in vivo mitochondrial dysmotility provoked

by a CMT2A mutant can be normalized by mitofusin agonism mechanistically links abnormal mitochondrial trafficking in experimental CMT2A to MFN2 dysfunction. Mitofusin agonists may also have therapeutic potential for neurological conditions other than CMT2A, such as Alzheimer's, Parkinson's, and Huntington's diseases, wherein mitochondrial dysmotility and fragmentation are contributing factors (*14*–*16*).

REFERENCES AND NOTES

1. D. C. Chan, *Annu. Rev. Genet.* **46**, 265–287 (2012).
2. M. Liesa, M. Palacín, A. Zorzano, *Physiol. Rev.* **89**, 799–845 (2009).
3. A. Franco *et al*., *Nature* **540**, 74–79 (2016).
4. Y. Chen, G. W. Dorn II, *Science* **340**, 471–475 (2013).
5. G. Gong *et al*., *Science* **350**, aad2459 (2015).
6. V. H. Lawson, B. V. Graham, K. M. Flanigan, *Neurology* **65**, 197–204 (2005).
7. K. Verhoeven *et al*., *Brain* **129**, 2093–2102 (2006).
8. R. H. Baloh, R. E. Schmidt, A. Pestronk, J. Milbrandt, *J. Neurosci.* **27**, 422–430 (2007).
9. A. Misko, S. Jiang, I. Wegorzewska, J. Milbrandt, R. H. Baloh, *J. Neurosci.* **30**, 4232–4240 (2010).
10. A. L. Misko, Y. Sasaki, E. Tuck, J. Milbrandt, R. H. Baloh, *J. Neurosci.* **32**, 4145–4155 (2012).
11. K. W. Chung *et al*., *Brain* **129**, 2103–2118 (2006).
12. C. Casasnovas *et al*., *J. Med. Genet.* **47**, 249–256 (2010).
13. F. Bombelli *et al*., *JAMA Neurol.* **71**, 1036–1042 (2014).
14. K. Hirai *et al*., *J. Neurosci.* **21**, 3017–3023 (2001).
15. K. F. Winklhofer, C. Haass, *Biochim. Biophys. Acta* **1802**, 29–44 (2010).
16. J. Kim *et al*., *Hum. Mol. Genet.* **19**, 3919–3935 (2010).

ACKNOWLEDGMENTS

We gratefully acknowledge discussions with P. Needleman and the assistance of L. Zhang, P. Erdmann-Gilmore, Y. Mi, and R. Connors. **Funding:** This work was supported by NIH grants R35HL135736 (G.W.D.), R01HL128071 (R.N.K. and G.W.D.), and R01CA178394 and P30CA013330 (E.G.); a McDonnell Center for Cellular and Molecular Neurobiology postdoctoral fellowship (A.F.); and the Washington University Proteomics Shared Resource, supported by National Center for Advancing Translational Sciences grants UL1TR000448, NIGMSP41, GM103422, and NCIP30 CA091842. G.W.D. is the Philip and Sima K. Needleman–endowed professor. **Author contributions:** G.W.D., A.M.K., D.M.-R, and R.R.T. conceived of or designed the research, except the initial in silico screen. E.G., R.N.K., N.B., and E.Z. conceived of the small-molecule screen, designed the original pharmacophore model, and performed the initial in silico screen. G.W.D. wrote the manuscript. J.M.R. and R.R.T. performed phosphoprotein mass spectroscopy analyses. A.M.K. performed peptide nuclear magnetic resonance studies. A.G.R. screened compounds for activity and characterized agonists. A.G.R., A.F., J.M.A., and W.C.K. performed mitochondrial studies. A.F. performed cultured neuron and ex vivo sciatic nerve studies. J.W.J. analyzed and purified compounds. R.N.K., E.G., and R.H.B. provided essential reagents. **Competing interests:** D.M.-R. and G.W.D. are inventors on patent application 15/710,696, submitted by Stanford University, which covers the use of peptide regulators of mitochondrial fusion and small-molecule peptidomimetics derived from them. G.W.D. is an inventor on provisional patent applications 62/488,787 and 62/584,515, submitted by Washington University, which cover the use of novel small-molecule mitofusin agonists to treat chronic neurodegenerative diseases. E.G., R.N.K., N.B., and E.Z. are inventors on patent application 62/573,217, submitted by Albert Einstein College of Medicine, which covers compositions of mitofusin agonists and their uses for the treatment of diseases and disorders. D.M.-R. is the founder of Mitoconix Bio, a company focused on improving mitochondrial health as a therapeutic approach for neurodegenerative diseases. None of the research conducted in D.M.-R.'s laboratory is supported by or performed in collaboration with Mitoconix. The other authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. There are no material transfer agreements associated with this study.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/336/suppl/DC1
Materials and Methods
Figs. S1 to S31
Table S1
References (*17*–*30*)
Movies S1 to S4
Data S1 to S4

22 June 2017; resubmitted 3 October 2017
Accepted 27 February 2018
10.1126/science.aao1785

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

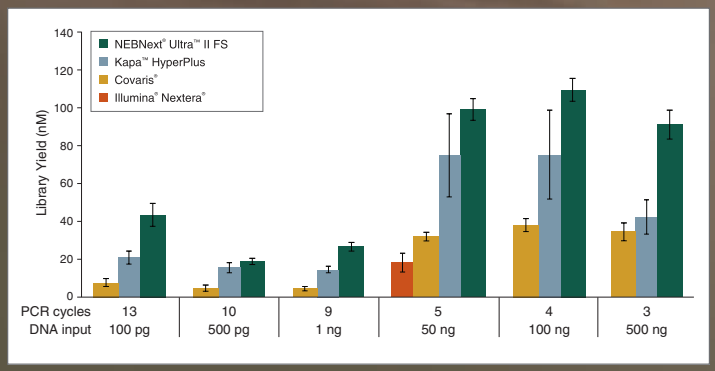
You'll be thrilled to pieces.

NEBNext® Ultra™ II FS DNA Library Prep Kit
Novel Enzymatic Fragmentation System

Do you need a faster, more reliable solution for DNA fragmentation and library construction? The NEBNext Ultra II FS DNA Library Prep Kit meets the dual challenge of generating high quality next gen sequencing libraries from ever-decreasing input amounts AND simple scalability. Enzymatic shearing increases library yields by reducing DNA damage and sample loss. Our novel fragmentation reagent is combined with end repair and dA-tailing reagents, and a single protocol is used for a wide range of input amounts (100 pg – 500 ng) and sample types.

You'll be thrilled to pieces with the result – a reliable, flexible, high-quality library prep that is fast and scalable.

