Can trauma leave molecular scars on later generations? *p. 212*

Pericytes constrict brain vessels in Alzheimer's *pp. 223 & 250*

A short stay for deep magma *p. 260*

# Science

$15
19 JULY 2019
sciencemag.org

AAAS

SPECIAL ISSUE

## APOLLO LANDING

50TH ANNIVERSARY

# CONTENTS

19 JULY 2019 • VOLUME 365 • ISSUE 6450

SPECIAL SECTION

# MOON LANDING AT 50



**ON THE COVER**




Neil Armstrong's shadow cast on the surface of the Moon. This image was taken by Armstrong on 21 July 1969, as part of a panorama of the Apollo 11 landing site, while Buzz Aldrin deployed scientific instruments. The Sun was directly behind Armstrong, producing the appearance of a bright halo around his helmet. Black crosses were introduced by the camera optics for distortion calibration. See page 232. *Credit: NASA*

SEE ALSO ▸ EDITORIAL P. 203; BOOKS *ET AL.* P. 226



## NEWS



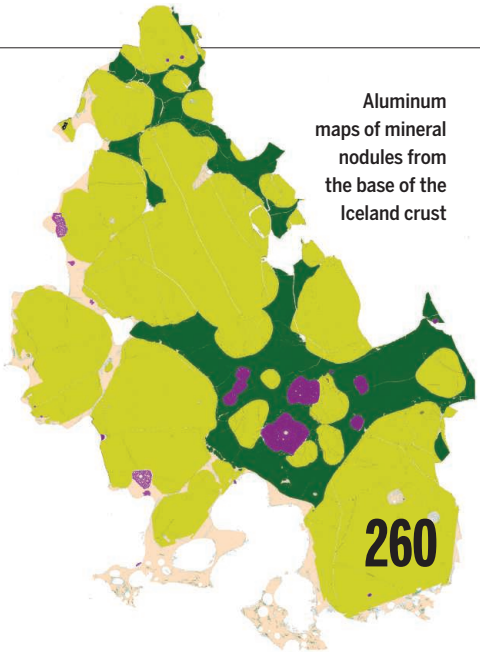
## INSIGHTS



PHOTO: NASA



Published by AAAS




Aluminum maps of mineral nodules from the base of the Iceland crust



# RESEARCH




**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): $1971; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $98. 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.





CREDIT: EUN J.F. MUTCH

# Science AAAS

**Editor-in-Chief** Jeremy Berg, jberg@aaas.org

**Executive Editor** Monica M. Bradford

**EDITOR, INSIGHTS** Lisa D. Chong **EDITORS, RESEARCH** Valda Vinson, Jake S. Yeston **DEPUTY EDITORS** Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Paula A. Kiberstis, Marc S. Lavine (Canada), Steve Mao, Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Keith T. Smith (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Hilary Stewart (UK), Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

**News Editor** Tim Appenzeller

**NEWS MANAGING EDITOR** John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad (Cambridge, UK), Paul Voosen, Meredith Wadman **INTERN** Alex Fox **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Shanghai), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein (Editor), Katie Langin **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

**Creative Director** Beth Rakouskas

**DESIGN MANAGING EDITOR** Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **PHOTO EDITOR** Emily Petersen **GRAPHICS EDITOR** Nirja Desai **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary

**CEO and Executive Publisher** Rush D. Holt

**Publisher, Science Family of Journals** Bill Moran

**DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS** Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy **CONTENT SPECIALIST** Jacob Hedrick **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **DIGITAL AD-OPS SPECIALIST** Patrick Gerrits **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

**ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **SENIOR DESIGNER** Kim Huynh **TRADE SHOW AND MEETINGS ASSOCIATE** Andrew Clamp

**DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING** Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

**DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **ASSOCIATE DIRECTOR, PRODUCT DEVELOPMENT AND SPJ** Hannah Heckner **SR. PRODUCT ASSOCIATE** Robert Koepke **DIGITAL PRODUCT STRATEGIST** Michael Hardesty **SPJ ASSOCIATE** Samantha Bruno Fuller

**DIRECTOR, INSTITUTIONAL LICENSING** Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGERS** Marco Castellan, Christos Skoutas **MANAGER, SYSTEMS AND OPERATIONS** Brian Holiahn **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

**DIRECTOR, GLOBAL SALES** Tracy Holmes **ASSOCIATE DIRECTOR, ADVERTISING SALES, US** Laurie Faraday **US EAST COAST AND MID WEST SALES** Glen Cox **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **PROJECT MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

**DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS** Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

---

**MAIN HEADQUARTERS**
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

**SCIENCE INTERNATIONAL**
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

**SCIENCE CHINA**
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

**SCIENCE JAPAN**
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

**EDITORIAL**
science_editors@aaas.org

**NEWS**
science_news@aaas.org

**INFORMATION FOR AUTHORS**
sciencemag.org/authors/science-information-authors

**REPRINTS AND PERMISSIONS**
sciencemag.org/help/reprints-and-permissions

**MEDIA CONTACTS**
scipak@aaas.org

**MULTIMEDIA CONTACTS**
SciencePodcast@aaas.org
ScienceVideo@aaas.org

**INSTITUTIONAL SALES AND SITE LICENSES**
sciencemag.org/librarian

**PRODUCT ADVERTISING & CUSTOM PUBLISHING**
advertising.sciencemag.org/products-services
science_advertising@aaas.org

**CLASSIFIED ADVERTISING**
advertising.sciencemag.org/science-careers
advertise@sciencecareers.org

**JOB POSTING CUSTOMER SERVICE**
employers.sciencecareers.org
support@sciencecareers.org

**MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS**
sciencemag.org/subscriptions

**MEMBER BENEFITS**
aaas.org/membercentral

**AAAS BOARD OF DIRECTORS**
**CHAIR** Margaret A. Hamburg
**PRESIDENT** Steven Chu
**PRESIDENT-ELECT** Claire M. Fraser
**TREASURER** Carolyn N. Ainslie
**CHIEF EXECUTIVE OFFICER**
Rush D. Holt
**BOARD** Cynthia M. Beall
May R. Berenbaum
Rosina M. Bierbaum
Ann Bostrom
Stephen P. A. Fodor
S. James Gates, Jr.
Laura H. Greene
Kaye Husbands Fealing
Maria Klawe
Robert B. Millard
William D. Provine









*Published by AAAS*

EDITORIAL

# A child of Apollo

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

I was 11 years old when my neighborhood gathered to watch Apollo 11 transport human beings to the surface of the Moon for the first time. We clustered around a black-and-white television when, at dusk in California, Neil Armstrong took his first lunar steps and said, "That's one small step for [a] man, one giant leap for mankind." This issue of *Science* celebrates the 50th anniversary of this landmark event. The path that led to this success transformed our understanding of humankind's place in the universe and demonstrated the power of an audacious goal in driving technology forward.

After the launch of Sputnik in 1957, the United States focused on the need for engineers and scientists to keep up technologically with the Soviet Union. In 1961, President John F. Kennedy proposed that the United States "should commit itself to achieving the goal, before this decade is out, of landing a man on the Moon and returning him safely to the Earth." The ambitious National Aeronautics and Space Administration (NASA) project had three phases—Mercury, Gemini, and Apollo. I first became aware of the space program with the last of the Mercury flights and was fully engaged by the time of the first crewed Gemini launch in March 1965. I remember watching the launches, splashdowns, and the spacewalk of Ed White tethered to Gemini 4 in June 1965. Gemini model kits soon joined my list of desired presents. As a child, the progress of the program seemed exciting but inevitable, particularly given the confident attitudes portrayed by the astronauts and NASA staff. I recall being shocked by the horrible fire in a sealed command module that killed three astronauts as they prepared for the first Apollo mission in January 1967. NASA spent a year extensively redesigning the command module, focusing on safety, possible modes of failure, and reliability. In December 1968, Apollo 8 orbited the Moon for the first time. This vantage point offered a new perspective of Earth, captured by an iconic color photograph of our home planet rising over the lunar surface. Such pictures depicted a unified, blue, cloud-decorated planet without national boundaries. Seven months later, after two additional flights, Apollo 11 headed to the Moon with the goal of placing two astronauts on the lunar surface. After a dramatic journey to find a suitable landing site, the lunar lander settled into the lunar soil and, less than 7 hours later, Neil Armstrong took his famous first step.

The bootprint on the lunar surface was made by astronaut Edwin "Buzz" Aldrin of the Apollo 11 mission.

The technology that enabled the Apollo missions was incredibly impressive at the time but is remarkably primitive by today's standards. The guidance computer was less powerful than a cheap, modern hand calculator. The successes of the Moon missions were primarily triumphs of rocket engineering (rather than rocket science). Nonetheless, the Moon missions enabled many new scientific directions, beginning with studies of the samples from the lunar surface. More experiments were performed on subsequent missions, facilitated by roving vehicles, setting the stage for unmanned rovers that are still exploring the surface of Mars today. Increasingly sophisticated scientific instruments have been built for probing the structures, atmospheres, and surfaces of other planets and various objects in our Solar System. Some people are now actively discussing possible missions involving human passengers to the Moon and then on to Mars. Such endeavors have the potential to capture the public's imagination but, as these are contemplated, it is important to articulate clearly the benefits (and challenges) of including oxygen-breathing, waste-producing, radiation-sensitive human beings along with robotic and computer systems that are now so powerful. If a major goal is to develop technologies for future human space travel, including astronauts may make sense. However, if the goal is scientific exploration, robotic extensions of humans offer many advantages; we can continue to make giant leaps for mankind with all the human steps occurring on Earth.

**–Jeremy Berg**

10.1126/science.aay6770

PHOTOS: (INSET) NASA; (TOP RIGHT) TERRY CLARK

"It was a perfect voyage."

**Koji Hara,** in Kyodo News, after his team paddled last week from Taiwan to Okinawa, Japan, in a dugout canoe, showing how early humans might have migrated there.

## IN BRIEF

*Edited by* **Jeffrey Brainard**

INFECTIOUS DISEASE

# Congo turns down new Ebola vaccines

The Democratic Republic of the Congo (DRC) said it will not allow new experimental vaccines to be deployed to combat an Ebola outbreak that has killed more than 1650 people since August 2018, the DRC's health minister announced on 11 July. An experimental Merck vaccine has already been given to 161,000 people in the DRC and was shown to be effective in those exposed to the virus during a West African epidemic that ended in 2016. Health minister Oly Ilunga Kalenga said he decided not to allow additional, experimental vaccines because their use might confuse the public. He also noted that other experimental vaccines—a Johnson & Johnson candidate has been pushed by outside experts—don't have proven efficacy in people exposed to Ebola virus. Outside experts worry that stocks of the Merck vaccine may run out before the current outbreak ends. On 14 July, health workers diagnosed the first Ebola case in the large city of Goma, near the Rwandan border; the patient died this week, the DRC Ministry of Health confirmed.

## NASA changes flight leaders

SPACE | In a bid to jump-start returning astronauts to the moon by 2024, NASA Administrator Jim Bridenstine in Washington, D.C., reassigned two top human spaceflight officials last week, surprising many in the space industry. Bridenstine removed William Gerstenmaier, who had led NASA's human exploration and operations directorate since its creation in 2011, and Bill Hill, deputy associate administrator for exploration systems development, and placed them in advisory roles. New leadership is required to meet the 2024 goal, Bridenstine said; a major challenge is reining in the cost of the delayed, powerful Space Launch System rocket. "The administration is interested in going fast, [and] we're interested in doing things in a different way," he said. NASA has estimated that it could take $20 billion to $30 billion in new funding to meet the 2024 goal; Congress has yet to approve any of this additional money.

## Mice eyed as cat substitutes

BIOMEDICINE | Researchers have shown that laboratory mice can be used as a substitute for cats in studies of the parasite *Toxoplasma gondii*, a leading cause of blindness and birth defects in people. A team led by Laura Knoll at the University of Wisconsin in Madison succeeded at getting *T. gondii* to sexually reproduce in mice. Until now, cats have been the only animals in which the parasite completes the sexual stages of its life cycle. The world's leading research lab for studies of *T. gondii*, part of the U.S. Department of Agriculture, was closed in April after animal activists protested because it used the cats for parasite production and then euthanized them (*Science*, 12 April, p. 109). The Wisconsin group reported its success, which depended on boosting the mice's blood levels of a fatty acid that is high in cats but low in other mammals, on 1 July on the bioRxiv preprint server. Editors at *PLOS Biology* have since accepted it for publication.

## Hawaii telescope moves ahead

ASTRONOMY | Hundreds of protesters confronted workers this week as managers moved to resume construction of the $1 billion Thirty Meter Telescope on Mauna Kea in Hawaii. Work on what would be one of the world's largest telescopes stopped in 2015, when opponents halted construction by blockading an access road; 4 years of litigation followed. Critics say the 18-story instrument would desecrate a site of cultural and religious significance. Hawaii's state Supreme Court ruled in October 2018 to allow construction, and in June, the state gave a final green light. This week, authorities worked to close access roads to the mountain. "This will bar Native Hawaiians from being able to access or enter their place or house of worship," says opponent Kealoha Pisciotta of Mauna Kea Anaina Hou. On 8 July, opponents filed a last-ditch legal petition arguing that the project has not provided a sufficiently large security bond to guarantee the cost of construction if it does not receive full funding.

## Probe samples asteroid interior

PLANETARY SCIENCE | Japan's Hayabusa2 spacecraft has landed on asteroid Ryugu and, in a space exploration first, probably captured material from the interior. Earlier this year, the mission blasted a crater in the asteroid that left subsurface rock scattered around the crater's rim. Last week, Hayabusa2 briefly touched down and fired a bulletlike projectile at that material to knock samples into a collection horn. Launched in 2014, Hayabusa2 retrieved surface samples in February. Having both surface and interior samples should give clues to space weathering and the evolution of the solar system, says project scientist Seiichiro Watanabe of Nagoya University in Japan. The samples are to be brought to Earth in December 2020.

## Harvard suspends economist

#METOO | Harvard University on 10 July placed Roland Fryer, a prominent economics professor, on administrative leave without pay for 2 years after a university investigation found he had engaged in unwelcome

PHOTO: TUI DE ROY/MINDEN PICTURES

CONSERVATION

## Kenya vows to save mountain bongo

The Kenyan government has launched a plan to save the mountain bongo (*Tragelaphus eurycerus ssp. isaaci*), one of the most endangered mammals in the world, from extinction. Only 100 individuals of the brown-and-white antelope subspecies, also known as the eastern bongo, survive in the wild, most of them in Kenya's Aberdare mountain range. But the Mount Kenya Wildlife Conservancy in Nanyuki has bred a captive population (above) of 77 animals, from 14 females and four males donated by the White Oak Conservation Center in Yulee, Florida, in 2004. On 8 July, Najib Balala, Kenya's minister of tourism and wildlife, pledged to set aside a 314-hectare sanctuary for the animals in Mount Kenya Forest that the Kenya Wildlife Service will protect against poachers and predators. The conservancy could start to release the first bongos there this year. The population is projected to grow to 730 animals over the next 50 years, Balala said.

sexual conduct. Several employees accused him of making sexual comments in a professional setting and fostering a hostile work environment. Fryer will be allowed to return to his post after 2 years with restrictions. He rose to prominence from a low-income background and received tenure at age 30. He has been recognized for his research on closing the racial gap in educational achievement and other social issues. Fryer's suspension comes as economists have been reckoning with many instances in their field of harassment and discrimination against women and minorities.

## EPA appointments questioned

POLICY | The U.S. Environmental Protection Agency (EPA) repeatedly sidestepped regular procedures for appointing members of two influential science advisory committees, says a report released this week by congressional auditors. Appointments to the Science Advisory Board and Clean Air Scientific Advisory Committee were drawn from the full roster of nominees submitted to the agency instead of a short list prepared by career agency staff, the Government Accountability Office found. Critics have charged that EPA has moved to remove independent-minded academic experts and replace them with members in sync with President Donald Trump's administration's campaign for environmental deregulation. EPA said its senior management rigorously evaluated all appointees.

## Turing to appear on £50 note

SCIENCE HISTORY | The Bank of England announced this week it will release an updated £50 note bearing the image of Alan Turing, the mathematician and World War II codebreaker. Turing, a pioneer in computer science and artificial intelligence, was instrumental in cracking the Nazi's Enigma code, which vastly aided Allied war efforts. But Turing, who during a burglary investigation acknowledged having a physical relationship with a man, was arrested in 1952 under a law forbidding homosexuality, leading to his forced chemical castration. He died in 1954 in an apparent suicide. In 2013, Turing received a royal pardon.

## Cities' climate shifts projected

CLIMATE SCIENCE | If you live in Toronto, Canada, by 2050 the climate may resemble that of steamy Washington, D.C., today. Of the world's 520 major cities, 77% will experience a similar shift, most of them becoming warmer, wrote researchers in an article published last week in *PLOS ONE*. They quantified cities' current and future climates on a scale and considered projected effects of climate change. On average, by 2050 the climates of Northern Hemisphere cities were estimated to resemble those of cities 1000 kilometers to the south; cities in the tropics will shift to drier conditions. Ecologist Thomas Crowther of ETH Zurich in Switzerland and co-authors contend that describing such city-level climate shifts can help citizens and government planners visualize specific effects of climate change and prepare for them.

**◀ In 2050, the climate in many cities will resemble that now found in warmer cities:**

| NOW | 2050 |
|---|---|
| Seattle, Washington | San Francisco, California |
| Toronto, Canada | Washington, D.C. |
| New York City | Virginia Beach, Virginia |
| Rio de Janeiro, Brazil | Havana |
| Marseille, France | Algiers |
| Stockholm | Budapest |
| Kathmandu | Shenzhen, China |

## Biden cancer program on hold

POLITICS | The Biden Cancer Initiative announced last week that it will suspend operations to eliminate any conflict of interest with the campaign of Joe Biden, its founder and the Democratic front-runner for president. The former vice president and his wife, Jill, launched the program, based in Washington, D.C., in June 2017 to build on the White House cancer moonshot program created the year before under President Barack Obama; the initiative's mission included encouraging cancer researchers to collaborate, but it did not give scientists grants. Biden and his wife had stepped down from the board last month after concerns were raised about whether Biden, if elected president in the 2020 election, would give favorable treatment to the initiative's donors.

**IN FOCUS** A material called aerogel (above) is an excellent candidate to insulate greenhouses on Mars, scientists report in the July issue of *Nature Astronomy*. They showed that just a few centimeters of the wispy material, which is 99% air, could trap enough warmth to melt martian surface ice, which may allow astronauts to grow food. Making aerogel on Mars or transporting it there presents challenges, however.

PHOTO: NASA/JPL-CALTECH

## Accused astrophysicist fired

WORKPLACE | The board of ETH Zurich in Switzerland this week fired astrophysicist Marcella Carollo, saying she did not maintain "respectful relations," after being accused of bullying students. Carollo, one of only a few women in the university's physics department, helped found ETH Zurich's institute for astronomy in 2002 with her husband, cosmologist Simon Lilly. The university dissolved the institute in 2017, after the misconduct allegations came to light. In a previous statement from her lawyer, Carollo said she was the target of a smear campaign by a graduate student whose work she was dissatisfied with; Carollo also said ETH Zurich has not taken comparable actions in response to complaints against male professors.

## NASA adopts #MeToo rules

WORKPLACE | NASA announced on 10 July new requirements for institutions to report sexual harassment by the agency's grantees. NASA will require organizations to inform it of "any findings or determinations of sexual harassment, other forms of harassment, or sexual assault" involving a NASA-funded principal investigator (PI) or co-PI, whether on campus, online, at field sites, or at conferences, the agency wrote in a *Federal Register* notice. NASA will also require notification if a PI or co-PI is placed on administrative leave or if a university takes other administrative action resulting from a sexual harassment or assault investigation. In response, the agency could terminate grants.

## Butterfly named for biodiversity

CONSERVATION | Costa Rican scientists announced this month that they named a new butterfly species they recently identified in their country *Philtronoma cbdora*, after the United Nations's Convention on Biological Diversity, to honor the wildlife treaty, which took effect in 1993.

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

IN DEPTH

Monocultures of the Cavendish banana variety are highly susceptible to *Fusarium*.

PHOTO: ISTOCK.COM/MG81

PLANT HEALTH

# *Banana fungus puts Latin America on alert*

## Apparent detection of a devastating *Fusarium* strain in Colombia threatens exports

*By* **Erik Stokstad**

In a long-feared development, an extremely damaging banana disease has apparently reached Latin America. Late last week, the Colombian Agricultural Institute (ICA) in Bogotá confirmed that four plantations in northern Colombia have been quarantined because of suspected infection with Fusarium wilt tropical race 4 (TR4), a fungus that kills plants by clogging their vascular system. Already widespread in Asia, the disease can wipe out entire plantations.