NEBNext Ultra II FS DNA produces the highest yields, from a range of input amounts



Libraries were prepared from Human NA19240 genomic DNA using the input amounts and numbers of PCR cycles shown. For NEBNext Ultra II FS, a 20-minute fragmentation time was used. For Kapa HyperPlus libraries, input DNA was cleaned up with 3X beads prior to library construction, as recommended, and a 20-minute fragmentation time was used. Illumina recommends 50 ng input for Nextera, and not an input range; therefore, only 50 ng was used in this experiment. "Covaris" libraries were prepared by shearing each input amount in 1X TE Buffer to an insert size of ~200 bp using a Covaris instrument, followed by library construction using the NEBNext Ultra II DNA Library Prep Kit (NEB #E7645). Error bars indicate standard deviation for an average of 3–6 replicates performed by 2 independent users.

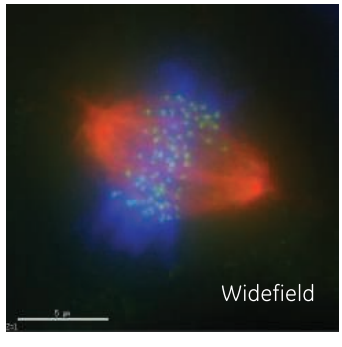
Visit **NEBNextUltraII.com** to request your sample today.

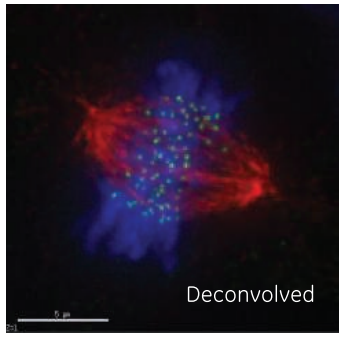
GE Healthcare

Focus on discovery with the DeltaVision Ultra™

An easy-to-use, automated widefield microscope, DeltaVision Ultra enhances data quality so you can see the structures that matter in time-lapse live cell imaging, multi-point cell tracking, and multi-well plate scanning.

Learn about the microscope that supports all your research needs at **gelifesciences.com/deltavisionultra**





imagination at work

eppendorf & Science

PRIZE FOR NEURO BIOLOGY

2017 Winner
Flavio Donato, Ph.D.
Kavli Institute
Norwegian University of
Science and Technology

For research on how neural networks mature during development to represent space in the brain

Application Deadline
June 15, 2018

Call for Entries

Eppendorf & Science Prize for Neurobiology
The annual Eppendorf & Science Prize for Neurobiology is an international award which honors young scientists for their outstanding contributions to neurobiological research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by *Science's* Senior Editor, Dr. Peter Stern. To be eligible, you must be 35 years of age or younger.

As the Grand Prize Winner, you could be next to receive
> Prize money of US$25,000
> Publication of your work in *Science*
> Full support to attend the Prize Ceremony held in conjunction with the Annual Meeting of the Society for Neuroscience in the USA
> 10-year AAAS membership and online subscription to *Science*
> Complimentary products worth US$1,000 from Eppendorf
> An invitation to visit Eppendorf in Hamburg, Germany

It's easy to apply! Learn more at:

www.eppendorf.com/prize

eppendorf Science AAAS

eppendorf & Science

PRIZE FOR NEURO BIOLOGY

2017 Winner
Flavio Donato, Ph.D.
Kavli Institute
Norwegian University of
Science and Technology

For research on how neural networks mature during development to represent space in the brain

Application Deadline
June 15, 2018

Call for Entries

Eppendorf & Science Prize for Neurobiology
The annual Eppendorf & Science Prize for Neurobiology is an international award which honors young scientists for their outstanding contributions to neurobiological research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by *Science's* Senior Editor, Dr. Peter Stern. To be eligible, you must be 35 years of age or younger.

As the Grand Prize Winner, you could be next to receive
> Prize money of US$25,000
> Publication of your work in *Science*
> Full support to attend the Prize Ceremony held in conjunction with the Annual Meeting of the Society for Neuroscience in the USA
> 10-year AAAS membership and online subscription to *Science*
> Complimentary products worth US$1,000 from Eppendorf
> An invitation to visit Eppendorf in Hamburg, Germany

It's easy to apply! Learn more at:

www.eppendorf.com/prize

eppendorf Science AAAS

eppendorf & Science

PRIZE FOR NEURO BIOLOGY

2017 年获奖者
Flavio Donato, Ph.D.
卡弗里研究所
挪威科技大学

专注发育期间神经网络成熟如何构建大脑空间的研究

申请截止日期
2018 年 6 月 15 日

欢迎参选

Eppendorf & Science 神经生物学奖
一年一度的 Eppendorf & Science 神经生物学奖是一项国际奖项，授予用分子与细胞生物学方法在神经生物学领域取得非凡成果的青年科学家。冠军及入围候选人均是由《Science》杂志高级编辑 Peter Stern 博士领衔的独立科学家所组成的委员会评出。年龄不超过 35 岁的科学家可以申请参选。

您可能就是下一位获奖者并获得：
> $25,000 美元奖金
> 获奖论文发表在《Science》杂志上
> 得以全额资助参与美国神经科学协会年会和颁奖仪式
> 10 年 AAAS 会员和《Science》电子期刊
> 赠送价值 $1,000 美元的 Eppendorf 产品
> 获邀参观 Eppendorf 位于德国汉堡的总部

申请非常容易！
要了解更多信息，请登陆：

www.eppendorf.com/prize

eppendorf Science AAAS

eppendorf & Science

PRIZE FOR NEURO BIOLOGY

2017 Winner
Flavio Donato, Ph.D.
Kavli Institute
Norwegian University of
Science and Technology

For research on how neural networks mature during development to represent space in the brain

申込期限
2018年6月15日

Call for Entries

Eppendorf & Science for Neurobiology
Eppendorf & Science神経生物学賞は、分子生物学や細胞生物学に基づく神経生物学研究において、卓越した貢献をなされた若手科学者に贈呈している国際賞です。最終選考選出者ならびに受賞者は、Science誌編集主任Dr. Peter Sternをはじめとする科学者達の独立委員会によって選出されます。35歳までの研究者が応募条件できます。

As the Grand Prize Winner, you could be next to receive
> 賞金：25,000米ドル
> Science誌に研究内容を掲載
> 米神経科学学会年次総会に併せて開催される授賞式への参加を全面サポート
> 10年間のAAAS会員資格およびScience誌のオンライン購読権
> 1,000米ドル相当のエッペンドルフ製品を提供
> ドイツ、ハンブルクのエッペンドルフ本社にご招待

応募は簡単です！ 詳しくはこちら：

www.eppendorf.com/prize

eppendorf Science AAAS

The 6 sharpest perspectives

for focused science and technology solutions in life science

Our mission is to collaborate with the global scientific community to solve the toughest problems in life science. To achieve this, we have brought together the world's leading Life Science brands to create a world-class portfolio: Millipore®, SAFC®, BioReliance®, Sigma-Aldrich®, Milli-Q® and Supelco®.