The finding has yet to be confirmed, but countries in the region are on high alert. Neighboring Ecuador is the largest banana exporter in the world; Colombia, Costa Rica, and Guatemala are big producers as well. A major outbreak of TR4 could ruin many farmers and drive up banana prices globally. "It poses a big threat," says Rob Reeder, a plant pathologist at CABI, a nonprofit research and outreach center for plant diseases in the developing world, based in Egham, U.K. "This should really start raising alarm bells." "We should take this extremely seriously," adds Gert Kema, a plant pathologist at Wageningen University in the Netherlands.

TR4 is a variant of Panama disease, which wiped out banana plantations across Latin America in the mid-20th century. The industry recovered after it replaced the most widely cultivated banana variety at the time, Gros Michel—also known as the Big Mike—with a new one, the Cavendish, that is resistant to Panama disease and now dominates the export industry.

TR4, which easily overcomes the defenses of the Cavendish and many other banana varieties, emerged in Indonesia in the 1960s and has spread to many other countries since then. It surfaced in Jordan in 2013, in Mozambique 2 years later, and also in India, the world's largest banana producer. Scientists dreaded its jump to the Americas, suspecting it was only a matter of time: "I wasn't surprised, but I was shocked," Kema says.

> ***"This is a turning point for the industry."***
> **Gert Kema**, Wageningen University

In June, staff at a large Colombian banana plantation spotted suspicious symptoms on trees and alerted ICA. After an initial polymerase chain reaction test for the fungus DNA came back positive, ICA launched its contingency plan, closing four farms and destroying all plants within 10 meters of samples that tested positive. ICA officials also established checkpoints to disinfect vehicles and boots and expanded disease surveillance in another 1100 hectares. So far, samples from the wider area have come back negative.

To confirm the presence of TR4, samples from the four farms will be analyzed by Wageningen University and KeyGene, a plant breeding company also in Wageningen; they expect to have the strain's genome sequence in early August. "We are trying to do this as fast as possible, but it takes time," says Fernando Garcia Bastidas, a Colombian researcher at KeyGene.

The sequence may also shed light on the origins of the fungus and how it arrived in Colombia. *Fusarium* is spread largely by contaminated soil and infected plant materials. It's possible that the strain arrived with farm machinery from abroad, or was carried by traveling farm workers or tourists. Banana leaves, used for wrapping food in many countries, are another potential infection route. (Bananas themselves do not spread the disease.)

Fungicides can't save plants that are already infected with TR4, and the fungus's

spores persist in soil for decades. The only way to contain the pathogen, once a plantation is infested, is to destroy all of the plants and take the farm out of production for many years—while trying to prevent the spores from escaping in runoff.

Data on TR4's economic toll are scarce, but one estimate puts the damage for Indonesian farmers at $120 million annually, and for farmers in Taiwan at double that amount. After TR4 was detected in Queensland in Australia in 2015, "most farmers implemented extensive biosecurity measures which are very expensive and very restrictive," says James Dale of the Queensland University of Technology in Brisbane. Meanwhile, he says, "Areas in which there is TR4 are quarantined and therefore out of production. There is a double economic hit." Filipino farmers have tried to cope with the disease by abandoning infested plantations and planting on clean soil elsewhere. But that strategy only lasts until clean land runs out—and often, farmers have contaminated the new fields in the process.

TR4 may be harder to control in Central and South America than in Australia, Dale says, because many more hectares are in production and many smallholder farmers don't know about, or can't afford, control measures.

New ways to battle the scourge are on the horizon. Adding certain types of biomass to the soil and covering it in plastic can kill the spores, as the material decomposes and releases gas toxic to bacteria and fungi. In trials conducted by Kema and his colleagues in the Philippines, the technique significantly reduced the number of spores, suggesting it might help contain the disease.

The longer-term solution is the same one that saved plantations decades ago: replacing the vulnerable plants with a resistant variety. The Honduras Foundation for Agricultural Research in La Lima has spent decades breeding TR4-resistant bananas, but so far the results have not lived up to the Cavendish in properties such as taste and resistance to blemishes. A genetically modified Cavendish produced in Dale's lab has shown resistance to the fungus in early field trials (*Science*, 24 November 2017, p. 979). Dale says that banana is now in larger trials; he hopes it can be commercialized in 2023. But whether consumers will buy transgenic bananas remains a question. "It's down to public perception," Reeder says.

What's already clear from TR4's apparent arrival in Latin America, Kema says, is that banana cultivation cannot continue without major changes. "This is a turning point for the industry," he says. ■

A graduate student from the University of Alaska in Fairbanks takes an Arctic ice core earlier this year.

PHOTO: UNIVERSITY OF ALASKA IN FAIRBANKS

RESEARCH FUNDING

# *University of Alaska braces for draconian budget cuts*

## Arctic science programs could suffer, faculty fear

*By* **Michael Price**

University of Alaska (UA) administrators are scrambling to decide how to impose deep mandatory spending cuts that could hobble research programs at one of the world's premier Arctic science institutions. The UA Board of Regents this week began to consider declaring a "financial exigency" that would allow officials to take extraordinary cost-cutting measures, which are expected to include laying off some tenured faculty and unionized staff, as well as eliminating or downsizing campuses and departments. The discussion followed a 28 June decision by Governor Mike Dunleavy (R) to reject a proposed $8.7 billion state operating budget and insist on a reduction of $444 million, including a $136 million cut to the UA system.

The cut, which applies to the fiscal year that began 1 July, amounts to a 40% decrease in UA's state funding, and a 17% reduction overall. Officials at the university, which operates three flagship and 13 community campuses, has some 1200 full-time faculty, and serves about 26,000 students, say they will decide how to proceed later this month.

Dunleavy has said the cut is needed to balance the state's budget and boost annual payments to residents from oil drilling revenue. But researchers are worried about the impact on UA, a prominent player in studying climate change in the Arctic, the planet's fastest-warming region. UA in Fairbanks (UAF) is among the world's top Arctic science institutions in terms of funding and publications, according to an analysis by the University of the Arctic in Rovaniemi, Finland.

"I'm extremely frustrated," says UAF geophysicist Nettie LaBelle-Hamer. "It's not just about climate, it's also about the socioeconomics, politics … we need to be part of that."

Paul Layer, UA's vice president for academics, students, and research in Fairbanks, says one of his highest priorities "is to maintain our status in Arctic research. It's the one thing we do better than anybody."

UAF's International Arctic Research Center, which partners with scientists across the United States and Japan to study weather, ocean acidification, and other topics, is funded largely by grants from nonstate sources. But it relies on state funding to pay for support staff and operations, as well as work requested by state agencies. And at UAF's Center for Alaska Native Health Research, state funds often pay for sending researchers to remote villages, says Deputy Director Diane O'Brien. "Even when we are bringing in millions of dollars of [nonstate] support, these are research services that we depend on the university to provide from their state allocation," she says.

Others fear the uncertainty will prompt scientists to leave the university—or top candidates to reject job offers. Ironically, the cuts could make it harder to win funds from other sources. That's because faculty can often only use state or university funds to pay for the time they spend drafting grant proposals. "In [the governor's] view, all we need to do is get more federal funding and we'll be fine," LaBelle-Hamer says. "He doesn't understand the research model." ■

# *Pinpoint brain stimulation probes perception*

## Triggering a handful of neurons in a mouse brain may conjure a visual experience

*By* **Kelly Servick**

How many neurons does it take to spark a memory, a sensation, or a movement? Neuroscientists have struggled to answer this question with relatively crude methods that don't allow them to fire up individually selected brain cells. Two teams, however, have recently adapted optogenetics—a technology for stimulating neurons with light—to precisely awaken particular cells in the visual cortex of a mouse. They showed that zapping just a few neurons could trigger the same brain activity as showing the animals a visual pattern and could make them react as if they had seen that pattern. "Essentially, they take control over the internal world of the brain," neuroscientist Thomas Knöpfel of Imperial College London says of the new experiments.

"We don't know how many cells it might take to trigger a more elaborate thought, sensory experience, or emotion in a person," says Karl Deisseroth, a neuroscientist and psychiatrist at Stanford University in Palo Alto, California, who led one of the new studies, published online this week in *Science*, "but it's likely to be a surprisingly small number, given what we're seeing in the mouse."

That observation might help explain why disordered states—hallucinations, unwanted thoughts, and harmful actions—arise so readily in the brain, Deisseroth says. And single-neuron optogenetics may someday point researchers toward highly targeted ways of stamping out these states and treating symptoms of brain diseases.

Neuroscientists have spent decades watching how mice behave when parts of their brains are stimulated with electrodes or, more recently, with optogenetics, which involves introducing a gene for one of several light-sensitive proteins called opsins into neurons. In most experiments, researchers awaken opsin-bearing neurons of a specific cell type with a pulse of diffuse blue-green light. But Deisseroth's group and others have been targeting optogenetics more precisely with a red light–sensitive opsin and the sharp, penetrating beam of a near-infrared laser.

"Imagine every neuron in the brain like a key on the piano," says Rafael Yuste, a neuroscientist at Columbia University who has pioneered such experiments. "You can literally choose which neurons to turn on."

In the two new studies, Deisseroth's and Yuste's groups targeted predefined sets of cells by sculpting the laser beam into a hologram with a device called a spatial light modulator. Along with an opsin gene, they injected the gene for a molecule that fluoresces when neurons fire, allowing them to discern what cells were active. They showed the mice a pattern of drifting parallel lines on a screen and trained them to lick at a water spout when those lines were in one of two orientations (horizontal or vertical), but not in the other. They identified the cells "tuned" to fire preferentially for either the horizontal or vertical pattern.

Yuste's group, which published its experiments last month in *Cell*, found that stimulating as few as two particularly well-connected neurons made the mouse more likely to lick when the vertical bars on screen were hard to discern. In some trials, the stimulation even prompted the animals to lick when there was nothing on the screen.

The results, Yuste says, support the long-standing theory that ensembles of coactivated neurons—not individual cells—form the basic building blocks of our perceptions and memories. That's still a controversial suggestion, says Michael Brecht, a neuroscientist at Humboldt University in Berlin. It's also possible that individual neurons "just do their thing and contribute incrementally" to brain function, he says—that cells don't have to form these defined groups in order to collectively represent experiences. But future studies of precisely triggered neurons may yet resolve the role of ensembles, Brecht notes.

Deisseroth's group, meanwhile, activated larger sets of vertically or horizontally tuned neurons than in the *Cell* study, and evaluated whether mice could distinguish between the two possible perceptions. Using a newly discovered gene from a single-celled marine organism that produces a highly sensitive opsin, they found that zapping sets of roughly 10 to 20 cells that were tuned to one visual pattern or the other improved a mouse's ability to distinguish increasingly dim on-screen bars. Eventually, this stimulation alone prompted accurate "lick" or "don't lick" decisions.

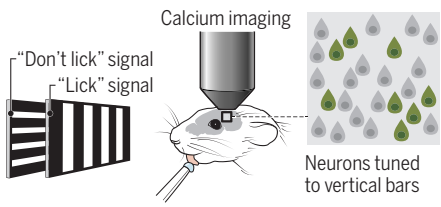
It's impossible to know whether the mice really "saw" the absent bars, but both the behavioral tests and imaging suggest "the brain is doing what it does during natural perception," Deisseroth says.

"It's probably a bit too early" to claim that optogenetic stimulation can fully recreate real vision, which is much more complex than simple moving bars, says Valentina Emiliani, a physicist at a vision institute affiliated with CNRS, the French national research agency in Paris. Still, she says, it's exciting that hitting a few neurons can call up an entire pattern of brain activity related to vision.

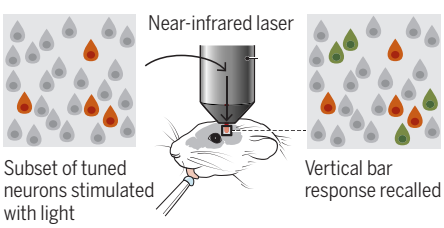
The Deisseroth and Yuste labs now plan to use single-neuron optogenetics to find neurons underlying more complex behavior—including symptoms of brain disease. Yuste has launched experiments in mice that aim to reverse symptoms of schizophrenia and Alzheimer's disease by stimulating ensembles of neurons that don't activate as strongly in the diseased mice as healthy ones. ■

### Replaying perception

In a new study, researchers used light to precisely activate cells in a mouse's visual cortex, re-creating the brain activity involved in seeing specific patterns.




**Recording**
Mice saw one of two on-screen patterns while a microscope captured which neurons were "tuned" to respond to one pattern or the other.




**Playback**
Stimulating some of those cells with light reactivated the brain's response to a pattern and made mice act as if they were seeing it again—with nothing on the screen.

GRAPHIC: L. CARRILLO-REID *ET AL., CELL* (2019) ADAPTED BY V. ALTOUNIAN/*SCIENCE*

CONSERVATION BIOLOGY

# Genomics guides help for dwindling species

## Long-term studies, modeling help clarify how to do genetic rescue

*By* **Elizabeth Pennisi**, *in Providence*

The expanding global human footprint is dividing the world's flora and fauna into ever-smaller, more isolated populations that could wink out because of inbreeding, disease, or environmental change. For decades, conservationists have proposed revitalizing those holdouts by bringing in new blood from larger populations. But they've wondered whether it really works—and how to do it without swamping the genetic identity and unique adaptations of the group at risk. Last month at Evolution 2019 here, researchers described how genomic tools are refining what is known as genetic rescue.

Although zoos have worked to maintain genetic diversity in endangered species by carefully matching individual animals for breeding, the strategy has rarely been tried in nature. Genetic rescue "should be attempted more frequently," Andrew Whiteley, a conservation genomicist at the University of Montana in Missoula, and his colleagues wrote last week in *Trends in Ecology and Evolution*. But showing that it works requires tracking multiple generations for years, something few studies have attempted. And researchers have only recently been able to detect what happens on a molecular level. Now, says Sarah Fitzpatrick, an evolutionary biologist at Michigan State University's (MSU's) W. K. Kellogg Biological Station in Hickory Corners, "We have genomic tools to study these populations … in ways we never could before."

Adding new blood to small populations really does help, a long-term experimental evolution study of wild guppies in Trinidad has demonstrated, says Brendan Reid, an MSU conservation biologist who works with Fitzpatrick. Decades ago, researchers seeded the headwaters of two streams in the mountainous country with guppies taken from a distant habitat. In one stream, the displaced fish had to travel a long way and only slowly made their way downstream to a small, isolated population. In the other stream, the fish more quickly joined another isolated group. Every month for 2.5 years, Fitzpatrick and her colleagues caught, marked, and studied all the fish they could find at the isolated groups' territories before returning the fish to the streams. They tracked the growth, survival, and genetic diversity of the fish over about seven generations.

In both streams, the populations increased 10-fold and genetic diversity doubled. Later generations were more fecund, with many of the most fit offspring being hybrids of the local and introduced fish, Reid reported at the meeting. But the findings also sounded a note of caution. In the second stream, the rapid infusion of new fish almost completely eliminated pure residents—an outcome conservationists usually hope to avoid. That result suggests "a slow trickle of immigration might be preferable," Fitzpatrick says.

A Florida scrub jay population relies on birds from other groups to sustain its genetic diversity.

PHOTO: TIM ZUROWSKI/SHUTTERSTOCK.COM

Another genomic study showed some small populations experience natural genetic rescue—and benefit from it. Nancy Chen, a population geneticist at the University of Rochester in New York, and her team study the threatened Florida scrub jay (*Aphelocoma coerulescens*), whose numbers are down to a few thousand individuals, split among a few hundred sites. For 50 years, researchers have regularly counted and assessed all the jays found at Archbold Biological Station near Lake Placid, Florida. More recently, they've collected blood samples from each bird, which enabled Chen and her colleagues to track genetic changes over time.

The team discovered that the population naturally gets a slow infusion of new blood. Typically, birds trickle in from smaller groups a few kilometers away. The newcomers are less genetically diverse than those already there, but because they are from a different population, they help maintain the resident group's diversity. However, with fewer birds arriving in recent years because of population declines, that diversity is declining, putting the population at risk of dying out. "Gene flow from small populations may be really important," she concluded at the meeting.

Most biologists have assumed that larger populations are better sources of new blood. But Chris Kyriazis, a graduate student at the University of California, Los Angeles, used computer models to study the impact of deleterious mutations hidden in a source population. Because such mutations tend to be harmful only when both parents pass the mutation to offspring, they are likely to be eliminated from historically small, inbred populations and to persist in larger ones. Kyriazis's modeling suggests intermediate-size populations, not the biggest ones, could be the best source for genetic rescues, he reported at the meeting and in a preprint posted 21 June on bioRxiv.

Sometimes, genomic results suggest the rescue strategy may backfire. Just 1000 island foxes (*Urocyon littoralis*) are left on California's Santa Catalina Island, and 60% of them have a cancer that affects their ears. Paul Hohenlohe, an evolutionary biologist from the University of Idaho in Moscow, had identified many genes that make the foxes susceptible to the cancer and wondered whether they were a candidate for genetic rescue. But he found that the Santa Catalina foxes have a genetic advantage over neighboring populations that might be sources of new blood: They have more variation throughout their genome, including in the cancer genes, he reported at the meeting. Furthermore, the Santa Catalina foxes are better adapted to the island's hot, arid climate than the other foxes, many of which live on wetter, cooler islands. So, he recommends letting nature take its course and monitoring whether the foxes eventually evolve resistance to the cancer.

These studies are helping invigorate a strategy that many believe is sorely needed. Fitzpatrick says, "The urgency of the problem and the availability of the tools makes it a really exciting time." ■

A question about citizenship won't be on the 2020 census, President Donald Trump said last week.

PHOTO: CHEN MENGTONG/CHINA NEWS SERVICE/VCG VIA GETTY IMAGES

2020 U.S. CENSUS

# *Census citizenship question is dropped, but challenges linger*

## Methods for filling in missing data will get major test

*By* **Jeffrey Mervis**

President Donald Trump's decision last week to forgo a citizenship question on the U.S. census lets the Census Bureau disappear from the headlines and resume preparations for next spring's decennial head count.

But the end of that bitterly partisan, 15-month fight doesn't mean an easy road ahead for the agency. Social scientists say little-noticed changes in how the government will conduct the 2020 census will create a major statistical challenge.

The key change, in response to a congressional mandate to hold down costs, is to hire many fewer field workers. That will result in less chasing of people who do not initially respond to the 10-question census form, which includes queries about the age, sex, race/ethnicity, and number of people living at a particular address.

Census officials say tapping information in administrative records—the vast files kept by other government agencies, including Social Security, tax, and health records—will allow the smaller army of enumerators to be more efficient. The Census Bureau will also increase its use of whole-person imputation—a method that enables statisticians to predict the characteristics, such as sex and age, of people believed to be living at a specific address who haven't responded.

Neither imputation nor administrative records has ever been used on such a large scale for the U.S. census, statisticians and demographers note, and some worry how well the techniques will perform.

The size of the challenge is staggering. Overall, census officials hope 60% of households will respond when invited to fill out the questionnaire. That's down from its earlier estimate of 65%, and a steep drop from the 74% who responded in 2010. For the first time, residents will also be able reply online or by calling a toll-free number. But a 40% nonresponse rate would mean roughly 50 million households will go uncounted.

In the past, the bureau deployed more than 500,000 census workers to visit nonresponsive households up to six times in a bid to collect missing information, sometimes by asking neighbors or landlords for help. The bureau wants to shrink that number by nearly half in 2020, keeping with direction from Congress not to exceed the $12.5 billion price tag of the 2010 count. (The bureau estimates tracking a larger, harder to reach, and more mobile population in 2020 will cost $15.7 billion.)

Experts believe high nonresponse rates will force the bureau to expand its use of whole-person imputation. In the last census, in 2010, the agency used the technique to fill in information for just about 1% of households. But that could rise to 12% of households in 2020 because of the reduced follow-up, according to a 2017 analysis by Andrew Keller of the agency's decennial statistical studies division in Suitland, Maryland.

Imputation "has served [the Census Bureau] well in the past," says Robert Santos, vice president and chief methodologist at the Urban Institute in Washington, D.C. But the potential scale of its use in 2020 is daunting, and the Census Bureau is flying blind. It "has never really done research on whole-person imputation," he says, "because they say it can't be done."

As for using administrative records to fill gaps, officials have said the approach could help target follow-up efforts by identifying residences that are vacant or don't actually exist. But researchers warn the records can be inaccurate, outdated, incomplete, or not suited to the needs of the Census Bureau. The Social Security Administration, for example, doesn't retain the addresses of people to whom it mails a Social Security card.