A science and technology ecosystem

MilliporeSigma has brought together the world's leading Life Science brands to solve the toughest problems in life science. In particular, you may be interested in the following brands:

Sigma-Aldrich® continues to develop its broad portfolio of state-of-the-art lab and production materials, paired with technical support and scientific partnerships.
Milli-Q® offers a continually pioneering range of intuitive, easy-to-use lab water instruments that seamlessly integrate into your daily work.
Millipore® provides proven preparation, separation, filtration and testing products and technologies.
Supelco® continues to be a trusted resource for accurate and reliable analytical products, developed by analytical chemists, for analytical chemists.

Support you can count on

At MilliporeSigma, we offer a world-class portfolio with over 300,000 high-quality products and services – from research to commercial manufacturing.

With 19,000 life science employees in 66 countries across the globe, we have 130 distribution centers and 10 global customer collaboration facilities across North America, Europe, Asia and Latin America.

Customer-centric innovation

We are dedicated to putting our customers first, and by collaborating we can help advance life science. Together, we can improve and expand global access to health with unparallelled services and support under one banner.

Seamless access

With 24/7 customer support, an efficient ordering and delivery service, and an unmatched e-commerce platform, you get exactly what you need, when you need it.

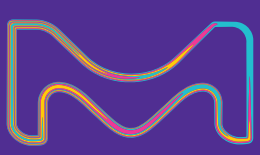
A sharper look

As part of the exciting life science restructure, we are refreshing our existing packaging design and labeling to better serve you and reflect the vibrant nature of our business and vision.

Let us help you

Find out how we collaborate to solve your toughest problems at SigmaAldrich.com/advancinglifescience

#howwesolve



The life science business of Merck KGaA, Darmstadt, Germany operates as MilliporeSigma in the U.S. and Canada.

MilliporeSigma

THE 6 SHARPEST PERSPECTIVES

for focused science and technology solutions in life science

Our mission is to collaborate with the global scientific community to solve the toughest problems in life science. To achieve this, we have brought together the world's leading Life Science brands to create a world-class portfolio: Millipore®, SAFC®, BioReliance®, Sigma-Aldrich®, Milli-Q® and Supelco®.

A science and technology ecosystem

Merck has brought together the world's leading Life Science brands to solve the toughest problems in life science. In particular, you may be interested in the following brands:

Sigma-Aldrich® continues to develop its broad portfolio of state-of-the-art lab and production materials, paired with technical support and scientific partnerships.
Milli-Q® offers a continually pioneering range of intuitive, easy-to-use lab water instruments that seamlessly integrate into your daily work.
Millipore® provides proven preparation, separation, filtration and testing products and technologies.
Supelco® continues to be a trusted resource for accurate and reliable analytical products, developed by analytical chemists, for analytical chemists.

Support you can count on

At Merck, we offer a world-class portfolio with over 300,000 high-quality products and services – from research to commercial manufacturing.

With 19,000 life science employees in 66 countries across the globe, we have 130 distribution centres and 10 global customer collaboration facilities across North America, Europe, Asia and Latin America.

Customer-centric innovation

We are dedicated to putting our customers first, and by collaborating we can help advance life science. Together, we can improve and expand global access to health with unparallelled services and support under one banner.

Seamless access

With 24/7 customer support, an efficient ordering and delivery service, and an unmatched e-commerce platform, you get exactly what you need, when you need it.

A sharper look

As part of the exciting life science restructure, we are refreshing our existing packaging design and labelling to better serve you and reflect the vibrant nature of our business and vision.

Let us help you

Find out how we collaborate to solve your toughest problems at SigmaAldrich.com/advancinglifescience

#howwesolve

The life science business of Merck operates as MilliporeSigma in the U.S. and Canada.

MERCK

6 powerful propellers

to accelerate expert science and technology solutions in life science

Millipore®
Proven Preparation, Separation, Filtration & Testing Products

SAFC®
Tailored Pharma & Biopharma Raw Material Solutions

BioReliance®
Expert Pharma & Biopharma Manufacturing & Testing Services

Sigma-Aldrich®
State-of-the-Art Lab & Production Materials

Milli-Q®
Pioneering Lab Water Solutions

Supelco®
Trusted Analytical Products

MilliporeSigma has brought together the world's leading Life Science brands, so whatever your life science problem, you can benefit from our expert products and services.

To find out how MilliporeSigma can help you work, visit SigmaAldrich.com/advancinglifescience

#howwesolve

The life science business of Merck KGaA, Darmstadt, Germany operates as MilliporeSigma in the U.S. and Canada.

MilliporeSigma

6 powerful propellers

to accelerate expert science and technology solutions in life science

Millipore®
Proven Preparation, Separation, Filtration & Testing Products

SAFC®
Tailored Pharma & Biopharma Raw Material Solutions

BioReliance®
Expert Pharma & Biopharma Manufacturing & Testing Services

Sigma-Aldrich®
State-of-the-Art Lab & Production Materials

Milli-Q®
Pioneering Lab Water Solutions

Supelco®
Trusted Analytical Products

Merck has brought together the world's leading Life Science brands, so whatever your life science problem, you can benefit from our expert products and services.

To find out how the life science business of Merck can help you work, visit SigmaAldrich.com/advancinglifescience

#howwesolve

The life science business of Merck operates as MilliporeSigma in the U.S. and Canada.

MERCK

Tissue Imaging System
The Vectra Polaris Automated Quantitative Pathology Imaging System integrates both multispectral imaging and automated slide scanning to better visualize, analyze, quantify, and phenotype immune cells in situ in FFPE tissue sections and tissue microarrays (TMAs). The fully automated system provides researchers unparalleled speed, performance, and versatility for extracting proteomic and morphometric information from intact FFPE tissue sections and TMAs. Vectra Polaris is the only instrument on the market that goes beyond basic functions such as whole-slide, brightfield, and fluorescence imaging, integrating the power of multispectral imaging in a simplified, elegant workflow. Coupled with the proven power of inForm image analysis software, its multiplexing capability enables identification and quantification of multiple biomarkers and reveals spatial context within a digital workflow, to assist researchers with better, faster decisions.
PerkinElmer
For info: 800-762-4000
www.perkinelmer.com/cancer-immunology

Microscope Camera
The Lumenera INFINITY3-6UR is an ideal general-purpose camera for most microscopy applications due to its 6-MP resolution, excellent color reproduction, and speed, with the light sensitivity needed for low-light applications. Built on Sony's EXview HAD II sensor technology, this camera offers extremely high dynamic range, 4.54 µm x 4.54 µm pixels, and very low noise. With its 2x2 binning, there is a fourfold increase in sensitivity along with 1.5-MP (1376x1096) resolution. INFINITY3-6UR is designed for use in a variety of scientific, life science, clinical, and industrial applications requiring optimal color reproduction, extreme sensitivity, increased resolution, and high speed.
Lumenera
For info: 613-736-4077
www.lumenera.com

Non-Browning Zoom Lens
Resolve Optics' Model 357 non-browning zoom lens provides high-definition-quality images over an unrivaled 10X zoom range. Featuring specialist non-browning glasses, it produces clear, sharp images without the strong yellow tint that has traditionally been a limiting issue when using radiation-resistant lenses on color sensors. It is designed to operate in environments subject to high-level radiation, such as nuclear fuel handling sites, reactor active zones, and nuclear waste storage plants. Operating at f/2, Model 357 offers high image resolution and minimum geometric distortion from 400 nm to 770 nm, and can image objects from 800 mm to infinity without add-on adapters.
Resolve Optics
For info: +44-(0)-1494-777-100
www.resolveoptics.com

Enhanced Confocal Detector
The ZEISS Airyscan is an area detector. Unlike traditional confocal microscopes, which reject photons from outside of the focal plane at the pinhole, Airyscan detects all precious fluorescence emission photons of 1.25 a.u. Their information is then used to deliver higher sensitivity, superresolution, and high acquisition speeds. The new 2D Superresolution mode takes advantage of the fact that Airyscan captures information from the *x*, *y*, and *z* planes with the confocal point spread function. A new exclusive processing algorithm uses this inherent spatial information captured in a single image, specifically distinguishing between photons originating from the focal plane of 0.2 a.u. and photons from outside of this focal plane. In a traditional confocal microscope, a researcher could only close the pinhole to 0.2 a.u. to attempt the same optical sectioning. This would mean sacrificing many photons, even from the focal plane, thus reducing signal-to-noise drastically.
ZEISS
For info: 800-233-2343
www.zeiss.com

Biomolecular Imager
The Amersham Typhoon NIR Plus is a modular, upgradeable near-infrared (NIR) imaging system from GE Healthcare Life Sciences. Its three-laser, flexible configuration provides two essential NIR channels as well as a third visible laser channel for triplex protein detection, to facilitate normalization and the reduction of blot-to-blot variation. Ideal for multiuser environments, it delivers extremely sensitive detection, high image resolution, and a broad linear dynamic range with multifluorescent and stained samples. Able to scan up to nine microtiter plates simultaneously, it is also suitable for high-throughput screening. NIR Plus is an affordable, user-friendly, open imaging platform compatible with most commercially available NIR detection reagents. With high (10-µm) resolution, a dynamic range greater than 5 orders of magnitude, and sensitivity as low as 3 picograms, it accommodates a wide range of samples, including tissue sections, plant leaves, and cell colonies, in addition to traditional gels, membranes, multiwell plates, and dishes.
GE Healthcare Life Sciences
For info: 800-526-3593
bioresearch.gehealthcare.info

Biological Microscopes
The CX33 and CX43 microscopes enable users to remain comfortable during long periods of routine microscopy observations. In both instruments, the stage is 70 mm lower than in the previous model, resulting in a large workspace under the eyepiece that facilitates checking or swapping samples with just one hand. In addition, the stage and coaxial focus knobs have been lowered, so users can reach them while their forearms remain on the desk. The low-positioned revolving nosepiece with a finger rest accommodates up to five objectives, enabling users to quickly change magnification with minimal arm movement. To improve durability and reduce running costs, we have added a centering-free LED light source with a long 60,000-h lifetime, providing uniform illumination and constant color temperature at any brightness level. This means users get natural color representation from a wide range of stains and can perform analyses faster and more confidently.
Olympus
For info: 704-877-8801
www.olympus-lifescience.com/en/microscopes/upright

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.

Alexander von Humboldt
Stiftung/Foundation

more independence + more opportunities + more scope search

Sofja Kovalevskaja Award – Become a research group leader in Germany

€ 1.65 million for young researchers from all disciplines

The Alexander von Humboldt Foundation is now calling for applications for the Sofja Kovalevskaja Award, one of the most valuable academic awards in Germany. With funding of up to € 1.65 million, it creates unique opportunities for excellent researchers to conduct their own research in Germany: award winners are invited to spend five years working at the institute of their choice and setting up their own research groups – independently and virtually free of bureaucracy.

Scientists and scholars of all disciplines from abroad who have completed their doctorates within the last six years are eligible to apply. The award targets outstanding talent and a creative approach to research.

The Humboldt Foundation actively promotes equal opportunities and therefore particularly welcomes applications from women researchers.

The closing date for applications is 31 July 2018. More information: **www.humboldt-foundation.de/SKP_en**

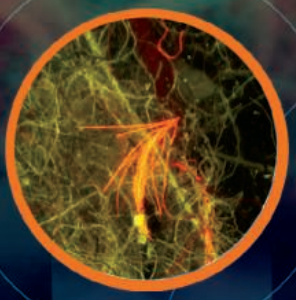
Exzellenz verbindet –
be part of a worldwide network.