The records "are OK for [characterizing] part of the population," says John Thompson, who stepped down as census director in June 2017 and now lives in Oregon. But he says the people who are more likely to snub the census are also more likely to be missing from other government records. (The Census Bureau declined requests to discuss its use of these methods.)

Such issues could also entangle Trump's hope of using administrative records to tally noncitizens. When Trump dumped the citizenship question on 11 July, he also issued an executive order that requires every federal agency to share with the Census Bureau all records relevant to citizenship. Commerce Secretary Wilbur Ross estimated last year that the bureau cannot determine the citizenship status of between 30 million and 35 million residents—roughly 10% of the U.S. population—because it can access only a subset of relevant federal records. Trump said his new order would remedy that situation "immediately," and lead to a tally what would cover "100%" of all citizens, noncitizens, and those he calls "illegal aliens."

But social scientists say it is not that simple. "The quality of the data is not what [Trump] thinks it is, and merging them is difficult," says demographer William O'Hare, a veteran census data cruncher based in Baltimore, Maryland. "But an even bigger problem" for Trump, he says, is that administrative records can't identify people who are living in the country without the proper documents. "Many of the people you want," he says, "are not there." ■

# A PAINFUL LEGACY

## Mice hint at how people's emotional trauma may affect the biology of their children—and their children's children

*By* **Andrew Curry**, *in Zurich, Switzerland*

The children living in SOS Children's Villages orphanages in Pakistan have had a rough start in life. Many have lost their fathers, which in conservative Pakistani society can effectively mean losing their mothers, too: Destitute widows often struggle to find enough work to support their families and may have to give up their children.

The orphanages, in Multan, Lahore, and Islamabad, provide shelter and health care and send kids to local schools, trying to provide "the best possible support," says University of Zurich (UZH) physician and neuroscientist Ali Jawaid. "But despite that, these children experience symptoms similar to PTSD [post-traumatic stress disorder]," including anxiety and depression.

Beyond these psychological burdens, Jawaid wonders about a potential hidden consequence of the children's experience. He has set up a study with the orphanages to probe the disturbing possibility that the emotional trauma of separation from their parents also triggers subtle biological alterations—changes so lasting that the children might even pass them to their own offspring.

That idea would have been laughed at 20 years ago. But today the hypothesis that an individual's experience might alter the cells and behavior of their children and grandchildren has become widely accepted. In animals, exposure to stress, cold, or high-fat diets has been shown to trigger metabolic changes in later generations. And small studies in humans exposed to traumatic conditions—among them the children of Holocaust survivors—suggest subtle biological and health changes in their children.

The implications are profound. If our experiences can have consequences that reverberate to our children or our children's children, that's a powerful argument against everything from smoking to immigration policies that split families. "This is really scary stuff. If what your grandmother and grandfather were exposed to is going to change your disease risk, the things we're doing today that we thought were erased are affecting our great-great-grandchildren," says Michael Skinner, a biologist at Washington State University in Pullman.

Skinner's own research in animals suggests changes to the epigenome, a swirl of biological factors that affect how genes are expressed, can be passed down through multiple generations (*Science*, 24 January 2014, p. 361). If trauma can trigger such epigenetic changes in people, the alterations could serve as biomarkers to identify individuals at greater risk for mental illness or other health problems—and as targets for interventions that might reverse that legacy.

But proving that emotional trauma, as distinct from physical stress, can be passed on to subsequent generations in people is a challenge. "The difficulty … is being able to disentangle what comes through social inheritance—which must be massive—and what doesn't," says neuroscientist Johannes Bohacek of ETH Zurich. "The jury is still out on humans."

Some of the field's biggest names also worry that the idea could have dangerous consequences. Rachel Yehuda of the Icahn School of Medicine at Mount Sinai in New York City studied the children of 40 Holocaust survivors and found lower baseline levels of the stress hormone cortisol as well as a distinctive pattern of DNA methylation, an epigenetic marker. But in a paper last year, she said it would be "premature" to conclude that trauma causes heritable changes, adding that hyped media coverage could promote a misleading narrative of hopelessness, suggesting that one generation's trauma permanently scars later generations.

"There's a lot of overinterpretation of initial results," says Columbia University biologist Katherine Crocker, who studies nongenetic inheritance in crickets. "What is out there in the public mind about epi-

PHOTO: PIOTR PIWOWARSKI

Isabelle Mansuy (shown in her lab in Zurich, Switzerland) found that mice with traumatized forebears took more risks when exploring open, elevated platforms (left).

PHOTO: PIOTR PIWOWARSKI

genetics probably can never be proved."

To investigate, Jawaid is collecting blood and saliva samples from the Pakistani orphans and from classmates who live with parents. As a researcher in the lab of Isabelle Mansuy of UZH and ETH Zurich, he hopes to learn whether the trauma of loss and forced separation has left identifiable marks at the cellular level. But to really prove transgenerational inheritance, he'd have to study the orphans for years—until they have children of their own. That's why Mansuy herself has turned to mice.

**ONE RECENT AFTERNOON**, Mansuy donned a fresh lab coat and blue sanitary booties and gently cracked the door of a darkened room at her lab at UZH. A powerful smell—something like dog chow mixed with animal musk—wafted out on a gust of warm air. Inside were hundreds of mice in 40 breeding cages. "We keep it dark during the day to preserve their circadian rhythm when we work with them," Mansuy says in a hushed voice. "This is our 31st cohort."

The idea Mansuy is exploring—that not all inherited characteristics are rooted in DNA—dates back more than half a century. Tantalizing early results came from maize, in which plants with identical DNA had variations in traits such as kernel color that persisted for hundreds of generations. The work was initially controversial, as geneticists saw it as a revival of the non-Darwinian ideas of 19th century scientist Jean-Baptiste Lamarck.

But experiments in many organisms suggested epigenetic inheritance was real. In simple creatures like *Caenorhabditis elegans* worms, researchers found that genes turned off once by altering the RNA they produced remained silenced for 80 generations or more. Some examples were even more dramatic: Water fleas exposed to the scent of a predator have offspring with spiky, armored heads. And in mice, researchers including Skinner found that parents exposed to altered diets, low temperatures, or toxins had descendants with behavioral changes and weight gain.

Epidemiological studies of people have revealed similar patterns. One of the best-known cases is the Dutch hunger winter, a famine that gripped the Netherlands in the closing months of World War II. The children of women pregnant during the food shortages died earlier than peers born just before, and had higher rates of obesity, diabetes, and schizophrenia. Studies of other groups suggested the children of parents who had starved early in life—even in the womb—had more heart disease. And a look last year at historical records showed the sons of Civil War soldiers who had spent time as prisoners of war (POWs) were more likely to die early than the sons of their fellow veterans. (The researchers controlled for socioeconomic status and maternal health.)

But the human studies faced an obvious objection: The trauma could have been transmitted through parenting rather than epigenetics. Something about the POW experience, for example, might have made those veterans poor fathers, to the detriment of their sons' lives. The psychological impact of growing up with a parent who starved as a child or survived the Holocaust could itself be enough to shape a child's behavior. Answering that objection is where mouse models come in.

Mansuy began in 2001 by designing a mouse intervention that re-creates some aspects of childhood trauma. She separates mouse mothers from their pups at unpredictable intervals and further disrupts parenting by confining the mothers in tubes or dropping them in water, both stressful experiences for mice. When the mothers return to the cage and their pups, they're frantic and distracted. They often ignore the pups, compounding the stress of the separation on their offspring.

Mansuy says the mice's suffering has a

purpose. "We're applying a paradigm that is inspired by human conditions," she says. "We're doing it to gain understanding for better child health."

Unsurprisingly, the pups of stressed mothers displayed altered behavior as adults. But to Mansuy's surprise, the behavioral changes persisted in the offspring's offspring. Initially, she thought this could be a result of the offspring's own behavior: Mice traumatized as pups could have been bad parents, replicating the neglect they experienced in childhood. Thus they might simply be passing on a behavioral legacy—the same lasting psychological effect that might explain such findings in humans.

To rule out that possibility, Mansuy studied only the male line (see graphic, p. 215), breeding untraumatized, "naïve" female mice with traumatized males, and then removing males from the mother's cage so that their behavior did not impact their offspring. After weaning, she raised the mice in mixed groups to prevent litter mates from reinforcing each other's behaviors.

Her lab repeated the procedure, sometimes going out six generations. "It worked immediately," she says of the protocol. "We could see that there were symptoms [in descendants] that were similar to the animals that were themselves separated." Descendants of stressed fathers displayed more risk-taking behavior, like exploring exposed areas of a platform suspended off the ground. When dropped in water, they "gave up" and stopped swimming sooner than control mice, an indicator of depressivelike behavior in mice.

Mansuy is "definitely a pioneer," says Romain Barrès, a molecular biologist at the University of Copenhagen. Other researchers have developed conceptually similar models, for example giving male mice altered diets or exposing them to nicotine and tracing metabolic and behavioral changes out for generations.

"If you're asking, 'Does the experience of the parent influence the process of development?' the answer is yes," says epigenetics researcher Michael Meaney at McGill University in Montreal, Canada, whose own studies have shown that differences in maternal care can have epigenetic effects on brain development. "Isabelle and others have documented the degree to which the experience of the parent can be passed on. The question [is] how."

**THREE MASSIVE FREEZERS** down the hall from Mansuy's office are filled with samples of mouse blood, liver, milk, microbiome, and other tissues. These serve as a −80°C archive of more than 10 years of data. Mansuy estimates she's collected behavioral data and tissue samples from thousands of mice altogether.

She hopes the biological markers of trauma are hidden in those freezers, waiting to be revealed. Many of the early mammalian epigenetics studies focused on DNA methylation, which "tags" DNA with methyl groups that switch genes off. But those changes seemed unlikely to be directly inherited: In mammals, methylation is mostly erased when egg and sperm come together to form an embryo.

Mansuy and others still think methylation could have some role. But they are also zeroing in on tiny information-rich molecules called small noncoding RNAs (sncRNAs). Most RNA is copied from DNA, and then acts as a messenger to instruct the cell's ribosomes to produce specific proteins. But cells also contain short strands of RNA that don't produce proteins. Instead, these noncoding RNAs piggyback on the messenger RNAs, interfering with or amplifying their function, thus causing more or less of certain proteins to be produced.

Mansuy and others think stress may influence sncRNAs, along with the many other biochemical changes it causes, from higher levels of hormones like cortisol to inflammation. They have focused on the sncRNAs in sperm, which may be especially vulnerable to stress during the weeks that newly formed sperm spend maturing in a twisting tube on top of the testes. Later, when sperm and egg come together, altered sncRNAs could modify the production of proteins at the very beginning of development in a way that ripples through the millions and millions of cell divisions that follow. "Hosts of signals happen as those cells become a zygote," says epigeneticist Tracy Bale at the University of Maryland in Baltimore. "If dad brings small noncoding RNAs that have an effect on mom's RNAs, that can change the trajectory of embryo development."

Ali Jawaid (back right) works with children in a Pakistani orphanage. One boy has just given a blood sample for an epigenetics study.

PHOTO: ALI JAWAID

Bale found evidence that trauma can affect sncRNAs in sperm—and that the effects might be transmitted to offspring. She stressed mice during adolescence by barraging them for weeks at unpredictable intervals, with things like fox odors, loud noises, and bright light. Then, she examined the sncRNAs in their sperm and offspring. She found differences in nine types of sncRNAs, including one that regulates *SIRT1*, a gene that affects metabolism and cell growth.

She then created RNA molecules with similar alterations and injected them into early-stage embryos. When those embryos grew to adults, they carried RNA alterations like those seen in the sperm. This second generation also had lower levels of corticosterone, the mouse equivalent of cortisol, after a stressful spell inside a tight tube. "If you do the same RNA changes, you produce offspring with the same phenotype," Bale says.

Mansuy found similar RNA changes in her male mice traumatized as pups. They had higher levels of specific sncRNAs, including miR-375, which plays a role in stress response. Mansuy is convinced those molecular changes account for some of the inherited behavioral traits she documented. In one experiment, her team injected RNA from traumatized male sperm into the fertilized eggs of untraumatized

parents and saw the same behavioral changes in the resulting mice.

But although the cause, in the form of altered RNA, and the effect, in the form of altered behavior and physiology, are identifiable in mouse experiments, everything else remains maddeningly difficult to untangle, especially in people. "The field has come a long way in the last 5 years," Bale says. "But we don't know what's going on in humans because we don't have a controlled environment."

Still, mouse data in hand, Mansuy has been looking for similar epigenetic changes in people. She analyzed blood samples from Dutch soldiers, collected before and after deployment to Afghanistan between 2005 and 2008. And she's working with clinicians in Nice, France, to examine blood samples from survivors of a horrific 2015 terror attack.

Other researchers had found altered sncRNAs in the blood of the soldiers. In 2017, for example, Dutch researchers showed soldiers exposed to combat trauma had recognizable differences in dozens of sncRNA groups, some of them correlated with PTSD. But Mansuy couldn't find the same kinds of RNA changes that appeared in her lab's mice. That could be because the soldiers' samples were years old, or simply because mice and people are different, showing the limits of mouse models. But Mansuy hopes it means epigenetic changes are sensitive to the type of trauma and when it occurs in the life course. Mice can never perfectly replicate human suffering, but, she says, "the best approach" for research "is to select a population of humans who have gone through conditions which are as similar as possible to our model."

That's where the Pakistani orphans come in. The children's chaotic early years may have some similarities to what the mice in Mansuy's lab experience, she says, including unpredictable separation from their mothers.

Early results are promising. "We have overlapping findings with the mouse model," Jawaid says. In a preprint uploaded last month to bioRxiv, Mansuy and Jawaid documented changes in the levels of fatty acids in the orphans' blood and saliva that mimicked changes in the traumatized mice—as well as similar sncRNA alterations. The presence of similar biomarkers "suggests that comparable pathways are operating after trauma in mice and children," Mansuy says.

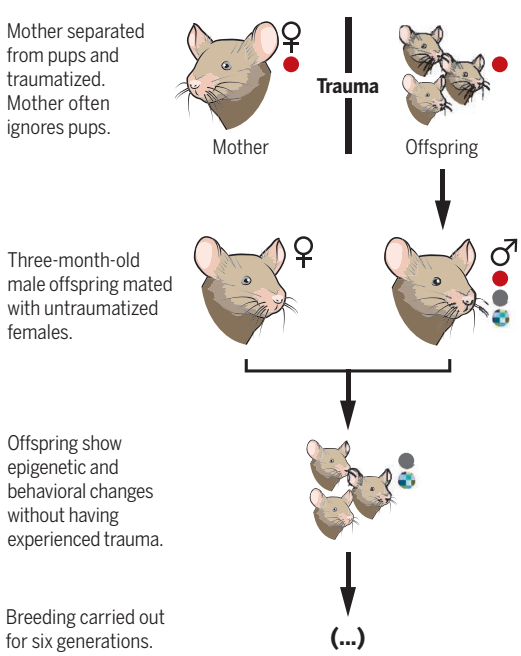
In a conceptually similar effort to go from mice to people, biologist Larry Feig at Tufts University in Boston exposed male mice to social stress by routinely changing their cage mates. Their sperm had altered levels of specific sncRNA groups—albeit different ones from those altered in Mansuy's mice—and their offspring were more anxious and less sociable than the offspring of unstressed parents.

Working with a sperm bank, Feig then looked for the same sncRNAs in human sperm. He also asked donors to fill out the Adverse Childhood Experience (ACE) questionnaire, which asks about abusive or dysfunctional family history. The higher the men's ACE score, the more likely they were to have sperm sncRNA profiles matching what Feig had seen in mice.

But this body of research hasn't convinced everyone. Geneticist John Greally at the Albert Einstein College of Medicine in New York City has been a vocal critic of the evidence for epigenetic inheritance of trauma, pointing at small sample sizes and an overreliance on epidemiological studies. For now, he says, "Mouse models are the way to go." He's not yet seen definitive experiments even in mice, he says. "I'd like to see us be more bold and brave and move from preliminary association studies to definitive studies—and be open to the idea that there may be nothing there."

## Troubled offspring

To explore how trauma affects generations of mice, such as the grandfather, father, and son below, researchers stressed mother mice. Their pups then exhibited both molecular and behavioral changes, such as taking more risks on an elevated maze. These changes persisted for up to five generations.




**IN A DARKENED ROOM** down the hall from Mansuy's office, just outside the mouse breeding area, two cages stand side by side on a table. One is a standard lab mouse enclosure, not much bigger than a shoebox. Wood chip–strewn cages like this are where most lab mice, including most of Mansuy's animals, spend their lives.

Next to it, black-furred, pink-tailed mice scurry up and down in a luxury two-story mouse house, equipped with three running wheels and a miniature maze. Their environment is designed to stimulate their senses and engage more of their brains in play and exploration (*Science*, 9 February 2018, p. 624).

In 2016, Mansuy published evidence that traumatized mice raised in this enriched environment didn't pass the symptoms of trauma to their offspring. The limited data—Mansuy says her lab is now working on an expanded study—suggest life experience can be healing as well as hurtful at the molecular level. "Environmental enrichment at the right time could eventually help correct some of the alterations which are induced by trauma," Mansuy says.

This and a few other studies suggesting epigenetic change is reversible have the potential to change the narrative of doom around the topic, researchers say. "If it's epigenetic, it's responsive to the environment," says Feig, who more than a decade ago found similar effects on brain function across generations by giving mice play tubes, running wheels, toys, and larger cages. "That means negative environmental effects are likely reversible."

In public talks and interviews, Mansuy says she's careful not to promise too much. As confident as she is in her mouse model, she says, there's lots more work to be done. "I don't think the field is moving too fast," Mansuy says. "I think it's moving too slow." ■

*Andrew Curry is a journalist in Berlin.*

CREDITS: (GRAPHIC) V. ALTOUNIAN/*SCIENCE*; (PHOTO) PIOTR PIWOWARSKI





CHEMISTRY

# How to make interlocked nanocarbons

## Interlocked nanocarbon rings have promising properties for molecular machines

*By* **Jeff M. Van Raden** *and* **Ramesh Jasti**

Carbon-rich materials such as fullerenes, carbon nanotubes, and graphene have a wide range of unusual physical properties resulting from their unique topography. For example, graphene, a two-dimensional sheet material consisting solely of carbon atoms, is a zero-gap semiconductor (*1*). When this same material is rolled into a cylindrical topology—a carbon nanotube—the resulting material can be either metallic or semiconducting, depending on the specific atom connectivity. Many other molecular entities can be synthesized entirely from carbon (*2*). On page 272 of this issue, Segawa *et al.* (*3*) report the synthesis of nanocarbons that are mechanically interlocked. These mechanically bound nanocarbons create a bridge between carbon nanoscience and research into molecular machines.

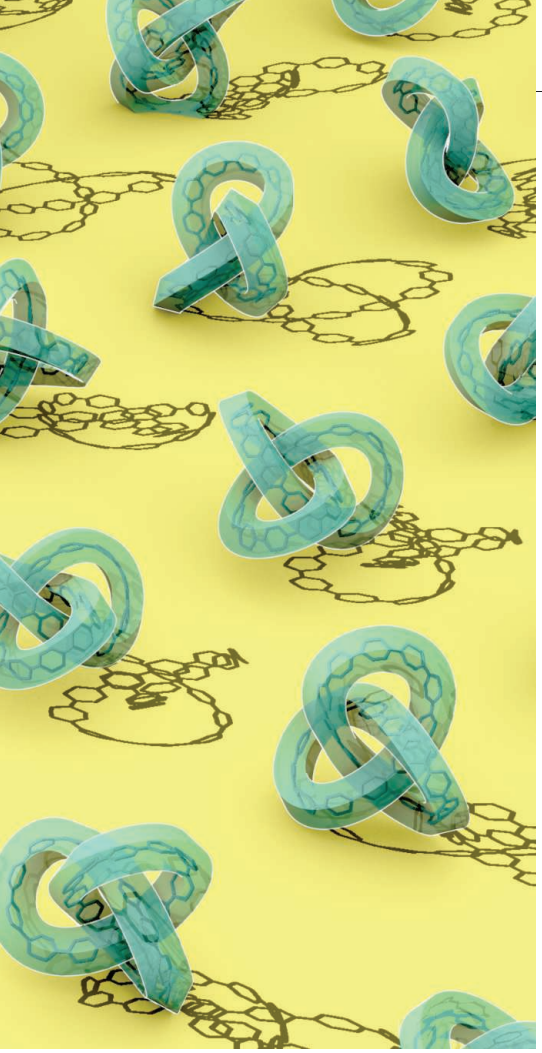
Segawa *et al.* created nanotube slices, that is, small fragments of a [12,12] armchair carbon nanotube called [12]cycloparaphenylene ([12]CPP) (see the figure). Initially described in 2008 (*4*), [*n*]CPPs consist of *n* benzene rings linked in the para position, resulting in a substantial amount of strain energy. To prepare these strained molecules, Segawa *et al.* used unstrained cyclohexadiene moieties that they then converted to benzenes in a late step of the synthesis.

The authors formed the key mechanical bond using a silicon-based template method, reminiscent of the original concept demonstrated by Sauvage and colleagues (*5*). In this method, a tetrahedral silicon atom is used to adjoin two neighboring CPP fragments in a crossing pattern (see the figure). After dual macrocyclization, the silicon tether can be removed and the interlocked structure remains. Finally, reductive aromatization converts the cyclohexadiene units into the all-benzene structure. Expanding on this method, Segawa *et al.* also prepared a molecular trefoil knot, the simplest knot that can only be untied by cutting one of the strands. This synthetic approach does not require metal coordination (*6*, *7*) or host-guest interactions (*8*) and thus greatly expands the types of mechanically interlocked structures that can be synthesized.

At the nanoscale, molecular motion is less affected by forces that are common to typical macroscopic objects, such as gravity and momentum (*9*). In contrast, factors such as Brownian motion and viscosity dominate in the operation of synthetic molecular machines. The class of molecules known as mechanically interlocked molecules provides a way to overcome these

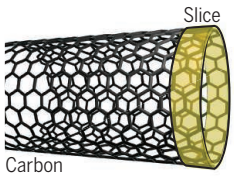
*Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403, USA. Email: rjasti@uoregon.edu*

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

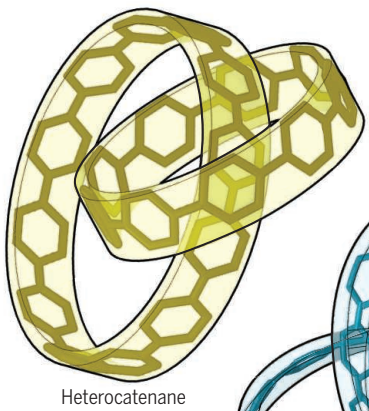


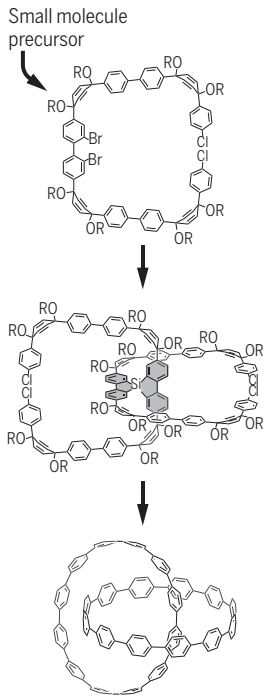
## Nanocarbons, interlocked

Mechanically interlocked molecules are important components of molecular machines. Segawa *et al*. report the synthesis of such molecules consisting entirely of nanocarbons formed by cycloparaphenylenes (CPPs).




**Interlocking nanotube slices**
The authors create interlocked rings (catenanes) and a structure called a trefoil knot, with the same chemical structure as a nanotube slice.




**Synthesis of [12]CPP catenane**
The synthesis starts from small molecule precursors and involves a silicon templating step. Trefoil knots are created in a similar stepwise manner.

challenges; thus, they are key building blocks of synthetic machines that operate on the molecular length scale, as envisioned by Richard Feynman.

The synthetic work of Segawa *et al.* differs from typical mechanically interlocked molecules in that the molecular composition of each molecule consists only of rigid, $sp^2$-hybridized carbon atoms and edge hydrogen atoms. Their findings show that these interlocked carbon nanostructures have various distinctive chemical properties. For example, the molecular trefoil knot shows only a single proton resonance in the proton nuclear magnetic resonance ($^1$H-NMR) spectrum. This is notable given that the molecular knot is composed of 24 individual benzene rings, each of which contains four protons. The observation of a single resonance rather than 96 individual resonances indicates that the molecular trefoil knot undergoes ultrafast motion on the NMR time scale. This ultrafast motion occurs even at −95°C—a temperature that would be expected to greatly reduce the motion of any typical molecule. Isobe and colleagues recently reported a similar observation in a carbon-rich molecular bearing (*10*). Taken together, these studies suggest that carbon-rich architectures may provide a platform to study frictionless motion at the nanoscale.

Beyond molecular motion, a particularly interesting consequence of interlocking molecules occurs in the case of [2]heterocatenane, in which [12]CPP and a smaller macrocycle, [9]CPP, are linked via a mechanical bond (see the figure). Excitation of the [2]heterocatenane with light results in emission only from the [9]CPP macrocycle, implying energy transfer from [12]CPP to [9]CPP via the mechanical bond. In contrast, excitation of a solution containing both noncatenated [12]CPP and [9]CPP results in emission from each macrocycle. These optoelectronic properties, in combination with the mechanical stiffness of the structure, render these catenanes promising candidates for use in new types of advanced sensing materials.

Many aspects of the electronic and optical behavior of these mixed catenane systems remain to be explored. Also, it is not yet clear through which mechanisms they can be switched. From a synthetic standpoint, the topological complexity can be furthered advanced, yielding interlocked structures made of three or more CPPs or even perhaps polymeric versions akin to molecular chain mail.

As the demand for miniaturized technology increases, the ability to shrink machines to the nanoscale represents a major challenge. The construction of mechanically interlocked molecules provides a strong foundation for addressing these challenges. As the work reported by Segawa *et al.* shows, it is the imagination and skill of synthetic chemists that will continue to play a key role in developing new structural frameworks. The carbon-rich structures that they report will provide focus points for nanocarbon research and studies of molecular machines. ■

### REFERENCES AND NOTES


1. A. K. Geim, K. S. Novoselov, *Nat. Mater.* **6**, 183 (2007).
2. M. A. Majewski, M. Stępień, *Angew. Chem. Int. Ed.* **58**, 86 (2019).
3. Y. Segawa *et al.*, *Science* **365**, 272 (2019).
4. R. Jasti, J. Bhattacharjee, J. B. Neaton, C. R. Bertozzi, *J. Am. Chem. Soc.* **130**, 17646 (2008).
5. C. O. Dietrich-Buchecker, J. P. Sauvage, J. P. Kintzinger, *Tetrahedron Lett.* **24**, 5095 (1983).
6. Y-Y. Fan *et al.*, *Nat. Commun.* **9**, 3037 (2018).
7. J. M. Van Raden, B. M. White, L. N. Zakharov, R. Jasti, *Angew. Chem. Int. Ed.* **58**, 7341 (2019).
8. Y. Xu *et al.*, *J. Am. Chem. Soc.* **140**, 13413 (2018).
9. S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, A. L. Nussbaumer, *Chem. Rev.* **115**, 10081 (2015).
10. T. Matsuno, Y. Nakai, S. Sato, Y. Maniwa, H. Isobe, *Nat. Commun.* **9**, 1907 (2018).


10.1126/science.aay2861

GRAPHIC: V. ALTOUNIAN/*SCIENCE*, CHEMICAL STRUCTURES FROM (3)

COLD-ATOM PHYSICS

# Polarons leave a trace

## Spin and charge interplay leads to stringlike excitations in the 2D Hubbard model

*By* **Peter Schauss**

The physics of large many-particle quantum systems is barely tractable with first-principles approaches but can often be explored through quantum simulation, in which a well-controlled system, such as interacting ultracold atoms, acts as an analog quantum computer for a problem of interest. Great progress has been made in the experimental preparation of low-entropy states of ultracold atoms in optical lattices for this purpose. Typical target systems are numerically intractable and not realized in nature in their ideal form, such as the Fermi-Hubbard model. On page 251 of this issue, Chiu *et al.* (*1*) used single-atom and single-site resolved imaging of ultracold atoms on a lattice to study the doped Fermi-Hubbard model and report the nature of its microscopic correlations in real space.

Hubbard models are a class of models in which interacting quantum particles move by tunneling between sites on a lattice. Quantum simulation of the Fermi-Hubbard model becomes particularly interesting in the case of strong interactions, in which calculations on classical computers are very hard and limited to small systems. Several mysteries remain unresolved, such as the nature of the low-temperature quantum phases.

Chiu *et al.* tackled the question of the interplay of spin and charge in the two-dimensional (2D) Hubbard model. They found string patterns, which are correlated 1D excitations in the 2D system. Ultracold atomic quantum systems are intriguing systems for quantum simulation because of their extremely good isolation from the environment and superior parameter tunability (*2*, *3*). Quantum gas microscopes can detect hundreds of individual ultracold atoms in optical lattices with single-site resolution. This enables the quantum simulation of increasingly complex phenomena in the 1D and 2D Hubbard model by use of quantum gas microscopes.

*Department of Physics, University of Virginia, Charlottesville, VA 22904-4714, USA. Email: ps@virginia.edu*

Although the particle interactions are simple, the Hubbard model contains a wide variety of interesting quantum effects. One recent breakthrough in quantum simulation of the Fermi-Hubbard model was the detection of antiferromagnetic correlations at one atom per site, called half-filling (*4*–*10*). These antiferromagnetic correlations in a spin-balanced Fermi lattice system arise as a consequence of Pauli blocking, the inability of two fermions to occupy the same quantum state. In perturbation theory, this leads to a reduction of kinetic energy of neighboring spin states with opposite spins compared with equal spins. These correlations can extend over the entire system (*11*).

In contrast to the 1D Fermi-Hubbard model in which spin and charge decouple (*12*, *13*), the 2D case adds fundamental difficulties because of the lack of an effective low-energy theory that describes the spin and charge interplay. In particular, Chiu *et al.* studied stringlike excitations on the antiferromagnetic background in the doped Hubbard model below half filling. The origin of the strings can be understood in a simple semiclassical picture (see the figure). Considering an antiferromagnetically ordered state with few hole-like excitations, the holes shift if a neighboring atom tunnels to the empty lattice site. If the hole moves over several lattice sites, a stringlike excitation stretches from the initial to the final hole position, and is characterized by inverted antiferromagnetic order compared with the background.

This model is highly simplified. First, the system has no preferred spin axis and cannot be mapped to classical spins. Also, the temperature in the real system is finite and leads to thermal excitations in the antiferromagnetic background. Additionally, the holes influence the magnetic correlations around them, and the resulting "dressed holes" can be described as quasiparticles, known as polarons (*14*). These polarons can then be thought of as leaving a trace in the magnetic correlations along the path they have taken in the 2D system. Because of the variety of different paths, these string correlations do not show up as correlation in two-point correlation functions and represent a type of hidden order.

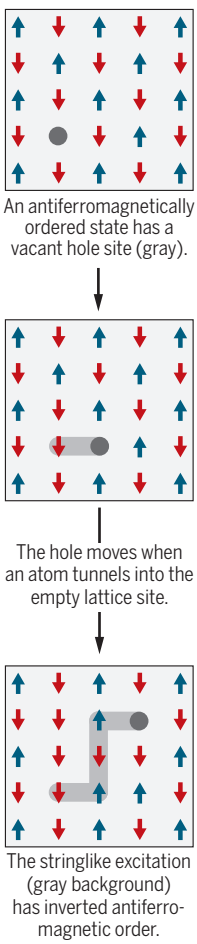
Chiu *et al.* used a variety of analysis techniques to study the statistical distribution of these stringlike traces. The possibility of overlapping strings and the projection of the quantum spins caused by the measurement complicates this analysis. To avoid detection bias, they used different string detection algorithms and compared these with simulated data created from different theory models. The large statistical samples needed to observe significant numbers of long strings in the systems made the data acquisition and evaluation challenging.

On the basis of these evaluations, Chiu *et al.* finally found evidence of the presence of correlated strings. The analytic string theory, which is consistent with the experimental findings, gives an intuitive view on the coupling of spin and charge excitations in the Fermi-Hubbard model. This study relies heavily on the repeatable preparation of low-entropy samples in the optical lattice that can be imaged with single-atom and single-site resolution. These findings alone are a demonstration of the wide range of strongly correlated effects that can be directly studied with quantum gas microscopes (*15*), including hidden order.

The results provide a real-space view on the interplay between spin and charge excitations in the Hubbard model that is complementary to condensed matter experiments. The theoretical models investigated could guide promising new descriptions of the low-energy behavior of the Fermi-Hubbard model. Furthermore, the measurements motivate quantum simulation using ultracold atoms in optical lattices to explore new types of strongly correlated states arising in Hubbard systems. ■

### Strings attached

Chiu *et al.* studied stringlike excitation in a lattice of two atomic fermions (red and blue).



An antiferromagnetically ordered state has a vacant hole site (gray).

The hole moves when an atom tunnels into the empty lattice site.

The stringlike excitation (gray background) has inverted antiferromagnetic order.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*


### REFERENCES AND NOTES

1. C. S. Chiu *et al.*, *Science* **365**, 251 (2019).
2. C. Gross, I. Bloch, *Science* **357**, 995 (2017).
3. I. Bloch, *Nat. Phys.* **14**, 1159 (2018).
4. D. Greif *et al.*, *Science* **340**, 1307 (2013).
5. R. A. Hart *et al.*, *Nature* **519**, 211 (2015).
6. M. F. Parsons *et al.*, *Science* **353**, 1253 (2016).
7. M. Boll *et al.*, *Science* **353**, 1257 (2016).
8. L. W. Cheuk *et al.*, *Science* **353**, 1260 (2016).
9. J. H. Drewes *et al.*, *Phys. Rev. Lett.* **118**, 170401 (2017).
10. P. T. Brown *et al.*, *Science* **357**, 1385 (2017).
11. A. Mazurenko *et al.*, *Nature* **545**, 462 (2017).
12. T. Giamarchi, *Quantum Physics in One Dimension* (Oxford Univ. Press, 2004).
13. T. A. Hilker *et al.*, *Science* **357**, 484 (2017).
14. G. Salomon *et al.*, *Nature* **565**, 56 (2019).
15. M. Endres *et al.*, *Appl. Phys. B* **113**, 27 (2013).


10.1126/science.aax6486

# An attractive, reshapable material

## Jamming the nanoparticles of a ferrofluid on droplet surfaces creates a soft ferromagnet

*By* **Rémi Dreyfus**

Permanent magnets are made from ferromagnetic materials like iron. An external magnetic field can align the electron spins, and strong spin-spin interactions can keep the spins aligned even after the field is removed. This remanent magnetization remains unless the material is heated above its Curie temperature, at which point thermal agitation overcomes the spin coupling and spins become disordered (*1*). Materials with weaker spin-spin interactions, such as aluminum and ferrous oxide, can form paramagnets, which become magnetized only when an external field is applied. The strong spin-spin interactions underlying ferromagnetism would seem to require the close atomic proximity found in dense solids. However, on page 264 of this issue, Liu *et al.* (*2*) present a way to prepare soft, reconfigurable millimeter-size magnets prepared from small droplets of ferrofluid (*3*) in water.

Ferrofluids are liquid suspensions, typically of iron oxide nanoparticles, which can be ferromagnetic or paramagnetic. The fluid nature of the system limits the magnetic response to paramagnetic properties. The nanoparticles can move within the liquid state, so when the magnetic field is removed, thermal agitation disorders the particles and the fluid cannot support permanent magnetization (*3*).

However, in the experiments of Liu *et al.*, as droplets of ferrofluid are formed, a surfactant is added to the outer phase, which drives the nanoparticles in the droplets to the droplets' surface. They form a very closely packed crust of magnetic nanoparticles. This kind of emulsion, where the interface is saturated with nanoparticles, is called a Pickering emulsion (*4*), which has various potential applications as catalytic reactors (*5*), encapsulating agents (*6*), or even microsensors (*7*).

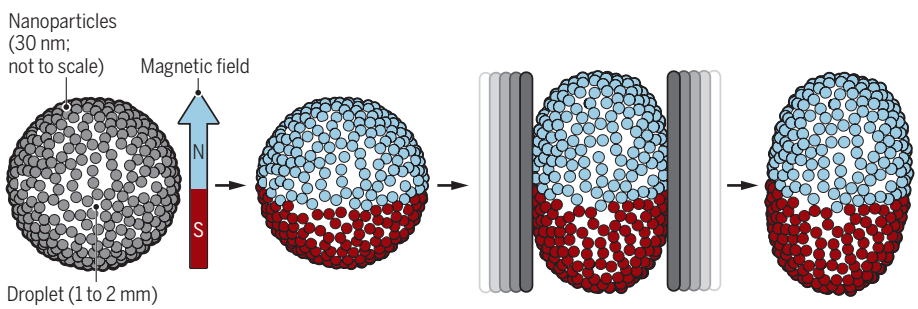
Liu *et al.* first formed a Pickering emulsion of a ferrofluid where nanoparticles covered the interface in a jammed state (see the figure). They then applied a magnetic field, which allowed the droplets to acquire magnetization. Unlike a conventional ferrofluid, this magnetization remained even when the external magnetic field was turned off. The formed droplets were therefore ferromagnetic. The authors measured the magnetization cycle, which exhibits the features of a soft ferromagnetic (*1*) material—that is, weak remanent magnetization and a narrow hysteresis loop. They hypothesized that the interactions between the nanoparticles pinned to the interface and the ones remaining in the bulk were responsible for this new phenomenon. Even more astoundingly, they showed how these magnets could be shaped and reshaped. These results now bring ferromagnetism to the world of reconfigurable soft matter, whereas until now ferromagnets have been limited to hard solids.

This approach suggests a number of potential applications. One direct application lies in using these materials for external actuation systems. For example, there is currently a very strong collective effort in designing robotic systems made from soft materials, which can enable new modes of locomotion and object manipulation (*8*). However, actuation of these systems remains a challenge, and requires rather bulky pneumatic equipment. The ferromagnetic materials created by Liu *et al.* would be good candidates as actuation systems if embedded within the soft robots, as they require only an external magnetic field to induce motion.

In addition to the central interest that this study raises concerning the fundamental physical origin of this kind of ferromagnetism, the work of Liu *et al.* should stimulate new research in materials science. Indeed, materials could be envisioned that would exhibit unusual mechanical and magnetic properties. For example, by concentrating the ferromagnetic droplets into very dense suspensions, it could be possible to synthesize porous ferromagnetic materials (*9*) such as magnetic sponges. It may also be possible to create elastic ferromagnetic polymer films, or even ferromagnetic bicontinuous emulsions (bijels) (*10*). The work of Liu *et al.* challenges the established physics that ferromagnetic magnets can only be made of hard materials, not only opening up new perspectives in direct applications of their material, but also enabling a cascade of innovations that have not yet been imagined. ■

### A magnet takes shape

Ferrofluids, suspensions of magnetic nanoparticles, stay magnetized only in a magnetic field. Liu *et al*. jammed these nanoparticles at droplet surfaces to create permanent ferromagnets that could be reshaped.




GRAPHIC: VERONICA FALCONIERI/*SCIENCE*

*Complex Assemblies of Soft Matter Laboratory (COMPASS), UMI 3254, CNRS-Solvay-University of Pennsylvania, RIC, Bristol, PA 19007, USA. Email: rdreyfus@sas.upenn.edu*

#### REFERENCES AND NOTES

1. C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, ed. 6, 1986).
2. X. Liu *et al.*, *Science* **365**, 264 (2019).
3. J. Popplewell, L. Sakhnini, *J. Magn. Magn. Mater.* **149**, 72 (1995).
4. S. U. Pickering, *J. Chem. Soc. Trans.* **91**, 2001 (1907).
5. M. Pera-Titus, L. Leclercq, J.-M. Clacens, F. De Campo, V. Nardello-Rataj, *Angew. Chem. Int. Ed.* **54**, 2006 (2015).
6. A. D. Dinsmore *et al.*, *Science* **298**, 1006 (2002).
7. C. A. S. Burel *et al.*, *Small* **13**, 1701925 (2017).
8. D. Trivedi, C. D. Rahn, W. M. Kier, I. D. Walker, *Appl. Bionics Biomech.* **5**, 99 (2008).
9. A. Imhof, D. J. Pine, *Nature* **389**, 948 (1997).
10. M. E. Cates, P. S. Clegg, *Soft Matter* **4**, 2132 (2008).