Alexander von Humboldt Foundation
Jean-Paul-Str. 12
53173 Bonn
Germany

info@avh.de

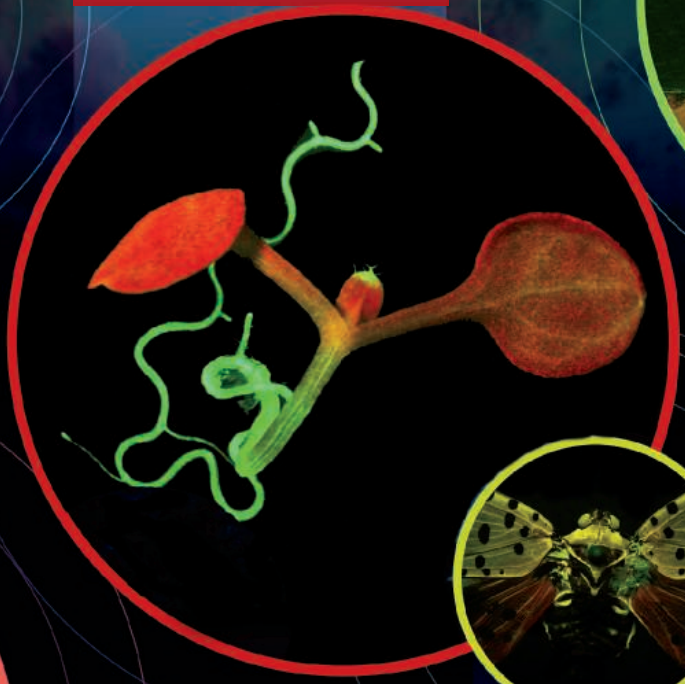
www.humboldt-foundation.de

there's more to see...
WITH FLUORESCENCE

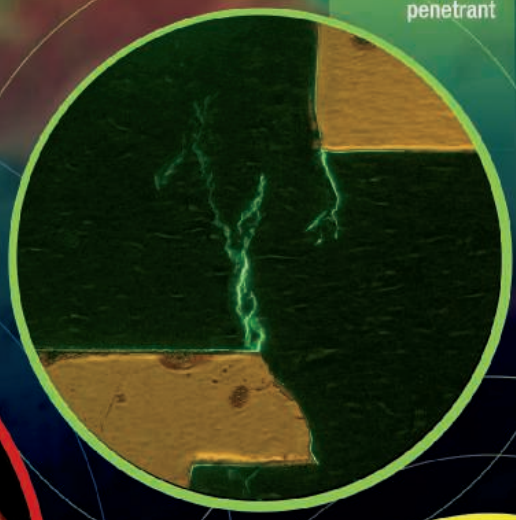


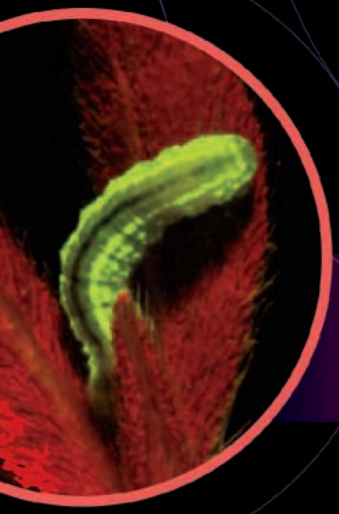
FORENSIC SCIENCE
'Dust bunny' under blue excitation

LIFE SCIENCE
Arabidopsis sp. expressing green fluorescent protein (GFP) in roots



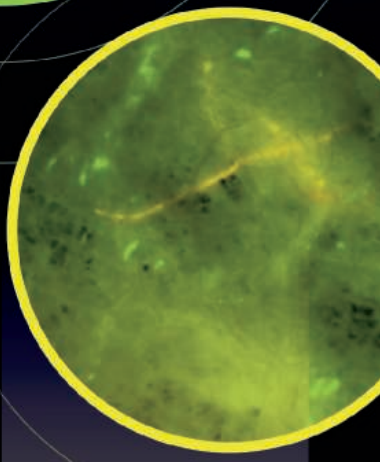
INDUSTRY
Crack feature, fluorescent penetrant





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.

POSITIONS OPEN

Duke University School of Medicine

Postdoctoral research positions are now available to study the impact of manipulating the estrogen signaling axis on immune function and how hormonal status impacts breast tumor immunity. This translational project will provide incumbents with the opportunity to train with a multidisciplinary team of scientists in a new and exciting area of cancer therapeutics. Successful applicants will have training in immunology, hormone action, cancer biology or genetics. Motivated individuals with a recent Ph.D., M.D. or M.D./Ph.D. should send a *curriculum vitae*, a brief description of research interests and the names of three references to: **Dr. Donald McDonnell, Department of Pharmacology and Cancer Biology, Duke University School of Medicine, Durham NC 27710. E-mail: Trena.martelon@duke.edu**

北京大学
120th ANNIVERSARY OF PEKING UNIVERSITY
1898–2018

Innovation and Services — The Development of Information Science at Peking University

The discipline of Information Science at Peking University (PKU) covers two majors: Electronics Engineering (EE) and Computer Science (CS). The EE major, founded in the 1950s, covers microelectronics and solid-state electronics, physical electronics, circuits and systems, electromagnetic fields and microwaves, quantum electronics, communication and information systems, etc. The CS major, founded in 1978, focuses on computer architecture, software, artificial intelligence, computational linguistics and digital media, etc.

While developing state-of-the-art mechanisms for cultivating top talents in information science, we are also dedicated to applying cutting-edge technologies to industrial and social development. Our current research work focuses on hardware, software and system technology for big data, cloud computing, artificial intelligence, internet of things, new generation communication, intelligent integrated microsystem and quantum computing, etc. We also encourage cross-disciplinary research, and have established multiple research centers to facilitate the integration and development of related disciplines, e.g., the Research Center for Medical-Information Science, the Research Center for Computational Social Science, and the Research Center for Brain-inspired Computing.

In the field of Microelectronics and Integrated Circuit (IC), as far back as 1975, the researchers in the Institute of Microelectronics, led by Prof. Yangyuan Wang, successfully developed the first 1024-bit n-channel silicon-gate MOS Dynamic Random Access Memory (DRAM) in China. More recently, Prof. Ru Huang and her team have been focusing on the low-power semiconductor devices with new structures/materials/mechanisms and the key technologies of brain-inspired computing chips, aiming at pushing IC technology to its power limit and expanding its functionality in the Post-Moore Era. In particular, they have proposed and fabricated novel nanoscale multi-gate transistors and steep-slope ultralow-power transistors with record figures-of-merit (FoMs), as well as new neuromorphic devices and array architectures which can greatly reduce the energy consumption of artificial intelligence (AI) chips. In the past 11 years, they have published 45 papers at the IEEE International Electron Devices Meeting (IEDM), the top-tier conference in the field of microelectronic devices. The relevant research results have been transferred to or applied in leading IC companies, such as SMIC, Huawei HiSilicon, Cadence, and Synopsys, etc.

In the field of software, Peking University has always been the pioneer and leader of fundamental research as well as technology transfer in China. The team led by Prof. Fuqing Yang developed several operating systems for computers made in China before 1980s, and a series of Integrated Development Environments (IDEs) for structured, object-oriented and component-based software in 1990s. The team led by Prof. Hong Mei proposed a new software paradigm, Internetware, to describe the new and even unconventional properties of software in the Internet Age. Internetware became the flagship of software research in China in the 2010s, revealing and addressing many interesting and valuable issues. For example, Prof. Mei's team invented an automated approach to enabling the efficient interoperability between information systems, sharing the data and functionalities of an information monolith to others without the source code, data schema or back-end permission, increasing the efficiency of interoperability by hundreds of times. This approach and its operating platform have been transferred to Digital China, Lenovo, AliCloud, and iFLYTECH, etc. with the biggest information technology patent royalty (15 million US dollars) in Chinese universities. This new interoperability technology has

been widely used in the development of software for data sharing, big data application, smart cities and system integration in China. With the vision of Internetware, the team will focus on intelligent software development, big data and social-cyber-physical computing in the next decade.

In the field of Chinese information processing, Prof. Wang Xuan led the revolution of Chinese character storage, display and laser phototypesetting technologies as early as the 1970s. Since the 1990s, Peking University has established the unique Chinese language model and knowledge bases. These achievements enable Chinese people to continue using and developing Chinese languages and scripts in the information era.

In the field of intelligent science, Prof. Qingyun Shi developed the highest performance fingerprinting system in 1980s. In the last 20 years, Prof. Wen Gao steered the development of national video coding standards in China, the IEEE standard and the ISO / IEC standard on visual feature coding, which have been widely deployed in High-definition Television broadcasting in China and other countries. As chairman of Artificial Intelligence Industry Technology Innovation Strategic Alliance (AITISA) approved by the Ministry of Science and Technology of China (MOST), Prof. Gao emphasized the leading role of PKU in establishing an open source AI platform with focus on the new intelligent computing architecture and brain-like computers, so as to promote the transformation and upgrading of human society from the information era to the intelligence era.