10.1126/science.aax8979

MEDICINE

# Making precision medicine personal for cystic fibrosis

## Molecular defects in the cystic fibrosis gene prompt creative approaches to treatment

By **Candela Manfredi[1], Janice M. Tindall[2], Jeong S. Hong[1], Eric J. Sorscher[1]**

Cystic fibrosis (CF) is an inherited, life-threatening disease that primarily involves exocrine tissues (such as lungs, pancreas, and liver) for which highly active pharmacotherapies have recently emerged. More than 1700 disease-associated variants are described in the CF transmembrane conductance regulator (*CFTR*) gene, which encodes an epithelial cell ion channel that is defective in patients with CF. On the basis of classifying CFTR mutant proteins according to pathogenic mechanisms, the disease has been viewed as a model for personalized therapeutics. However, *CFTR* variants may have pleiotropic effects, which complicates assignment of specifically tailored drugs to discrete mechanistic subcategories. In addition, the cost of new CFTR modulators constrains third-party reimbursement and has delayed drug availability for certain patient groups, including individuals with ultrarare *CFTR* variants for which the treatments are not formally approved but may still be effective. Issues such as these are being addressed by innovative and powerful approaches to promote CF precision medicine.

Failure of the CFTR ion channel causes altered composition and volume of exocrine secretion, giving rise to thick, hyperviscous mucus that obstructs secretory organs including the lungs, pancreas, and liver, and diagnostic findings such as increased amount of chloride in sweat. The resulting inflammation, chronic infection, and fibrotic scarring of respiratory parenchyma represent the major causes of morbidity and mortality. For many years, a conceptual approach to CF intervention has pursued tailored small molecules (modulators) designed to rescue specific CFTR defects (*1*–*3*). These are grouped according to errors in CFTR protein synthesis (class I), maturation processing of the protein (class II), ion channel opening or gating (class III), conductance through the ion-selective pore (class IV), or steady-state protein concentrations (class V) (see the figure). A separate group (class VI) is sometimes used to specify class V *CFTR* variants that disrupt plasma membrane stability of the encoded protein (*4*).

Personalized treatment strategies based on this early annotation have led to impressive therapeutic progress. Ivacaftor, for example, is a "potentiator"-type modulator (a drug that helps open the CFTR ion channel gate). The compound is suitable for overcoming certain CFTR class III (gating) defects and improves lung function (*5*). The drug has gained U.S. Food and Drug Administration (FDA) approval for 38 *CFTR* variants (it is approved for fewer variants elsewhere), comprising ~15% of the patient population. In addition, ivacaftor in combination with lumacaftor (a "corrector" of decreased CFTR biogenesis) is marketed for individuals with two copies of the class II Phe508del CFTR protein maturation abnormality (~45% of patients; this is the most common CFTR mutation) (*2*). Tezacaftor, a corrector that functions similarly to lumacaftor, was approved in 2018 by the FDA in combination with ivacaftor to treat individuals with two copies of Phe508del, as well as those who carry one of 26 other ivacaftor-responsive CFTR mutations (*3*). These modulators, developed from studies of molecular pathogenesis and a personalized therapeutic strategy, can markedly improve respiratory manifestations of CF and have conferred new optimism worldwide among patients, families, and caregivers.

It was originally anticipated that well-established disease subclasses would serve as an organizing principle for precision CF treatments and help identify specific compounds for targeting CFTR mutants in a mechanism-directed manner. However, it has become increasingly clear that most *CFTR* variants result in not just one, but numerous subclasses of molecular defects in the CFTR protein, which makes personalized approaches complex. For example, the Phe508del mutant exhibits not only inadequate biogenesis (the traditional class II grouping), but also improper gating (class III) and increased plasma membrane turnover (class V) (*6*). Class I mutants, such as Glu831X (where X indicates premature stop), might be predicted to show no response to currently available modulators owing to *CFTR* messenger RNA (mRNA) instability and protein truncation. However, a fraction of Glu831X *CFTR* mRNA can produce full-length protein lacking only amino acid 831, which maintains residual activity (*7*). Accordingly, individuals with the Glu831X mutation are approved by the FDA to be treated with ivacaftor with or without tezacaftor. Numerous *CFTR* mRNA splicing defects (class V) generate proteins with large deletions or insertions and might otherwise be expected to exhibit negligible response to drugs such as ivacaftor. However, certain of these class V CFTR mutants produce alternatively spliced mRNA (e.g., 2789+5G→A, 3272-26A→G, 3849+10kbC→T) and reduced CFTR protein with residual function, and patients with these mutations are approved for ivacaftor treatment with or without tezacaftor. Thus, there is a need to recast some original assumptions that underlie CFTR mutation–tailored therapies for CF.

In agreement with the observation that *CFTR* variants are mechanistically pleiotropic, modulator drugs developed for a specific CFTR mutation or *CFTR* variant subcategory typically exhibit a broad spectrum of activity (*8*). For example, ivacaftor as single agent or in combination with tezacaftor leads to clinical benefit across all five mechanistic categories (see the figure). Similarly, emerging triple drug combination therapies (TCTs, e.g., ivacaftor in combination with tezacaftor and elexacaftor) have undergone extensive clinical testing (*9*). Elexacaftor appears to work through a Phe508del corrector mechanism independent from tezacaftor and enhances overall clinical effectiveness. TCTs have the potential to benefit a sizable majority of individuals with CF worldwide because the agents are directed toward patients with at least one allele encoding Phe508del CFTR.

Even if the promise of TCT is fully realized, thousands of patients will continue without effective modulators owing to refractoriness of the underlying mutant protein (e.g., those with untreatable mechanistic defects such as premature truncation, mRNA splicing defects, abnormal ion conductance, or aberrant protein folding). These patients highlight the continuing need to better understand intransigent CFTR pathophysiology and advance treatment of disease sequelae such as glandular obstruction by mucus (mucostasis), or respiratory infection and inflammation. Other individuals with CF have been unable to obtain modulator treatment because the mutations they possess are exceedingly rare.

[1]Emory University School of Medicine, Atlanta, GA, USA. [2]University of Alabama at Birmingham, Birmingham, AL USA. Email: esorscher@emory.edu

Recent estimates describe over 1000 *CFTR* variants represented by fewer than five patients each (*10*). Establishing processes so that individuals with poorly characterized or ultrarare *CFTR* variants can access effective modulator treatment remains one of the most vexing challenges in the field.

Many drugs approved by government regulatory agencies for one condition also show robust activity in other clinical settings. Despite lack of formal regulatory approval for these other indications, physicians are generally allowed flexibility to prescribe the treatments (with third-party insurance reimbursement) if patient benefit is anticipated. A difficulty arises in the case of CFTR modulators, however, because of high treatment cost (over $300,000 annually for certain regimens). Barriers to reimbursement have blocked modulator access in the United Kingdom and other countries, and drug expense can be a limiting economic burden for U.S. third-party payers, where insurance coverage for modulators has generally been restricted to patients with genotypes for which the drugs are formally approved. For example, variants with FDA approval comprise 39 CFTR mutations from among more than 1700 associated with the disease. Although many patients with ultrarare mutations would likely benefit from modulators already approved for other genotypes, access is constrained. Moreover, an attempt to broaden approval and include additional *CFTR* genotypes presents a formidable challenge. The number of patients with a particular ultrarare mutation is typically so small that a phase 3, double-blind placebo-controlled trial (often required by regulatory agencies) is not possible.

Innovative strategies have been devised to address the complex issue of drug authorization for ultrarare *CFTR* variants. For example, in part because pronounced clinical benefit from modulator treatment can often be obtained within a matter of weeks, N-of-1 trials have been evaluated. Such studies are designed so that individual patients with ultrarare mutations are monitored before, during, and after experimental treatment for evidence of improved disease manifestations. N-of-1 trials may be combined to encourage expanded drug approvals, but it can be difficult to show consistent modulator benefit in small cohorts, even those with identical *CFTR* genotypes. This is due to differences in, for example, age, past environmental exposures, chronic lung scarring, inflammation, severity of infection, and disease trajectory. As an alternative approach, patients with similar molecular phenotype (e.g., residual CFTR function determined by in vitro testing), or evidence of mild clinical disease (pancreatic sufficiency or sweat chloride value in an intermediate range), have been evaluated in a manner less bound by specific gene defect or mutation subcategory (*3*). For a cohort of patients with rare variants that can be meaningfully classified according to strong mechanistic rationale, a pronounced clinical response to CFTR modulators may be useful for broadening the regulatory indication for the entire group (*11*).

In recent, path-breaking decisions from the FDA, ivacaftor approvals were extended to comprise rare variants by applying in vitro data (*10*). The idea that compelling in vitro findings could contribute to modulator approval represents an innovative shift in CF personalized medicine. In vitro systems capable of predicting CF clinical improvement provide a compelling means to rationally expand drug access to patients with ultrarare genotypes. Well-validated cell models expressing recombinant CFTR mutants, CF intestinal organoids (cell cultures that form three-dimensional systems), and other cell-based strategies have been advanced for this purpose (*10*, *12*).

A single modulator or combination treatment can markedly improve disease phenotype for a large number of divergent CFTR molecular defects. A global potentiator of CFTR gating, for example, might favorably enhance ion transport across numerous *CFTR* variant mechanistic subcategories. Corrector agents that augment CFTR biogenesis typically confer improvement across multiple classes of variants. Based on the experience from CF, it is reasonable to imagine that similar drug versatility might be expected for other inherited conditions, including those with a considerable array of genetic abnormalities, such as adrenoleukodystrophy, certain muscular dystrophies, Pompe disease, etc. Specialized tools and leading-edge patient protocols are already being applied in the clinical setting toward the objective of making precision medicine more personal, and less strictly focused on subclasses of disease. Thus, emerging tailored therapies will be refined (for CF, cancer, inflammatory disorders, neurodegenerative conditions, and others) with increasingly informative data directed toward knowing what works best rather than relying on genotype alone. ■

## Disease-causing mutation subclasses in cystic fibrosis

There are over 1700 cystic fibrosis (CF) transmembrane conductance regulator (CFTR) mutations, which are classically divided into five categories based on pathogenic mechanism. Thirty-nine CFTR mutations are approved by the U.S. Food and Drug Administration for modulator treatment. The proportion of patients with CF and at least one mutation in a subclass and the proportion of those patients who can receive modulators are indicated. Figures are based on publicly available data (*13*).

| Class | I | II | III | IV | V |
|---|---|---|---|---|---|
| Type of mutation | Protein synthesis | Maturation processing | Ion channel gating | Ion channel conductance | Reduced protein |
| Example of modulator-approved mutation | Glu831X | Phe508del/ Phe508del | Gly551Asp | Arg117His | 3849+ 10kbC→T |
| Representative cellular compartment where defect occurs | Ribosome | Golgi/ER | Plasma membrane CFTR | | Plasma membrane, spliceosome |
| Patients with mutation class* | 22% | 88% | 6% | 6% | 5% |
| Patients with a modulator-approved genotype | <0.5% | 39.2% | 4.6% | 2.6% | 3.5% |

*Patients with heterozygous *CFTR* variants in two classes are counted twice.

GRAPHIC: N. CARY/*SCIENCE*

### REFERENCES AND NOTES


1. S. M. Rowe *et al.*, *N. Engl. J. Med.* **352**, 1992 (2005).
2. C. E. Wainwright *et al.*; TRAFFIC Study Group; TRANSPORT Study Group, *N. Engl. J. Med.* **373**, 220 (2015).
3. S. M. Rowe *et al.*, *N. Engl. J. Med.* **377**, 2024 (2017).
4. M. Haardt *et al.*, *J. Biol. Chem.* **274**, 21873 (1999).
5. F. J. Accurso *et al.*, *N. Engl. J. Med.* **363**, 1991 (2010).
6. G. Veit *et al.*, *Mol. Biol. Cell* **27**, 424 (2016).
7. A. Hinzpeter *et al.*, *PLOS Genet.* **6**, e1001153 (2010).
8. S. T. Han *et al.*, *JCI Insight* **3**, e121159 (2018).
9. D. Keating *et al.*; VX16-445-001 Study Group, *N. Engl. J. Med.* **379**, 1612 (2018).
10. A. G. Durmowicz *et al.*, *Ann. Am. Thorac. Soc.* **15**, 1 (2018).
11. R. N. Schuck *et al.*, *Clin. Pharmacol. Ther.* **104**, 282 (2018).
12. G. Berkers *et al.*, *Cell Reports* **26**, 1701 (2019).
13. The Clinical and Functional Translation of CFTR (CFTR2), available at https://cftr2.org/.



### ACKNOWLEDGMENTS

We thank M. Konomos for help with the figure and H. Evans and B. Sorscher for useful suggestions. Funding was obtained through grants from NIH (R01HL139876, R01HL136414) and CFF (SORSCH13XX0). E.J.S. serves as a board member, medical advisory council chair, and consultant for the CF Foundation.

10.1126/science.aaw0553

DEVELOPMENTAL BIOLOGY

# Tongues untied

## Fossil evidence reveals how jaws, ears, and tongues evolved in mammals and their relatives

*By* **Simone Hoffmann**[1] *and* **David W. Krause**[2]

The early embryos of mammals and other vertebrates typically have six pharyngeal arches, tissue bands under the early brain that develop into structures of the head and neck. The first arch gives rise to the mammalian malleus and incus (middle ear bones), mandible (part of the lower jaw), and tympanic bone (which supports the ear drum); the second, the stapes (middle ear bone) and part of the hyoid bone (a horseshoe-shaped structure in the neck); and the third, the remainder of the hyoid. Although the evolutionary transition of the first pharyngeal arch is well documented by fossil evidence, those of the second and third arches have received little attention in the developmental and paleontological literature. On page 276 of this issue, Zhou *et al.* (*1*) report on a newly discovered 165-million-year-old fossil from China in which the bones of the first three pharyngeal arches are preserved. The fossil defines a new taxon named *Microdocodon*.

The separation of the middle ear bones from the lower jaw is one of the more astonishing examples of convergent evolution in mammaliaforms—mammals and their closest fossil relatives. This transition has occurred independently at least three times (see the figure). The tiny incus and malleus bones that transmit sound in today's mammals were once large and attached to the lower jaw, and aided in jaw movement. This complex morphological transition is chronicled by fossil evidence (*2*–*5*), and the developmental and genetic processes that drove this transition are well characterized (*6*–*9*).

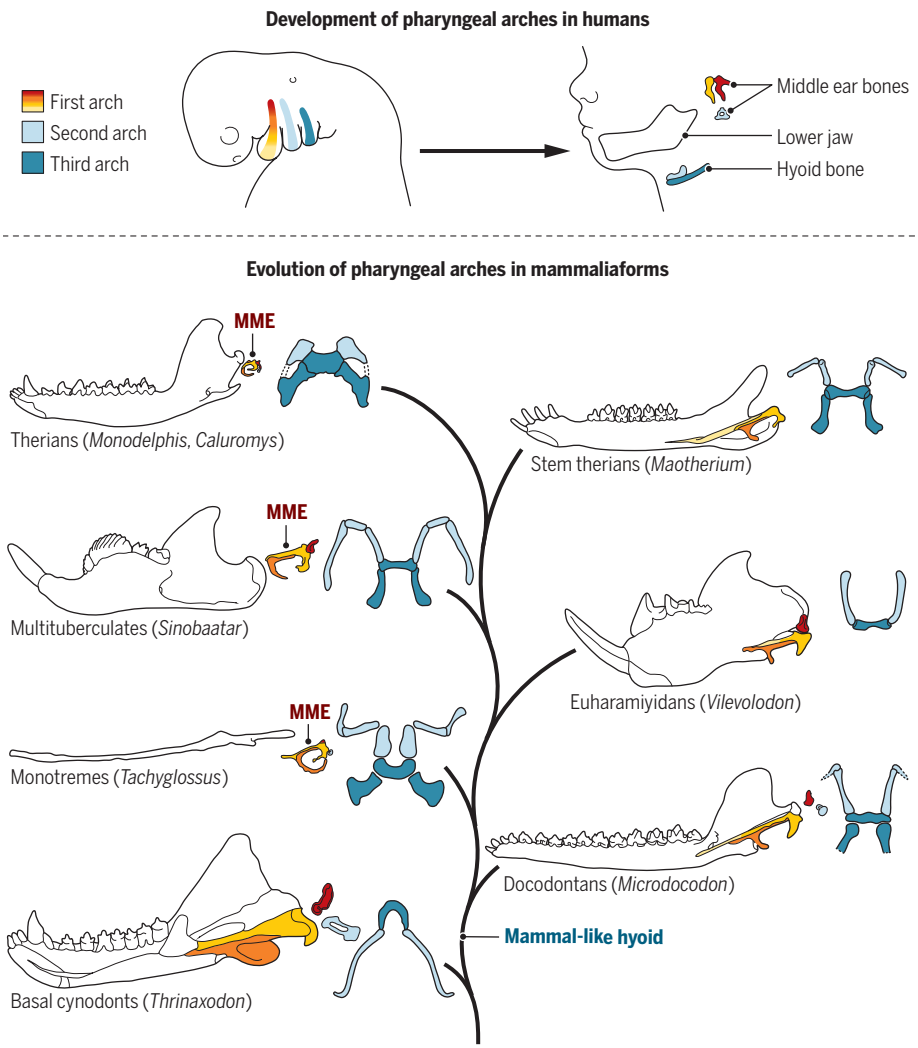
*Microdocodon* belongs to an early mammaliaform offshoot called docodontans, an extinct group widely distributed on northern continents in the Middle Jurassic through Early Cretaceous epochs, 174 to 100 million years ago (*10*). Like reptiles, the middle ear bones of docodontans were fully attached to the lower jaw (*1*, *10*). However, *Microdocodon* has a mammal-like hyoid. The nearly complete preservation in *Microdocodon* allowed Zhou *et al.* to identify poorly characterized hyoid elements in other fossil mammaliaforms, all of which exhibit a mammal-like pattern (see the figure).

The findings of Zhou *et al.* are fascinating from a developmental and functional perspective. In mammals, the hyoid is positioned at the base of the tongue and is crucial for tongue movements used in suckling or in swallowing chewed food. As a result, the hyoid of most mammals has a relatively complex configuration, with several mobile joints and large attachments for tongue and throat muscles (*11*). In contrast, the hyoid in reptiles is simple and has limited utility in pushing unchewed food through a relatively nonmuscular throat (*1*). *Microdocodon* demonstrates that the transformation of the second and third arches to a mammal-like hyoid occurred early in mammaliaform evolution and likely allowed for a more complex intake of food, enabled by a highly mobile tongue. Zhou *et al.* note that the mammal-like hyoid

[1]*Department of Anatomy, New York Institute of Technology College of Osteopathic Medicine, Old Westbury, NY, USA.*
[2]*Department of Earth Sciences, Denver Museum of Nature & Science, Denver, CO, USA. Email: shoffm04@nyit.edu*

### Development and evolution of pharyngeal arches

Transformation of the first pharyngeal arch into the mandible and mammalian middle ear (MME) bones occurred multiple times in mammaliaform evolutionary history, whereas development of the second and third pharyngeal arches into the hyoid bone occurred early in mammaliaform evolution and possibly only once (*1*).




GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

morphology might have coevolved with complex teeth that were able to masticate food, possibly as early as in the Late Triassic epoch (205 million years ago), as seen in fossils of the early mammaliaform *Morganucodon*.

Evidence from the newly discovered fossil further suggests that formation of the second and third arches into a mammal-like hyoid predates the ontogenetic separation of the middle ear from the lower jaw. Embryological growth of selected modern mammals supports the independent development of arches. In marsupials and monotremes, the hyoid is fully functional and is used in suckling immediately after birth, whereas the separation of the middle ear from the lower jaw does not occur until much later, after day 20 (*6*–*9*, *12*). The pattern holds true in placental mice as well; the hyoid is fully functional at birth, but the separation of the middle ear from the lower jaw starts at postnatal day 2, earlier than in monotremes and marsupials. Both fossil and developmental evidence suggest that mammal-like chewing and swallowing, with relatively complex tongue and hyoid movement, preceded the evolution of a fully functional middle ear.

The mammal-like hyoid appears to have evolved early and only once, whereas the separation of the middle ear bones from the lower jaw occurred multiple times convergently and relatively late in mammaliaform evolution (see the figure). The transformation of the second and third pharyngeal arches in mammaliaforms is decoupled and independent from the morphological changes of the first arch. Zhou *et al.* have opened the door to more thorough investigations of the interdependence or independence of pharyngeal arch evolutionary development. ■


### REFERENCES AND NOTES

1. C.-F. Zhou, B.-A. S. Bhullar, A. I. Neander, T. Martin, Z.-X. Luo, *Science* **365**, 276 (2019).
2. T. H. Rich, J. A. Hopson, A. M. Musser, T. F. Flannery, P. Vickers-Rich, *Science* **307**, 910 (2005).
3. Q. Ji, Z.-X. Luo, X. Zhang, C.-X. Yuan, L. Xu, *Science* **326**, 278 (2009).
4. J. Meng, Y. Wang, C. Li, *Nature* **472**, 181 (2011).
5. Z.-X. Luo, J. A. Schultz, E. G. Ekdale, in *Evolution of the Vertebrate Ear—Evidence from the Fossil Record*, J. A. Clack, R. R. Fay, A. N. Popper, Eds. (Springer, 2016), p. 139.
6. T. Rowe, *Science* **273**, 651 (1996).
7. N. Anthwal, L. Joshi, A. S. Tucker, *J. Anat.* **222**, 147 (2013).
8. N. Anthwal, D. J. Urban, Z.-X. Luo, K. E. Sears, A. S. Tucker, *Nat. Ecol. Evol.* **1**, 0093 (2017).
9. D. J. Urban *et al.*, *Proc. R. Soc. B* **284**, 20162416 (2017).
10. T. Martin, in *Mammalian Evolution, Diversity and Systematics*, Z. Frank, R. Asher, Eds. (de Gruyter, 2018), p. 199.
11. K. M. Hiiemae, A. W. Crompton, in *Functional Vertebrate Morphology*, M. Hildebrand, D. M. Bramble, K. F. Liem, D. B. Wake, Eds. (Belknap, 1985), p. 262.
12. U. Zeller, *Abh. Senck. Naturf. Ges.* **545**, 1 (1989).



### ACKNOWLEDGMENTS

We thank S. Beery for help with the figure. Supported by NSF grant EAR-1664432.

10.1126/science.aay2061


NEURODEGENERATION

# The vascular side of Alzheimer's disease

## Protein aggregates restrict cerebral blood flow, which causes neural injury

*By* **Arthur Liesz**

Alzheimer's disease (AD) is largely considered to be caused by aberrant amyloid β (Aβ) protein accumulation in the brain. Despite the current focus on Aβ (*1*), multiple lines of evidence indicate an important vascular contribution to AD. Aβ accumulation induces neurovascular dysfunction, which results in morphological as well as functional changes of the microvasculature, such as an impaired reactivity of capillaries in response to neuronal activity. Yet the mechanisms of these associations are largely unclear. On page 250 of this issue, Nortley *et al.* (*2*) demonstrate that pericytes—contractile cells that wrap around the endothelial cells of capillaries—link Aβ to vascular dysfunction in AD. This suggests pericyte-driven mechanisms as new avenues of therapeutic approaches for AD.

During the lifetime of Alois Alzheimer in the early 20th century, dementias were primarily perceived as a vascular dysfunction; this was because of the prevalence of focal behavioral symptoms (e.g., speech deficits) in dementia patients and the evidence for vascular pathologies in these patients. However, the vascular contribution to dementias is currently considered controversial, owing to the prevailing idea of "Alzheimerization" (*3*), whereby the deposition of Aβ (rather than vascular dysfunction) is predominantly attributed as the proposed molecular basis for neurodegeneration and cognitive impairment in AD.

By contrast, accumulating evidence highlights the relevance of vascular dysfunction in AD: Genetic studies have identified several cardiovascular risk genes that substantially increase the risk for AD (*4*); neuropathological studies show that up to 80% of diagnosed AD patients have vascular pathologies such as microinfarcts and atherosclerosis of cerebral arteries (*5*, *6*); and Aβ-driven animal models also show morphological changes at the microvascular level (*7*). Besides these morphological changes in the cerebral vasculature, AD is also associated with functional impairment: Cerebral blood flow and cerebrovascular reactivity (the response to a vasodilatory stimulus) are reduced (*7*). Moreover, cardiovascular risk factors (CVRFs) are associated with increased Aβ in the brain; indeed, the therapeutic control of CVRFs is currently the most efficient available therapy for mild cognitive impairment and AD (*8*, *9*).

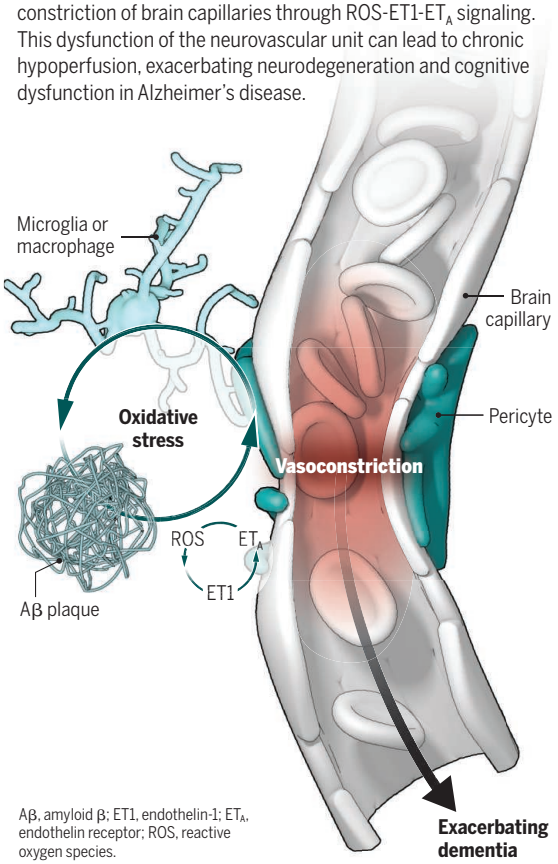
These findings suggest a pathophysiological concept in which Aβ accumulation and injury to the neurovascular unit (NVU; the functional and anatomical unit composed of the endothelium with the surrounding pericytes, smooth muscle cells, glia, and neurons) are cofactors in AD. NVU injury—independent of Aβ toxicity—has multiple causes, such as hypertension, diabetes, and atherosclerosis, which affect the cerebrovascular system at different sites ranging from large-artery changes to microvascular dysfunction. In this two-hit model of AD, Aβ and NVU injury each represent independent predisposing factors for vascular and cognitive dysfunction, respectively (*10*). This model provides an elegant explanation for the observation that not all individuals with Aβ accumulation develop clinical symptoms—that is, because vascular dysfunction might also be required (*6*). Neurovascular dysfunction and the resulting chronic hypoperfusion can have deleterious effects on neuronal homeostasis leading ultimately to neuronal cell death. Moreover, acknowledging the multifactorial etiology of AD poses an opportunity for extending criteria to improve individualized diagnosis and develop future therapies beyond Aβ-targeted strategies by including, for example, imaging-based biomarkers of morphological (e.g., microinfarcts) or functional (e.g., vascular reactivity) vascular impairment in the diagnostic workup.

Pericytes play a major role in cerebral blood flow regulation because the pericyte-wrapped capillaries have the highest resistance in the cerebrovascular system. Hence, targeting pericyte function could have a substantial impact on brain perfusion and metabolism, and thereby may reduce NVU injury and potentially AD symptoms. Nortley *et al.* identify oxidative stress—mediated by Aβ-induced increase of reactive oxygen species (ROS)—as a key molecular pathway


*Institute for Stroke and Dementia Research, LMU Munich, 81377 Munich, Germany. Email: arthur.liesz@med.uni-muenchen.de*

## Pericytes constrict capillaries

Oxidative stress induced by Aβ toxicity causes pericyte constriction of brain capillaries through ROS-ET1-$ET_A$ signaling. This dysfunction of the neurovascular unit can lead to chronic hypoperfusion, exacerbating neurodegeneration and cognitive dysfunction in Alzheimer's disease.




Aβ, amyloid β; ET1, endothelin-1; $ET_A$, endothelin receptor; ROS, reactive oxygen species.

leading to pericyte constriction (see the figure). ROS induced the release of endothelin-1 (ET1), a molecule that induces pericyte constriction via $ET_A$ receptors to cause vascular dysfunction in AD (*2*). Oxidative stress has been consistently described as a hallmark pathophysiological cascade contributing to AD and mild cognitive impairment. ROS are druggable targets, and a large body of evidence demonstrates neuroprotective effects of ROS inhibition in AD (*11*). It is very likely that Aβ-mediated oxidative stress has pleiotropic effects on the NVU, including effects on endothelial and innate immune cells at the NVU. These immune cells (such as microglia and perivascular macrophages) are the main producers of ROS and have previously been implicated in the pathogenesis of vascular dysregulation in NVU injury (*12*), although their communication with pericytes in AD and their involvement in NVU injury is largely unclear.

Beyond the direct interaction of Aβ with NVU function, additional vascular mechanisms have been described in AD animal models that could contribute to NVU dysfunction, Aβ burden, and neuronal stress. Amyloid drainage via interstitial flow and meningeal lymphatics is reduced in AD (*13*, *14*). Impaired drainage could lead to substantially increased local Aβ concentrations in the perivascular space in late-stage AD. The compartmentalized interactions of Aβ deposition within the impaired NVU as both a cause and a consequence of locally increased Aβ burden are still insufficiently understood. This will require further detailed analysis in preclinical AD models to elucidate the triad of the immune-vascular-Aβ function at the NVU.

Nortley *et al.* studied the impact of exogenous Aβ in living human brain tissue obtained during neuro-oncological biopsies. In this model, Aβ induced pericyte constriction, leading to reduced capillary diameters. Interestingly, Aβ load in human biopsies correlated with the reduction of vascular diameter close to pericytes, suggesting a causal function of Aβ burden in vasoconstriction. However, these findings were obtained mainly in ex vivo cultures and in few patient samples. Recent clinical trials aiming to directly reduce Aβ burden using Aβ monoclonal antibodies would have been ideal scenarios to directly test the impact of Aβ load on vascular dysfunction in living AD patients. However, none of these trials included vascular dysfunction as an outcome parameter.

The failure of Aβ antibody therapies to improve cognitive outcome in large clinical trials might indicate that the reduction in Aβ per se might not be the key or perhaps not even a direct cause of neurodegeneration and dementia (*15*). Might vascular dysfunction indeed be an intermediate pathophysiological process between Aβ accumulation and neurodegeneration? It is now time to reassess vascular dysfunction as a relevant therapeutic target for combination therapies to improve the treatment of AD. ■

### REFERENCES AND NOTES


1. M. D. Sweeney *et al.*, *Alzheimer's Dement.* **15**, 158 (2019).
2. R. Nortley *et al.*, *Science* **365**, eaav9518 (2019).
3. D. J. Libon *et al.*, *Cogn. Behav. Neurol.* **19**, 112 (2006).
4. I. J. Broce *et al.*, *Acta Neuropathol.* **137**, 209 (2019).
5. J. B. Toledo *et al.*, *Brain* **136**, 2697 (2013).
6. M. C. Power *et al.*, *Ann. Neurol.* **84**, 10 (2018).
7. C. Iadecola, *Acta Neuropathol.* **120**, 287 (2010).
8. J. S. Rabin *et al.*, *JAMA Neurol.* **75**, 1124 (2018).
9. SPRINT MIND Investigators, *JAMA* **321**, 553 (2019).
10. B. V. Zlokovic, *Nat. Rev. Neurosci.* **12**, 723 (2011).
11. M. Dumont, M. F. Beal, *Free Radic. Biol. Med.* **51**, 1014 (2011).
12. G. Faraco *et al.*, *J. Clin. Invest.* **126**, 4674 (2016).
13. K. G. Mawuenyega *et al.*, *Science* **330**, 1774 (2010).
14. S. Da Mesquita *et al.*, *Nature* **560**, 185 (2018).
15. D. J. Selkoe, *Nat. Rev. Neurol.* 10.1038/s41582-019-0205-1 (2019).


### ACKNOWLEDGMENTS

Supported by the Vascular Dementia Research Foundation and the Munich Cluster for Systems Neurology (EXC 2145 SyNergy).

10.1126/science.aay2720

NEUROSCIENCE

# *Stress signaling in pain control*

## Endoplasmic reticulum stress can drive chronic pain, revealing a new target

*By* **Tony Avril**[1,2] *and* **Eric Chevet**[1,2]

The control of chronic pain has become a major public health concern because frequent prescriptions for opioids to control pain leads to increasing risks for abuse and overdose-related deaths. Therefore, in order to develop new therapeutic approaches that are as efficient as current treatments but with reduced adverse effects, a better understanding of the underlying cellular and molecular mechanisms involved in chronic pain is required. On page 248 of this issue, Chopra *et al.* (*1*) demonstrate that signaling by inositol-requiring enzyme 1α (IRE1α)–X-box binding protein spliced (XBP1s), which is induced by endoplasmic reticulum (ER) stress, enhances prostaglandin synthesis and release by myeloid cells, contributing to chronic pain in mouse models. This study reveals a new therapeutic approach for attenuating chronic pain behavior.

Chronic pain is a major source of suffering and disability, occurring after an initial injury. Interactions between the nervous and immune systems regulate the development of chronic pain (*2*). Indeed, it has become evident that these systems communicate closely to protect tissues and nerves from injury. Immune cells release extracellular factors that in turn can activate nociceptive neurons and the neural pain pathway. As such, in response to injury, macrophages (which are sentinel myeloid cells present throughout the body), and monocytes (which are blood-borne myeloid cells), are prominently recruited to inflammatory sites during tissue injury. A role for macrophages and monocytes in chronic pain has been extensively demonstrated (*3*); however, the precise molecular mechanisms by which these cells exert their functions remain to be fully characterized. For

[1]*Inserm U1242, University of Rennes, Rennes, France.* [2]*Centre de Lutte Contre le Cancer Eugène Marquis, Rennes, France. Email: eric.chevet@inserm.fr*

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

example, how these cells produce inflammatory cytokines, growth factors, and lipids that can act directly on nociceptor neurons to transduce electrical signals to the central nervous system that are perceived as pain is not fully understood.

A previously unknown role of IRE1α-XBP1s on the synthesis of the lipid prostaglandin E2 ($PGE_2$) is now provided by Chopra *et al.* $PGE_2$ is a well-known lipid mediator that contributes to inflammatory pain (*4*). Therefore, most of the nonsteroidal pain killers currently used, including aspirin and ibuprofen, inhibit enzymes involved in $PGE_2$ biosynthesis. However, their long-term effects and efficiency remain limited. In myeloid cells, $PGE_2$ is synthesized by microsomal prostaglandin E-synthase-1 (mPGES1), which is the terminal enzyme downstream of cyclo-oxygenase 2 (COX2), the other key enzyme involved in the inducible $PGE_2$ synthesis cascade (*5*). The enzymatic reactions leading to the production of $PGE_2$ occur at the perinuclear membrane and the ER.

The functionality and homeostasis of the ER is ensured by an adaptive signaling pathway called the unfolded protein response (UPR), which is activated when ER homeostasis is disrupted, known as ER stress (*6*). The UPR is mediated by three sensors: IRE1α, PRKR–like ER kinase (PERK), and activating transcription factor 6α (ATF6α) (*7*). The most conserved arm of the UPR is mediated by IRE1α, a dual kinase and ribonuclease that is resident in the ER membrane (*8*). IRE1α exerts its functions in the UPR mostly through its ribonuclease activity by either contributing to nonconventional splicing of the messenger RNA (mRNA) encoding the potent transcription factor XBP1s or by promoting degradation of select mRNAs and microRNAs (miRNAs) that contain a specific consensus sequence for cleavage, likely in a cell type–specific manner. This activity is called regulated IRE1α-dependent decay (RIDD) of RNA (*6*, *9*).

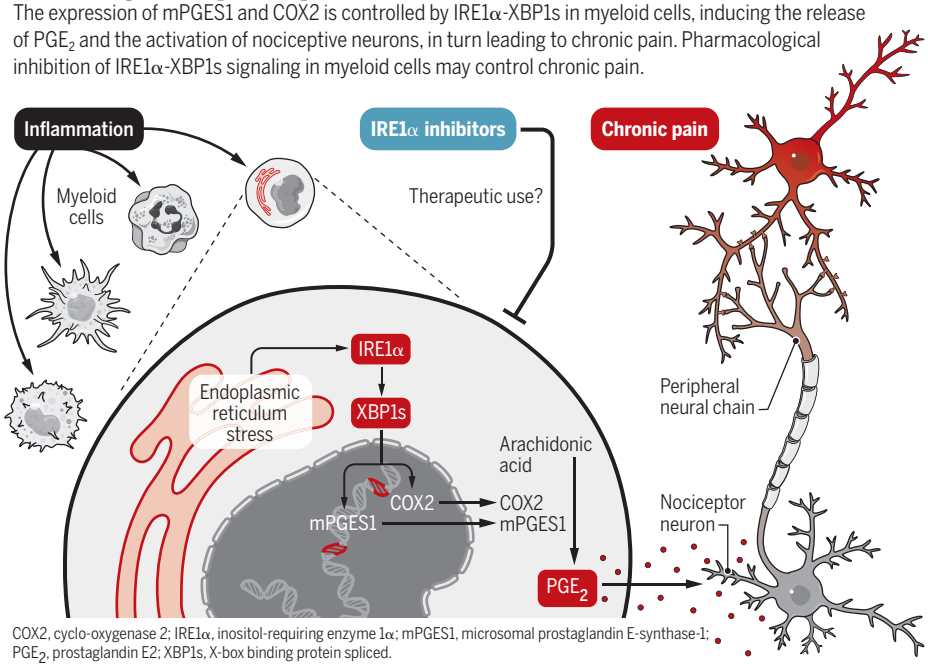
IRE1α signaling is also involved in many physiological processes beyond the UPR, including cell differentiation, lipid homeostasis, inflammation, and pathological processes such as diabetes, degenerative diseases, and cancer (*10*). Chopra *et al.* show that IRE1α-XBP1s signaling is induced by ER stress in myeloid cells after engagement and activation of Toll-like receptors (which sense microbial molecules) and regulates the expression of both COX2 and mPGES1 to control their production of $PGE_2$. Using $PGE_2$-dependent mouse models of pain, the authors demonstrate that both genetic and pharmacologic impairment of the IRE1α-XBP1s signaling pathway leads to an attenuation of the pain behavior (see the figure).

Some of the most used chronic pain treatments lead to addictive behaviors (*11*). For example, opioids (such as oxycodone, hydrocodone, codeine, and morphine) act through specific G protein–coupled receptors (GPCRs) expressed on neuronal cell membranes (such as from dorsal root ganglion neurons), to reduce signal transmission of nociceptive neurons mostly by attenuating neurotransmitter release (*12*). When administered at appropriate doses, opioids are effective not only at eliminating pain but preventing its recurrence in long-term recovery. However, long-term administration of these drugs can lead the way for opioid misuse, leading to addiction and overdose-related deaths. As such, pharmacological alternatives to control chronic pain could be an opportunity to reduce the burden of opioid dependency and fatality. Drugs that directly target the $PGE_2$ synthesis pathway are currently under development, such as the Celecoxib family of mPGES1 inhibitors, two of which are undergoing evaluation in clinical trials (*13*).

As proposed by Chopra *et al.*, targeting IRE1α-XBP1 might provide new therapeutic approaches to tackle chronic pain. In this context, the increasing number of IRE1α inhibitors [targeting either the kinase or the ribonuclease domain (*14*)] developed for the treatment of pathologies including neurodegenerative or metabolic diseases as well as cancer could represent appealing candidates for chronic pain control to be used alone or in combination or alternation with lower doses of opioids. Currently, it is difficult to define whether kinase or ribonuclease inhibitors would be more appropriate because in the study by Chopra *et al.*, both types of inhibitors led to similar pain attenuation in mice. Moreover, one might also anticipate that pharmacologic inhibition of IRE1α could reduce neurotransmitter production by nociceptive neurons by disturbing the homeostasis of the secretory pathway (which is regulated by the ER). Consequently, such inhibitors would not only reduce pain through preventing $PGE_2$ production by myeloid cells but also by reducing the activity of nociceptive neurons. The latter requires validation in relevant animal pain models. ■

## Chronic pain signaling

The expression of mPGES1 and COX2 is controlled by IRE1α-XBP1s in myeloid cells, inducing the release of $PGE_2$ and the activation of nociceptive neurons, in turn leading to chronic pain. Pharmacological inhibition of IRE1α-XBP1s signaling in myeloid cells may control chronic pain.




COX2, cyclo-oxygenase 2; IRE1α, inositol-requiring enzyme 1α; mPGES1, microsomal prostaglandin E-synthase-1; $PGE_2$, prostaglandin E2; XBP1s, X-box binding protein spliced.