PKU Information Science will continue promoting the multi-disciplinary interaction and integration within the university, while contributing to information science research and industry in China and around the world at multiple levels, including theory, technology, system and industrial services. Prof. Ru Huang, Dean of EECS School, warmly welcomes top researchers from all over the world to join or participate in our promising platform at Peking University. Feel free to contact us: dean.eecs@pku.edu.cn or visit http://eecs.pku.edu.cn/EN/.

Micro/nanofabrication facility for researches on Semiconductor device, M/NEMS and IC/MEMS integration with feature size down to sub 10 nm.

The Institute of Computer Science and Technology at Peking University

The Institute of Computer Science and Technology (ICST) was established in 1983 by professor WANG Xuan, the winner of the Top National Science and Technology Award in 2001.

ICST has made significant contributions to computer science and technologies by advancing applications in printing and media industries. Specifically, the Laser Typesetting System, adopted by more than eighty percent of Chinese newspaper publishers, has led a nationwide revolution; the Newspaper Digital Asset Management System has created a paperless working environment for the media industry; the Digital Rights Management System has promoted the popularity of e-books; and the Digital Broadcast Control System has made digital TV automation possible.

Expanding the scope from digital to intelligent media, ICST is dedicated to the research of media intelligence. The important progress achieved recently includes cross-modal content recognition, machine writing, and automatic generation of Chinese fonts.

Prof. Xiaojun Wan specializes in natural language generation and machine writing. His team invented a series of text summarization methods to automatically generate news reports. Their innovative technologies have been transferred into several online AI reporters (e.g., Xiaomingbot and XiaoNan) in various domains, including sports and lifestyles. For instance, Xiaomingbot published more than ten thousand news articles for Rio 2016, Football League and NBA events.

Prof. Yuxin Peng, has led his team to achieve several important breakthroughs on cross-modal content analysis and recognition. They proposed spatial topology based attention learning, which for the first time solved the difficult problem of fine-grained image classification without object annotation. They also proposed collaborative learning of spatial-temporal attention to achieve the video recognition accuracy of 95.7%, and won first place six times in video instance search of International evaluation campaign TRECVID. The graph regularized shared semantic space projection they proposed, for the first time broke through cross-modal retrieval up to 5 modal types including image, video, audio, text and 3D model. Their research achievements have been successfully applied to over 100 organizations, and gained first prize of Beijing Science and Technology Award in 2016.

The group headed by Professor Zhouhui Lian, and previously headed by Professor Jianguo Xiao, has been working for decades to advance the design and production of Chinese fonts. Adopting state-of-the-art CG/AI/CV techniques, they have proposed a series of solutions significantly improving the designing/producing efficiency of high-quality printing/compressed/colorful Chinese fonts. The group was also the first to make possible the automatic generation of practical handwriting fonts with arbitrarily large numbers of Chinese characters. Their techniques have been authorized and transferred into commercial products for hundreds of millions of users, producing great commercial benefits and significant social impact.

The Institute welcomes applications from top researchers and, in particular, young talents from around the world. Feel free to contact us:

Website: http://www.icst.pku.edu.cn/

Email: icst748@pku.edu.cn

Address: No. 128 Zhongguancun North Street, Haidian District, Beijing, 100871, P. R. China

Intelligent Control of Networked Dynamical Systems at Peking University

With the dramatic development of multiple robot systems, intelligent traffic systems, Internet of Things (IoT), smart grids, and mobile sensor networks in the last decade, classic control theory is no longer adequate. A new design theory and control methods are needed to deal with such distributed systems, each of which is typically composed of many subsystems connected as a large network.

Long Wang, a Cheung Kong Chair Professor of Dynamics and Control, has spearheaded a series of pioneering programs by using Evolutionary Game Theory (EGT), which is powerful in solving the competing interactions among different agents. Some fundamental studies include when the popular embedded Markov chain method is valid, the intrinsic differences between the two widely used evolutionary rules, and what is the asymptotic behavior of mobile agents. Professor Wang and his colleagues have found a simple yet effective control protocol to make all agents coordinate with each other, which has proven to be efficient and energy-saving over a wide range of network structures. In addition, they have also been successful in studying the emergent mechanism of human intelligent behaviors, which is fundamental to artificial intelligence. These novel and insightful theoretical results have successfully found their applications in the coordination and control of multiple mobile robots.

Feel free to contact us: Center for Systems and Control, Peking University.

Website: http://www.mech.pku.edu.cn/robot/index.htm

Email: longwang@pku.edu.cn

北京大学定量生物学中心
CENTER FOR QUANTITATIVE BIOLOGY

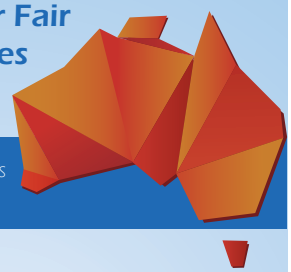
Faculty Positions in Biological Physics and Systems Biology

The Center for Quantitative Biology (CQB) at **Peking University (PKU)** invites applications for faculty positions at all ranks. We seek for creative individuals in all areas of biological physics and quantitative systems biology, broadly defined.

CQB (http://cqb.pku.edu.cn/) is dedicated to research and education at the interface between the physical and biological sciences. Current research areas include physical biology, mathematical biology, systems biology, synthetic biology and computational biology.

Applications are welcome all year round. For qualified applicants, PKU can sponsor their applications to the national (Young) Thousand Talents Program. Application materials (cover letter, CV, summary of research achievements and future research plan, all in a single PDF file) and three letters of reference should be sent to **Ms. Wei Xiao (gsmkyb@pku.edu.cn)**.

2018 Overseas Career Fair for Chinese Universities in Australia



Thousands of academic job vacancies are in fast-developing China.

Job Fairs in Australia		
May 25th	16:10-19:10 (Local Time)	The University of Adelaide
May 26th	13:30-16:30 (Local Time)	The University of Melbourne
May 27th	13:30-16:30 (Local Time)	Monash University
Global Virtual Job Fair		
May 27th	08:00-20:00 (Beijing Time)	http://www.edu.cn/cv

Participation Approach
Please submit your resume to zhaojia@eol.cn

Job Vacancies
in China's Universities and Institutes
Please visit http://www.acabridge.cn/

Pre-regisrtation Now Open

Tianjin Normal University International Elites' Forum

A sincere invitation to talents from all over the world to join us

An introduction to the forum

In accordance with the national strategy of Beijing-Tianjin-Hebei region's coordinated development, and following the principle of developing "An Insight for Discovering Talents, Sincerity for Cherishing Talents, a High-degree of Tolerance for Talents, Courage for Utilizing Talents, and Effective Measures for Assembling Talents", Tianjin Normal University extends a sincere invitation to international elites to join us. The university will hold the Second International Elites' Forum from 26th to 29th May, 2018, as part of our talent-recruiting programme, in an effort to gather the world's leading scholars and holders of national and provincial level talent titles. The forum will provide them with a platform where they can have in-depth consultation on the latest developments in their disciplinary fields through symposia, seminars and conferences; and learn more about the university and the city of Tianjin, both of which are endeavouring to attract innovative and enterprising young talents.

We sincerely welcome you to join us in our work in scientific research, teaching and training, public service, cultural innovation, and international communication and cooperation, so that together we will contribute to the economic and social development of our shared future.

Disciplinary fields

1. Psychology and related areas:
We are looking for experts in psychology and related areas, medical image processing, cognitive neuroscience, bioinformatics (or human genomics) and experimental psychology.

2. Political Sciences
We are looking for experts in political theory, comparative politics, international politics, the history and construction of the CPC, governance and public policy.

3. World History
We are looking for experts in the economic and social history of Europe, French history, German history, the late medieval and early modern history of Europe, and the history of European civilisation.

4. Marxist Theory
We are looking for experts in areas related to Marxist theory, politics and philosophy.

5. Education
We are looking for experts in the fundamental principles of education, curriculum and teaching theory, preschool education, and comparative education.

6. Chinese Language and Literature
We are looking for experts in comparative and world literature, linguistics and applied linguistics (including experimental phonetics).

7. Chemistry and Materials Science
We are looking for experts in inorganic chemistry, analytical chemistry, organic chemistry, physical chemistry and chemical biology.

8. Physics and Materials Science
We are looking for experts in physics and related areas, materials science and engineering (including energy-related studies), and astrophysics.

9. Geography
We are looking for experts in geography and related areas, earth sciences and environmental sciences.

10. Biology
We are looking for experts in cell biology, biochemistry and molecular biology, genetics, zoology, hydrobiology and ecology.

Application requirements

Applicants should be:
1) PhD degree holders;
2) Below the age of 45 (or the age of 40 for science disciplines);
3) With leading or outstanding potential, possessed with or have the qualifications close to national or provincial level talent title, such as National Thousand Talent Plan, National Thousand Talent Plan for Young Outstanding Scientists, Chang Jiang Scholars Program, Chang Jiang Scholars Program for Young Outstanding Scientists, Outstanding Young Scholars, and Excellent Young Scholars. See the Tianjin Normal University Double One-Hundred Talents advertising window for more information. (Website: http://www.tjnu.edu.cn/)
4) Holders of a teaching or research position at prestigious overseas universities, research institutions, or renowned enterprises or PhD/ Post-docs with more than 2 years of overseas research experience.

Forum programme

26th May Registration and travel reimbursement
27th May Morning: opening ceremony
Afternoon: seminars, observation in laboratories, campus tour
28th May symposium, applicants' question and answer session, tour of living and working environments
29th May closing session

Application procedure

CVs should be sent to gjgdlt@tjnu.edu.cn (with the email subject heading: TNU 2nd International Elites' Forum + Name + Current Organization + Discipline) by the 10th May 2018. Successful applicants will receive an invitation from the university.

Contact person

Ms Ge Yanxia
Tel: 17302237889

Mr. Yang Bin
Tel: 15822871057

"Ask Big and Important Questions"

Alexey Amunts chose to start his lab at SciLifeLab because of the possibilities he saw in the research environment.

Alexey Amunts uses cryo-EM to visualize macromolecules at atomic resolution. His group explores macromolecules that play key roles in fundamental and medically relevant cellular processes, such as the mechanism of protein synthesis in mitochondria.

"The cryo-EM method has made it possible to reveal complex cellular systems that we lacked the tools to investigate before, and there is a whole new nanoscale universe waiting to be explored, so it is a very exciting time to be in this field."

He started out by doing his PhD at Tel Aviv University in Israel, where he used X-ray crystallography to visualise plant photosynthetic complexes. After that he moved on to a postdoc position at the MRC Laboratory of Molecular Biology in Cambridge, UK, to investigate ribosomes using cryo-EM.

When Alexey Amunts got the opportunity to set up a cryo-EM lab at SciLifeLab, he took it. One of the reasons was the many strong advantages that he sees with the Swedish research environment.

"The grant schemes in Sweden have a tradition of supporting ambitious initiatives, those where you don't expect the results to start appearing within 2-3 years. This is particularly important, because it allows to navigate a research group through difficult periods. This consideration is crucial when one is looking for a place to set up a lab because finding a productive direction takes time."

His lab is located close to the SciLifeLab cryo-EM, mass spectrometry, and drug discovery facilities.

Photo: Neil Grant

Alexey Amunts

"There were three main things that made SciLifeLab so attractive to me: the infrastructure support, the network of collaborations, and most importantly the wide spectrum of young talent."

When Alexey Amunts reflects on the Fellows Program he highlights it's snowball effects.

"SciLifeLab recruits fellows from leading institutions abroad and gives them an unprecedented degree of freedom as well as tools to create. But what I think is even more important is that by establishing new technologies the Fellows attract bright postdocs and PhD students that otherwise would not consider coming here. This makes not only a major contribution to the current research environment, but will also have a long-term effect on the Swedish science and society."

His advice to researchers who are about to become independent investigators is to ask big and important questions that will take them into uncharted territories.

"Even if pursuing those questions might be risky, you will get an opportunity to expose new worlds with even more interesting questions, and eventually have better chances to make a unique contribution to science."

SciLifeLab – a national resource

SciLifeLab is a Swedish research center within molecular biosciences with focus on health and environment. It is also a national resource with the mission to develop, use and provide advanced technologies. The center infrastructure encompasses a multitude of biomolecular technologies and bioinformatics services.

SciLifeLab

www.scilifelab.se

ASSOCIATE PROFESSOR/PROFESSOR AND VICE CHAIR, DEPARTMENT OF BASIC MEDICAL SCIENCES

The Western University of Health Sciences (WesternU), College of Osteopathic Medicine of the Pacific (COMP), invites applicants for the position of Associate Professor/Professor and Vice Chair of the Department of Basic Medical Sciences (BMS) to lead the development and operation of the Northwest campus in Lebanon, Oregon. The Vice Chair is a vital member of the College's leadership team. The candidate will be expected to provide visionary leadership to the department in its research and educational missions. Principal responsibilities will be: 1) mentorship of faculty, 2) participate in the development and delivery of pre-clinical medical education, 3) development of departmental research programs and integration of research and education, and 4) strategic planning to advance the Department in line with the mission of the Department, College and University. Successful candidates will have an internationally-recognized program of research, significant history of extramural funding, administrative and educational experience, and strong leadership skills. Salary and rank will be commensurate with qualifications and experience. The successful candidate will have expertise in a field of biomedical sciences. Interested individuals should submit a curriculum vitae, cover letter, a two page outline of their scholarly work and teaching philosophy and apply to posting number A00259 on **http://apptrkr.com/1199686** Inquiries can be addressed to the Chair of Search Committee: Dr. Michelle Steinauer, Tel #: 541-259-0233; e-mail: msteinauer@westernu.edu

Western University of Health Sciences is an Equal Opportunity/ Affirmative Action Employer and is committed to a policy of equal employment to all applicants.