GRAPHIC: A. KITTERMAN/*SCIENCE*

### REFERENCES AND NOTES


1. S. Chopra *et al.*, *Science* **365**, eaau6499 (2019).
2. Z. Hore, F. Denk, *Brain Behav. Immun.* 10.1016/j.bbi.2019.04.033 (2019).
3. R. R. Ji, A. Chamessian, Y. Q. Zhang, *Science* **354**, 572 (2016).
4. A. Kawabata, *Biol. Pharm. Bull.* **34**, 1170 (2011).
5. İ. Küçükgüzel, B. Bülbül, *Anticancer. Agents Med. Chem.* 10.2174/1871520619666190227174137 (2019).
6. Y. Kozutsumi, M. Segal, K. Normington, M. J. Gething, J. Sambrook, *Nature* **332**, 462 (1988).
7. C. Hetz, E. Chevet, S. A. Oakes, *Nat. Cell Biol.* **17**, 829 (2015).
8. M. Y. Wong *et al.*, *Curr. Top. Microbiol. Immunol.* **414**, 1 (2018).
9. M. Maurel, E. Chevet, J. Tavernier, S. Gerlo, *Trends Biochem. Sci.* **39**, 245 (2014).
10. A. Almanza *et al.*, *FEBS J.* **286**, 241 (2019).
11. N. Massaly, J. A. Morón, *Curr. Opin. Behav. Sci.* **26**, 69 (2019).
12. E. Darcq, B. L. Kieffer, *Nat. Rev. Neurosci.* **19**, 499 (2018).
13. A. Psarra, A. Nikolaou, M. G. Kokotou, D. Limnios, G. Kokotos, *Expert Opin. Ther. Pat.* **27**, 1047 (2017).
14. J. Obacz *et al.*, *Sci. Signal.* **10**, eaal2323 (2017).


### ACKNOWLEDGMENTS


T.A. and E.C. are funded by grants from ERANET ANR (ERAAT), Institut National du Cancer (INCa) and Fondation pour la Recherche Médicale (FRM équipe labelisée 2018).



10.1126/science.aay2721

BOOKS *et al.*

HISTORY OF SCIENCE

# *Apollo, in context*

## Two tomes paint vivid historical portraits of Apollo 11

*By* **Asif Siddiqi**

On the evening of 13 December 1972, about a quarter of a million miles from Earth, 38-year-old Eugene Cernan stepped off the surface of the Moon and clambered back into his fragile and temporary home, a four-legged landing vehicle named *Challenger*. The ungainly contraption was set down in Mare Serenitatis, a vast lava plain on the Moon, where Cernan and his colleague Harrison "Jack" Schmitt, a geologist, had spent the past 3 days collecting rocks. A day later, the two men lifted off from the Moon and headed back to Earth, collecting their third crewmember, Ron Evans, who had patiently waited in lunar orbit in the mother ship, *America*. So ended NASA's Apollo program, which having spent $25 billion (about $290 billion in today's money) put a dozen Americans on the Moon between 1969 and 1972.

Few technical achievements of the 20th century have been so mythologized in the popular imagination. Textbooks regularly mark the Moon landing as one of the greatest achievements of human civilization. Is there anything left to say about Apollo?

If the public perception of Apollo remains wedded to grand historical narratives about exploring the unknown, professional historians have, for the most part, resisted the triumphalist urge and instead interpreted Apollo as a manifestation of a Cold War battle between adversarial superpowers. In this reading, Neil Armstrong's first step on the Moon on the evening of 20 July 1969 was as much about the human spirit to explore as it was cold hard political gamesmanship.

John F. Kennedy's famous speech to Congress in 1961 calling on the nation to set a goal of "landing a man on the moon and returning him safely to the earth" has provided a convenient starting point for most histories of Apollo—and there have been many. In the decades since the landing, library shelves have filled with hundreds and hundreds of tomes on the topic. They have generally favored three basic frameworks, the most ubiquitous being the astronauts-centered history. Another school has focused on the high politics of Kennedy, Johnson, and Nixon. Last, a smaller canon has excavated the project from the perspective of managers and engineers.

Among literally dozens of new books on Apollo, all timed to coincide with the 50th anniversary of the Moon landing in July 2019, two books exemplify both old and new tropes. James Donovan's *Shoot for the Moon* has been touted by none other than Apollo 11 astronaut Michael Collins, who recently remarked, "This is the best book on Apollo that I have read." Donovan's strength is his breezy and journalistic writing style that weaves together a vast and complicated set of stories from all levels of Apollo.

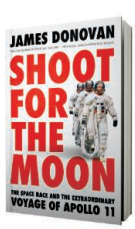
If Donovan's book is a by-the-numbers retelling that is also a familiar one, *Eight Years to the Moon* by Nancy Atkinson takes us into less-familiar aspects of the history, benefiting in many ways from a burst of recent scholarship. This book functions as a kind of corrective to the ubiquitous heroic narrative of politicians and astronauts.

The overall beats of Donovan's story will be familiar to many. For example, he captures the intimate world of Wernher von Braun, the visionary German rocket designer who was captured by Americans after World War II. Von Braun was brought to Texas and eventually to Alabama, where, in the 1960s, he directed the project to develop the giant Saturn V rocket that sent Apollo to the Moon.

Very succinctly, Donovan describes von Braun as "a handsome, charming ex-SS officer who had been the chief architect of an ambitious rocket program that had killed thousands during the war—and who now spread the gospel of space exploration to Americans in Walt Disney TV specials." What he leaves unsaid is as important: that more people died building the V2—as slave prisoners in Nazi concentration camps—than the rocket actually killed. Although Donovan glosses over this, von Braun's team's culpability for crimes against humanity is fundamental to any reckoning with the legacy of Apollo.

Donovan's narrative is a well-crafted one. It is one of the rarest of Apollo books that manages to weave together the political, the technical, and the heroic. We get vivid descriptions of managers struggling with difficult decisions, such as the directive that sent the first Apollo spacecraft around the Moon

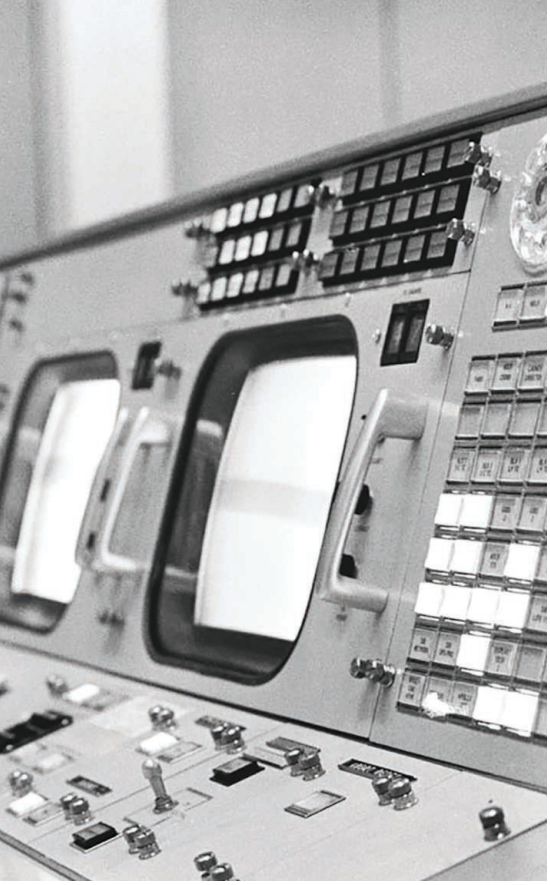

**Eight Years to the Moon**
*Nancy Atkinson*
Page Street Publishing, 2019. 240 pp.

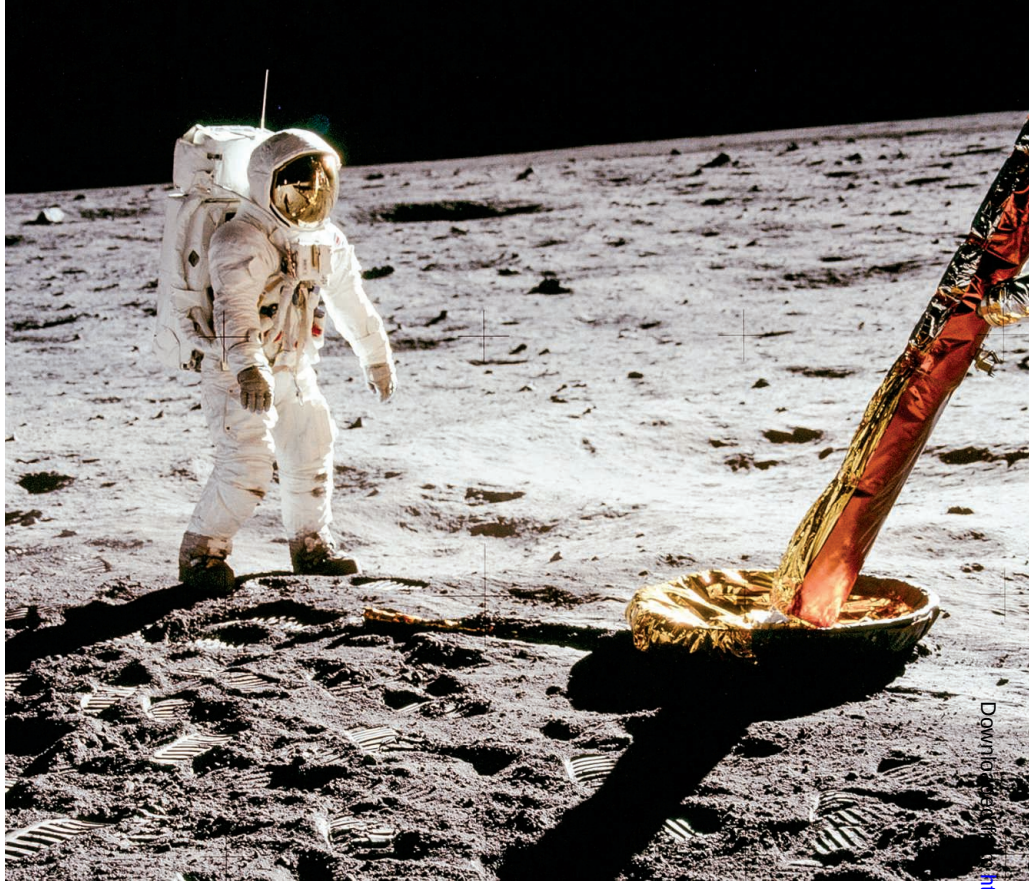



**Shoot for the Moon**
*James Donovan*
Little, Brown, 2019. 464 pp.

*The reviewer is at the Department of History, Fordham University, Bronx, NY 10458, USA. Email: siddiqi@fordham.edu*

PHOTO: NASA/REPRINTED WITH PERMISSION FROM *EIGHT YEARS TO THE MOON* BY NANCY ATKINSON, PAGE STREET PUBLISHING CO., 2019





Glynn Lunney participates in a simulation exercise at Mission Control in Houston, Texas (left). Buzz Aldrin stands near the lunar module on the Moon (right).

PHOTO: NASA

in late 1968, a recommendation that was one of the most risky in the entire program.

Donovan's account of this mission, Apollo 8, is one of the best in print, communicating the sheer hubris of sending three astronauts into deep space with only one chance—the foolproof firing of the Apollo command and service module's SPS engine—to get back to Earth. It is hard to think of a NASA mission that was more risky but even harder to imagine the Moon landing happening without it.

In Donovan's hands, the actual landing mission, Apollo 11, reads like a thriller. We know the ending, of course, but the contingencies of the mission—points at which the flight could have failed—are so many that their cumulative power communicates how miniscule the margins were. In a million possible scenarios, the *Eagle* lunar module lander crashes on the Moon with no way back for the astronauts. They did, of course, come back, and Donovan's narrative wraps up rather abruptly with a few cursory paragraphs about the postflight visits of the Apollo 11 astronauts to foreign countries. He concludes on a pithy note: "By the time the crew returned home, they understood that their lives would never be the same."

In *Eight Years to the Moon*, the astronauts and the technical details of the missions are still present, but we get insight into the tens of thousands of others who participated in and shaped Apollo. We hear, for example, about Ken Young, one of the first engineers hired at the Manned Spacecraft Center in Houston in the early 1960s. When he arrived, Young knew that he would be working on something to do with space but "didn't have a clue" exactly what his job would be. Stories such as this give the reader a sense of the team's commitment, its creativity, and often its confusion. When the Moon program was announced in 1961, many, such as rocket engineer Henry Pohl, were "dumbfounded" at the audacity of the challenge.

In weaving through the lives of engineers at NASA, Atkinson takes us to the drafting tables, engine test stands, smoky offices, and newly minted homes in Houston where the vast architecture of Apollo—its rockets, its spaceships, its tracking stations, its training centers, and its launch sites—was conceived. There are wonderfully vivid accounts of life in Houston, for example, where, lacking air conditioning in the sweltering heat, engineers' wives and kids would spend afternoons in bathtubs filled with cool water.

Atkinson is sensitive to those who were written out of the history of Apollo, particularly women. She integrates into the story the work of "computers"—women who made immeasurable contributions in ground support.

Atkinson's writing situates the story of the program's female African-American computers, whose story was documented in Margot Lee Shetterly's 2016 *Hidden Figures*, in the broader world of the technical development of Apollo. It would be an oversimplification to think of these women as functioning only as people who entered data, she argues. Some, including Dottie Lee, were involved in "designing and testing vehicles, not just running the numbers." Of course, this was largely a male-dominated and sexist engineering culture, but the role of women, both as technical support personnel and in unpaid roles as caregivers, cannot be overstated.

Like Donovan, Atkinson takes us into the heartbeat of the missions, but her narrative shows how the flights leading up to the landing were just the visible tip of a massive social, technical, and cultural enterprise.

Remarkably, her description of the Apollo 11 landing mission contains some previously unknown tidbits. She highlights, for example, the very dangerous uncontrolled reentry of the abandoned service module, which could have interfered with the reentry of Armstrong, Aldrin, and Collins as they sped toward Earth's atmosphere after completing the landing goal. Atkinson notes dryly that "somehow, the details and documentation of this anomaly were lost for nearly fifty years."

Based on a vast amount of original research and embellished with a suite of wonderful photographs, Atkinson's book is one of the best books published on Apollo. It is a surprisingly short work but a masterful one, communicating the enormity of the social and technical task that tens of thousands of Americans took on in the 1960s to achieve John F. Kennedy's goal. ■

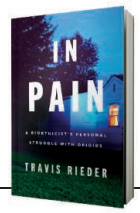
10.1126/science.aay4380

MEDICINE

# The ethics of opioids

## A bioethicist's harrowing encounter with prescription painkillers highlights systemic health care problems

**In Pain**
**A Bioethicist's Personal Struggle with Opioids**
*Travis Rieder*
HarperCollins, 2019.
314 pp.




*By* **Daniel Z. Buchman**

Imagine having your foot pinned between a moving van and a motorcycle and subsequently undergoing five surgeries to restore function to your mangled limb over a 1-month period. In order to endure the debilitating pain, you are prescribed high doses of opioid analgesics, only to discover that your care team has no exit strategy. Feeling abandoned by the medical establishment, you decide to subject yourself to a process of aggressive self-tapering and subsequently experience a gruesome withdrawal. Travis Rieder tells such a story—his story—in his new book, *In Pain: A Bioethicist's Personal Struggle with Opioids*.

*In Pain* situates Rieder's traumatic experience within the context of North America's "triple wave" opioid overdose crisis (*1*). As he explains, high prescribing volumes and aggressive marketing of opioids such as OxyContin initiated the crisis in the 1990s and 2000s. When their opioid prescriptions stopped or were scaled back, some patients turned to illicit sources. Most of the tens of thousands of overdose deaths that occurred in the United States and Canada within the past several years were due to a toxic street supply of heroin and fentanyl analogs.


Supervised injection sites could help mitigate the opioid crisis, argues Rieder.

Although opioid prescribing rates have fallen in recent years, the ethics of how and when these drugs are administered is still relevant. As Rieder shows, patients on long-term opioid therapy may be harmed when subjected to forced opioid tapering. And not prescribing an opioid when one is needed can be harmful as well.

Rieder's story exposes how the U.S. health care system is not designed to deal with chronicity and complexity. He argues that the errors in his care were systemic and offers a useful reframing: His physicians, he writes, were "part of a system that didn't help them properly identify or live up to their responsibilities." The difficulty of dealing with complicated pain cases is further intensified by additional challenges, such as physician reimbursement structures, financial conflicts of interest, and direct-to-consumer advertising of pharmaceuticals.

As an academic bioethicist, Rieder provides a philosophical lens to his lived experiences. It is this contribution where *In Pain* departs from other published narratives on pain and substance use, in which a discussion of the ethical issues tends to be implicit rather than at the forefront.

In "Pain and drugs," "Dependence and addiction," and "What doctors owe patients," Rieder discusses moral philosophy concepts such as first- and second-order desires, as well as duties and obligations, in digestible prose that is as profound as it is engaging. In chapter four, for example, he describes what it was like to experience the epistemic injustice of having the credibility of his testimony downgraded. His physician, it seems, thought his pain was not as severe as he claimed and believed that he was trying to secure drugs. This is a common experience among people who consume opioids (*2*).

Extending his analysis to the current overdose crisis, Rieder argues that the United States has a moral imperative to urgently adopt evidence-based harm reduction strategies (such as expanded access to naloxone and supervised consumption sites) that have been successful in other jurisdictions. These programs are based on values such as compassion and nonjudgment and have demonstrated effectiveness in the reduction of overdose deaths as well as infections such as HIV and hepatitis C. A harm reduction philosophy is counter to the "war on drugs" approach in the United States, which criminalizes people who use drugs and contributes to further harm. Given the scale of the crisis, Rieder submits that "It's time ... to abandon the idea that we will arrest our way out of drug use and addiction."

Despite his many privileges, Reider experienced tremendous suffering during his recovery and faced a number of challenges in accessing the care he needed. Those who are less privileged are often left far worse off.

*In Pain* is an invitation for compassion for the millions of people around the world suffering with pain and substance use disorders, but it is also a call to action, not just for more responsible opioid prescribing but for a fundamental change to how pain management is delivered. As Rieder argues, no advances in medical technology will solve the public health problem of pain. What he believes is required is a change in attitude toward the "pill for every pain" culture in North America and more sustained funding structures for evidence-based nonpharmacological treatments.

Pain patient advocacy groups and people on the front lines of the overdose crisis have been calling for action for decades. Rieder's thoughtful book joins their growing chorus. ■

*The reviewer is at the University Health Network, the University of Toronto Joint Centre for Bioethics, and the Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. Email: daniel.buchman@utoronto.ca*

### REFERENCES AND NOTES

1. D. Ciccarone, *Int. J. Drug Pol.* 10.1016/j.drugpo.2019.01.010 (2019).
2. D.Z. Buchman *et al.*, *Bioeth. Inq.* **4**, 1 (2017).

10.1126/science.aax8555

PHOTO: LAURENT VU THE/AFP/GETTY IMAGES

LETTERS

PHOTO: JOHN GOLLOP/ALAMY STOCK PHOTO

A boxwood shrub shows the first signs of blight.

*Edited by* **Jennifer Sills**

# *Fighting plant pathogens together*

Boxwood, a cornerstone species of American gardens, is currently threatened by a blight pathogen, *Calonectria pseudonaviculata*. The first blight epidemics in North America were seen in North Carolina and Connecticut in 2011 (*1*). The pathogen has since spread to 28 U.S. states (*2*) and several Canadian provinces (*3*). The disease also has severely affected boxwood gardens and forests in Europe and Asia (*3*). Efforts to contain the blight on boxwood must involve not only scientists, farmers, and retailers but also landscaping companies and home gardeners. As with any invasive pathogen, scientists must take every opportunity to engage laypeople who deal with these plants.

The blight pathogen spreads long distances primarily through the transport of infected plant stocks (*4*) and locally by contaminated animals, gardening tools, and stormwater (*2*). Thus, awareness campaigns to prevent the purchase of infected plant stocks are critical to protecting existing boxwood in a given garden, neighborhood, and region. When the pathogen is inadvertently introduced, all infected plants and planting materials must be properly contained in a timely manner to prevent pathogen local spread (*2*). This fight can be won only through community efforts in which consumers and homeowners are empowered with the knowledge required to identify and contain the disease, farmers and retailers work to produce and sell blight-free plants, and landscapers practice strict sanitation between job sites.

Boxwood blight containment can serve as a model of extensive stakeholder involvement in pursuing plant biosecurity. Science communities should work together with stakeholders, innovating disease mitigation technologies and developing a systems approach to saving the iconic landscape plant. A successful approach can be extended to other plants and crops.

**Chuanxue Hong**
Virginia Polytechnic Institute and State Univeristy, Virginia Beach, VA 23455, USA.
Email: chhong2@vt.edu

**REFERENCES AND NOTES**

1. K. L. Ivors *et al.*, *Plant Dis.* **96**, 1070 (2012).
2. C. X. Hong, *The Boxwood Bulletin, The Journal of the American Boxwood Society* **58**, 4 (2019).
3. N. LeBlanc, C. Salgado-Salazar, J. A. Crouch, *Appl. Microbiol. Biotechnol.* **102**, 4371 (2018).
4. T. Creswell, *Purdue Landscape Rep.* **18**, 18 (2018).