Cleveland Clinic
Lerner Research Institute

Postdoctoral Fellowships Available

The Lerner Research Institute is home to all basic, translational and clinical research at Cleveland Clinic, the No. 2 ranked U.S. hospital by *U.S. News and World Report*. With over $140M in federal grants and an annual research budget of $260M, the Lerner is consistently ranked among the top research institutes in the nation. Postdoctoral fellows routinely obtain grant funding and first-author publications in top-tier journals.

Postdoctoral Job Opportunities:
http://www.lerner.ccf.org/jobs/postdoctoral/
For further information email: RETC@ccf.org

The Lerner Experience

- Opportunity to train among world-class scientists and physician-scientists in a top-ranked healthcare institution
- Multidisciplinary, disease-focused research programs
- 175 principal investigators in 12 departments with over 700,000 square feet of research space
- Competitive salary and benefits package
- Active Postdoctoral Association and 250+ postdocs
- Career development opportunities and support through the Research Education and Training Center

City of Cleveland

Cleveland is a multicultural city with nationally acclaimed museums, sports, restaurants, and music and arts programs. Situated on Lake Erie, the area offers stunning views, beaches, and water sports. Low cost of living, with below average traffic and commute times for major cities.

Smithsonian
National Museum of Natural History

DEPUTY DIRECTOR

The National Museum of Natural History (NMNH), a science-based museum within the Smithsonian Institution's complex of museums and research organizations, seeks a dynamic candidate to lead the management of its research departments, collections programs, museum outreach activities on exhibits, education and communications, specialized laboratories, field stations, and other similar programs and partnerships. The successful candidate will bring extensive experience in natural history museum leadership, policy formulation and strategic planning to the NMNH that has 80 scientists in the fields of biology, botany, entomology, zoology, geology, paleobiology, anthropology and archeology an d over 450 total employees. Working in collaboration with the Sant Director, this high-profile leader will formulate and implement strategic directions for the Science programs of the Museum; assure the quality and breadth of all Museum activities; and provide institutional leadership in broader national and international programs.

The National Museum of Natural History is dedicated to understanding and explaining the natural world and is the most visited natural history museum in the world. With over 145 million specimens and objects, the Museum's collections represent over 90% of the holdings of the Smithsonian. These collections, and the work of the scientific staff, form the foundation for the Museum's extensive public programs, and include exhibitions, education, and outreach programs.

This is a full-time, permanent position to be filled as Trust (private sector, U.S. citizenship not required, proof of eligibility to work in the U.S. required). Salary commensurate with experience. The Smithsonian Institution offers a comprehensive package of benefits. For complete requirements and application procedures, please visit: **www.sihr.si.edu** and refer to Announcements **EX-18-07**. Applications must be received online by **May 8, 2018** and must reference the announcement number. Applicants will be notified by email when their applications are received.

We encourage all qualified candidates to apply.

The Smithsonian Institution is an Equal Opportunity Employer.

By **Arnav Chhabra**

Academia's forgotten footnote

In my third year of grad school, everything seemed to fall apart. I was dealing with my grandmother's death, and then my girlfriend and I broke up. I spent the following year in a painful feedback loop of depression and despair. Every day, I would trudge into lab and try to get excited about my projects. But when I encountered minor hurdles such as a failed replication or contaminated samples, I would become discouraged and give up. Even when my experiments went smoothly, I felt guilty about the time I had wasted being unproductive. I knew I was struggling, but I didn't ask for help. I thought I could deal with my state of mind just as I had dealt with every other problem in my life: Bottle up my emotions, attack the problem with logic, and iterate until I arrived at a solution.

This time, however, that approach didn't work. I wasn't sleeping well. I couldn't focus enough to read even a brief paper in one sitting. During a lab presentation, I got so dizzy from exhaustion that I had to stop midway through. Balancing work and my mental health had become untenable. I concluded that trying to persist with my lab work was not fair to myself or to my group. When I gathered the courage to discuss time off with my adviser, she encouraged me to take a hiatus.

In the beginning, it felt like a blessing to have time to focus on personal projects and self-care I had neglected, including a healthy dose of binge-watching basketball. But after 3 months, feelings of academic guilt crept back. Despite not feeling fully up to it, I convinced myself that it was time to return.

"I spent the following year in a painful feedback loop of depression and despair."

I tried to mask my depression, but the impact on my work was apparent again. A few months after my return, my adviser called me into her office and initiated a forthright discussion about why I was not being productive. I wondered whether it was time for me to leave my Ph.D. program. I felt like a failure, and I thought of suicide for the first time. These thoughts finally pushed me to see a therapist.

At first, therapy felt awkward and ineffectual. I was not accustomed to introspection, and I certainly was not accustomed to being probed about my problems. But over time, my counselor and I developed a rapport. She helped me figure out that while my girlfriend and I were together and my workload ramped up, I had socially isolated myself, which meant that I didn't have a support network to draw on when things got tough. She also encouraged me to connect with MindHandHeart (MHH), a coalition of students, faculty, and staff promoting mental health and well-being at the Massachusetts Institute of Technology.

A year and a half of counseling and MHH gave me the tools I needed to strengthen my relationships and establish a support structure to help cope with any future episodes. I revitalized friendships and opened up about my feelings. My lab work improved, and I am on track to graduate in the next year.

Studies have shown that 40% of Ph.D. students are depressed. But if it weren't for my own experiences, I would not be aware of this—and therein lies the problem. Academics tend to be averse to discussing mental health openly, and higher education's mental health safety net is patchy—a forgotten footnote that all too often fails its students. I am lucky that I have a supportive adviser and health insurance that covers mental health care. And I am lucky that I made it to therapy when I considered suicide. I could easily have waited too long.

Even if you aren't personally burdened by these issues, everyone must take action to support those who are struggling, and to ensure that institutions have support frameworks: advisers trained to deal with mental health issues; student-run support programs; and community events with a focus on diversity and inclusion, because students from underrepresented groups are more likely to experience depression. If your institution doesn't, what can you do to help before someone you care about becomes a different kind of footnote? ■

Arnav Chhabra is a grad student at Harvard-MIT Health Sciences and Technology in Cambridge, Massachusetts. Send your story to SciCareerEditor@aaas.org.