10.1126/science.aay4514

# *States lack endangered species reporting*

The U.S. Senate is currently considering a bill (S. 1051) that would cede some permitting decision authority to states relative to the U.S. Endangered Species Act (*1*). This bill is but one in a series of recent attempts to transfer authority for endangered species from federal to state agencies (*2*). The Endangered Species Act has long been criticized for its apparent inability to recover species (*3*). These efforts to give states more authority typically argue that recovery of endangered species would be more successful if states were granted more control. We find this argument problematic.

State endangered species programs do not, generally, report on recovery planning, expenditures, actions, outcomes, or status of endangered species. In contrast, federal agencies report on each of these issues and make the information readily accessible (*4*). If states are to assume more responsibility for endangered species in the United States, then transparency is critical. Under existing policies, the lack of public reporting by states results in a disquieting accountability gap. Without information, how can the public evaluate whether the gains that Congress envisions actually occur?

Information and accountability have been critical to the administration of, and the evaluation of, the federal application of the Endangered Species Act. This same transparency needs to be a component of any state Endangered Species Act responsibility. Before states take on tasks that are currently under federal jurisdiction, staffers, nongovernmental organizations, environmental lobbyists, and watchdog

groups such as the Center for Biological Diversity and Defenders of Wildlife should ask Congress to make sure that the relevant laws and regulations stipulate transparent reporting as a key part of the states' duties.

**Connor Bennett and Mark W. Schwartz**
Department of Environmental Science & Policy, University of California, Davis, Davis, CA 95616, USA.
*Corresponding author.
Email: mwschwartz@ucdavis.edu

#### REFERENCES AND NOTES

1. U.S. Senate Bill 1051, "A bill to amend the Endangered Species Act of 1973 to establish a program to allow States to assume certain Federal responsibilities under that Act with respect to agency actions applicable to highway projects within the States, and for other purposes" (2019); www.congress.gov/bill/116th-congress/senate-bill/1051/text.
2. Center for Biological Diversity, "Politics of extinction" (2019); www.biologicaldiversity.org/campaigns/esa_attacks/trumptable.html.
3. D. M. Evans *et al.*, "Species recovery in the United States: Increasing the effectiveness of the Endangered Species Act," *Issues in Ecology*, Rep. No. 20 (Ecological Society of America, 2017).
4. U.S. Fish and Wildlife Service, "Recovering threatened and endangered species" (www.fws.gov/southeast/endangered-species-act/recovery/).

10.1126/science.aay4403

# Replicable data for digit ratio differences

In his News Feature "The mismeasure of hands?" (7 June, p. 923), M. Leslie reports our finding that lesbians have, on average, a more masculine digit ratio than straight women (*1*) and immediately follows it by citing John Manning's BBC study finding no significant difference. This juxtaposition implies that our findings may not be replicable. However, a meta-analysis published almost a decade ago studied those two publications, in addition to—tellingly—14 more (*2*), and it concluded that the difference between lesbians and straight women was so strong that it would take 58 consecutive null findings to render the difference statistically nonsignificant.

I am aware of only one null finding since that time (in a woefully underpowered sample of 25 women) (*3*). Meanwhile, there have been two reports of monozygotic twins discordant for sexual orientation in which the lesbian twins have more masculine digit ratios than their sisters: one from Japan (*4*) and another from Europe (*5*), replicating an earlier U.S. report (*6*). I stand by our 2000 report and would be happy to hear any skeptics' speculations about what factor(s) might affect digit ratios in lesbians if the measures do not reflect prenatal androgen exposure.

**S. Marc Breedlove**
Neuroscience Program, Michigan State University, East Lansing, MI 48824, USA.
Email: breedsm@msu.edu

#### REFERENCES AND NOTES

1. T. J. Williams *et al.*, *Nature* **404**, 455 (2000).
2. T. Grimbos *et al.*, *Behav. Neurosci.* **124**, 278 (2010).
3. N. E. Charles, G. M. Alexander, *Arch. Sex. Behav.* **40**, 587 (2011).
4. K. Hiraishi, S. Sasaki, C. Shikishima, J. Ando, *Arch. Sex. Behav.* **41**, 711 (2012).
5. T. M. Watts, L. Holmes, J. Raines, S. Orbell, G. Rieger, *Arch. Sex. Behav.* **47**, 2435 (2018).
6. L. S. Hall, C. T. Love, *Arch. Sex. Behav.* **32**, 23 (2003).

10.1126/science.aay3385

## TECHNICAL COMMENT ABSTRACTS

### Comment on "Earth and Moon impact flux increased at the end of the Paleozoic"

**Stefan Hergarten, Gerwin Wulf, Thomas Kenkmann**

Mazrouei *et al.* (Reports, 18 January 2019, p. 253) found a nonuniform distribution of crater ages on Earth and the Moon, concluding that the impact flux increased about 290 million years ago. We show that the apparent increase on Earth can be explained by erosion, whereas that on the Moon may be an artifact of their calibration method.
Full text: dx.doi.org/10.1126/science.aaw7471

### Response to Comment on "Earth and Moon impact flux increased at the end of the Paleozoic"

**Sara Mazrouei, Rebecca R. Ghent, William F. Bottke, Alex H. Parker, Thomas M. Gernon**

Hergarten *et al.* interpret our results in terms of erosion and uncertain calibration, rather than requiring an increase in impact flux. Geologic constraints indicate low long-term erosion rates on stable cratons where most craters with diameters of ≥20 kilometers occur. We statistically test their proposed recalibration of the lunar crater ages and find that it is disfavored relative to our original calibration.
Full text: dx.doi.org/10.1126/science.aaw9895

## ERRATA

**Erratum for the Research Article "Aerosol-driven droplet concentrations dominate coverage and water of oceanic low-level clouds" by D. Rosenfeld *et al.*,** *Science* **364**, eaay4194 (2019). Published online 21 June 2019; 10.1126/science.aay4194

**Erratum for the Research Article "Opposing reactions in coenzyme A metabolism sensitize *Mycobacterium tuberculosis* to enzyme inhibition" by E. Ballinger *et al.*,** *Science* **364**, eaay3546 (2019). Published online 21 June 2019; 10.1126/science.aay3546

Cite as: S. Hergarten *et al.*, *Science* 10.1126/science.aaw7471 (2019).

# Comment on "Earth and Moon impact flux increased at the end of the Paleozoic"

**Stefan Hergarten*, Gerwin Wulf, Thomas Kenkmann**

Institut für Geo- und Umweltnaturwissenschaften, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany.

*Corresponding author. Email: stefan.hergarten@geologie.uni-freiburg.de

**Mazrouei *et al.* (Reports, 18 January 2019, p. 253) found a nonuniform distribution of crater ages on Earth and the Moon, concluding that the impact flux increased about 290 million years ago. We show that the apparent increase on Earth can be explained by erosion, whereas that on the Moon may be an artifact of their calibration method.**

Mazrouei *et al.* (*1*) analyzed the age distribution of lunar and terrestrial impact craters and found a nonuniform distribution. An apparent kink in the cumulative distributions at the same age of about 290 million years (Ma) led to the conclusion that the impact flux increased at this time.

Their analysis involves 38 terrestrial craters with diameters $D \geq 20$ km and requires that the crater inventory in this diameter range has not been affected by erosion. To support this hypothesis, they adopted a low estimate of long-term erosion rates (*2*). We argue that the vast majority of Earth's surface has been exposed to erosion rates much higher than the adopted 2.5 m $Ma^{-1}$. The record of the terrestrial craters with $D < 90$ km is more affected by erosion than by the age of the underlying crust (*3*).

Figure 1 illustrates the effect of erosion on the age-frequency distribution of the craters with $D \geq 20$ km exposed at the surface. The theoretical distribution was obtained by simulating $10^7$ hypothetical impacts following the most recent estimate of the terrestrial crater production rate (*4*). We assumed that crater production and erosion were constant over the entire age of Earth, and that the entire crust was as old as Earth. The spatial distribution of erosion rates was inferred from present-day erosion rates of 30 large drainage basins (*5*). Lifetimes of craters at a given erosion rate were estimated from their depth as a function of the diameter (*3*).

Although our simulation only predicts a statistical distribution and not absolute numbers of craters, it is visible from the shapes of the curves alone that constant crater production in combination with constant (over time, but spatially variable) erosion can reproduce the observed age-frequency distribution. We conclude that the curvature in the age-frequency distribution can be explained by degradation of the crater inventory by erosion and does not necessarily require an increase in impact flux.

The age distribution of the lunar craters is sensitive to the relationship between the rock abundance ($RA_{95/5}$) values and the crater ages. Mazrouei *et al.* assume a power-law relationship derived in a previous study (*6*) from a calibration dataset consisting of nine index craters of known ages.

Figure 2 shows the cumulative distribution of the $RA_{95/5}$ values from which ages were derived for 111 craters by Mazrouei *et al.* The distributions for diameter ranges $D \geq 10$ km and $D \geq 20$ km can be approximated by exponential distributions with almost the same decay constants ($\lambda = 59.7 \pm 5.7$ for $D \geq 10$ km, $\lambda = 58.3 \pm 12.1$ for $D \geq 20$ km, obtained from a maximum likelihood method).

This exponential distribution suggests an alternative model for the $RA_{95/5}$-age relation. If it was an exponential function of the form

$$\text{age} = \tau \exp\left(-\lambda RA_{95/5}\right) \quad (1)$$

with $\lambda$ found above and any arbitrary constant $\tau$, the $RA_{95/5}$ values would be consistent with a uniform age distribution, so there would be no requirement for any change in impact flux through time.

We fitted the nine index craters and found $\tau = 1213 \pm 330$ Ma. The resulting age distribution in both diameter ranges is shown in the inset of Fig. 2. Compared to figure 3 of Mazrouei *et al.*, the concave curvature indicating an increase in impact flux has vanished. The value of $\tau$ only defines the absolute time scale, not the shape of the curves.

It is unclear whether the nine index craters used for calibration (*6*) are sufficient to refute the exponential hypothesis. Our exponential fit yields $R^2 = 0.862$ and a root mean square deviation (RMSD) of 0.012 in the $RA_{95/5}$ values, while the power law suggested by Mazrouei *et al.* even has a slightly higher RMSD of 0.014. In the original data, the craters Jackson and Aristarchus differ in their $RA_{95/5}$ values by more than 0.01, and this difference is opposite to the trend

with age. Thus, the variation of the $RA_{95/5}$ values at given crater age should be larger than 0.01, and the RMSDs of both fits are in this order of magnitude.

There are two alternative models that are both consistent with the data, but the exponential model would be consistent with a constant impact flux. We do not claim that the exponential model is better than the power law suggested by Mazrouei *et al*., but given the uncertainties in the data, we are not convinced that the calibration dataset consisting of only nine craters can refute the exponential model.

In summary, we find that the apparent increase in the terrestrial impact flux is consistent with the removal of craters by erosion. The apparent increase in the lunar impact flux is very sensitive to the relationship between the values of $RA_{95/5}$ and the ages of craters, and other fitting functions are also consistent with the data under a constant impact flux.

## REFERENCES


1. S. Mazrouei, R. R. Ghent, W. F. Bottke, A. H. Parker, T. M. Gernon, Earth and Moon impact flux increased at the end of the Paleozoic. *Science* **363**, 253–257 (2019). doi:10.1126/science.aar4058 Medline
2. T. J. Blackburn, S. A. Bowring, J. T. Perron, K. H. Mahan, F. O. Dudas, K. R. Barnhart, An exhumation history of continents over billion-year time scales. *Science* **335**, 73–76 (2012). doi:10.1126/science.1213496 Medline
3. S. Hergarten, T. Kenkmann, The number of impact craters on Earth: Any room for further discoveries? *Earth Planet. Sci. Lett.* **425**, 187–192 (2015). doi:10.1016/j.epsl.2015.06.009
4. P. A. Bland, N. A. Artemieva, The rate of small impacts on Earth. *Meteorit. Planet. Sci.* **41**, 607–631 (2006). doi:10.1111/j.1945-5100.2006.tb00485.x
5. M. A. Summerfield, N. J. Hulton, Natural controls of fluvial denudation rates in major world drainage basins. *J. Geophys. Res.* **99**, 13871–13883 (1994). doi:10.1029/94JB00715
6. R. R. Ghent, P. O. Hayne, J. L. Bandfield, B. A. Campbell, C. C. Allen, L. M. Carter, D. A. Paige, Constraints on the recent rate of lunar ejecta breakdown and implications for crater ages. *Geology* **42**, 1059–1062 (2014). doi:10.1130/G35926.1


## ACKNOWLEDGMENTS


**Author contributions:** S.H. performed the calculations and wrote the paper; G.W. and T.K. contributed knowledge about terrestrial and lunar impact craters. **Competing interests:** The authors have no competing interests. **Data and materials availability:** The MATLAB codes for reproducing our results are provided in data S1.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/eaaw7471/suppl/DC1
Data S1


21 January 2019; resubmitted 12 February 2019
Accepted 11 June 2019
Published online 19 July 2019
10.1126/science.aaw7471

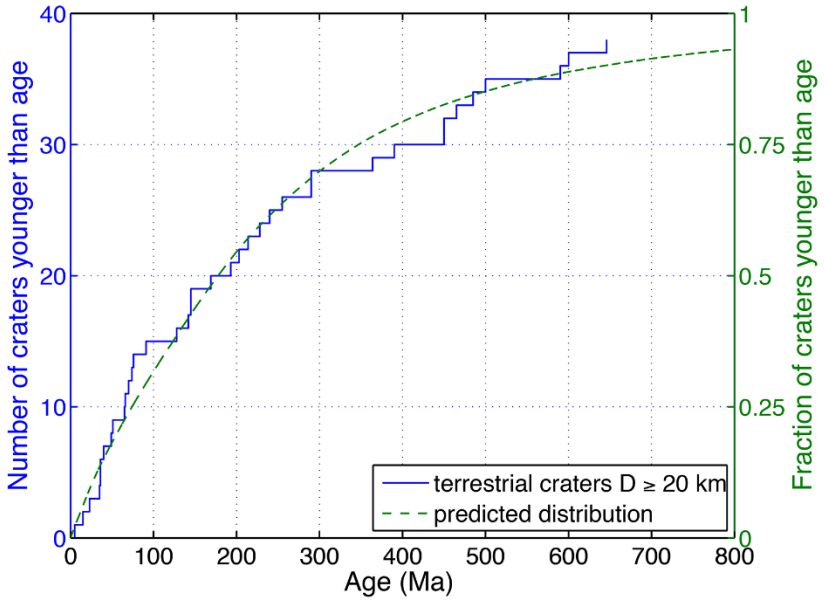



**Fig. 1. Age-frequency distribution of terrestrial craters with diameters $D \geq 20$ km.** The blue line (left axis) shows the ages of craters used in (*1*). Our predicted distribution (green line, right axis) is a simulation based on constant impact flux combined with erosion (see text).

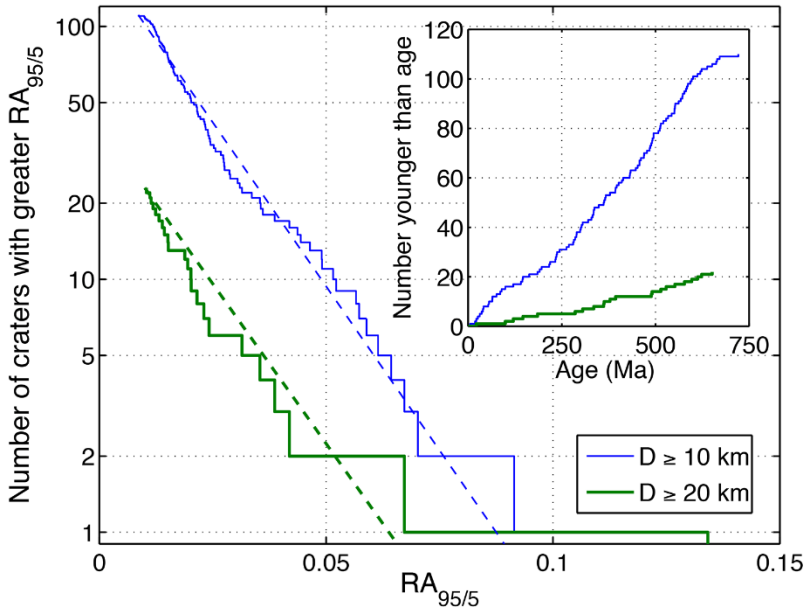



**Fig. 2. Cumulative distribution of the $RA_{95/5}$ values.** The distribution can be approximated by an exponential function (dashed lines). The inset shows the resulting age distribution if the $RA_{95/5}$-age relation follows Eq. 1 with $\lambda = 59.7$ and $\tau = 1213$ Ma.

Cite as: S. Mazrouei *et al.*, *Science* 10.1126/science.aaw9895 (2019).

# Response to Comment on "Earth and Moon impact flux increased at the end of the Paleozoic"

**Sara Mazrouei[1]*, Rebecca R. Ghent[1,2], William F. Bottke[3], Alex H. Parker[3], Thomas M. Gernon[4]**

[1]Department of Earth Sciences, University of Toronto, Toronto, ON, Canada. [2]Planetary Science Institute, Tucson, AZ, USA. [3]Southwest Research Institute, Boulder, CO, USA. [4]School of Ocean and Earth Science, University of Southampton, Southampton SO14 3ZH, UK.

*Corresponding author. Email: sara.mazrouei.seidani@mail.utoronto.ca

**Hergarten *et al.* interpret our results in terms of erosion and uncertain calibration, rather than requiring an increase in impact flux. Geologic constraints indicate low long-term erosion rates on stable cratons where most craters with diameters of ≥20 kilometers occur. We statistically test their proposed recalibration of the lunar crater ages and find that it is disfavored relative to our original calibration.**

Hergarten *et al.* (*1*) present an alternative interpretation of the impact record on Earth and the Moon. Their argument requires that Earth's cratons have been deeply eroded over the past 650 million years (Ma), removing many craters with diameters $D \geq 20$ km. The existing crater inventory has been taken to imply substantial global erosion rates of ~59 m $Ma^{-1}$ (*2*), more than 20 times cratonic rates (*3*).

Thermochronological data (*3*) provide consistent constraints on long-term cratonic erosion rates. Global mean erosion rates (*2*, *4*) cannot be applied to cratonic regions because such estimates include the high erosion rates prevalent in tectonically active regions today. Long-term integrated estimates do not preclude short-lived pulses of exhumation (*3*). Therefore, the application of present-day erosion rates over 100-million-year time scales is tentative and uncertain (*2*).

The global mean erosion rate (*2*) would imply 30 to 40 km of erosion on the cratons over the past 650 Ma. Such losses are inconsistent with geologic constraints (*5*), including the widespread cratonic occurrence of shallow (≤2 km) crustal features such as kimberlite pipes (*6*). Moreover, 85% of $D \geq 20$ km terrestrial craters are located on stable cratons (*6*), which have remained largely intact since the Proterozoic (*3*, *7*).

If the 20- to 30-km terrestrial craters have erosion lifetimes of 35 to 47 Ma (*2*) and erode at an exponential rate, the fraction of craters surviving longer than 290 Ma should only be 0.02 to 0.2% of the initial population. Instead, nearly one-third (5 of 16) have ages between 290 and 650 Ma (*3*).

We now address the alternative calibration of our lunar crater dating method (*1*). We argue that a goodness of fit based on $R^2$ and residual root mean square (RMS) scatter is not sufficient for model selection given the available information.

Our Bayesian model selection framework (*6*) can self-consistently accommodate uncertainty in the functional form proposed for the relationship between age and rock abundance, $RA_{95/5}$. There are parameter sets for a power-law relationship that provide crater age distributions consistent with an impact rate uniform in time, as there are for Hergarten *et al.*'s proposed exponential relationship. The issue is whether these parameter sets are likely given the ages of "index" craters, defined as $D \geq 20$ km craters whose ages were derived by independent means. Our analysis indicated that they are not (*6*), instead favoring parameter sets that include a change in the impact rate.

Using the same method (*6*), we performed a likelihood ratio test on the power-law and exponential models (Fig. 1). The relative likelihood of the exponential model to the power-law model is 0.004, disfavoring the former.

We can additionally include the choice of the power-law versus exponential models as part of the approximate Bayesian computation rejection (ABCr) analysis used to determine the Bayes factor for the broken impact rate versus the uniform impact rate (*6*). At the step of selecting parameters for the age-$RA_{95/5}$ from the posterior probability density function (PDF), we modified our method to select which model we are using for this trial, and choose each model with a probability proportional to the relative likelihood of each model as determined by the likelihood ratio test (that is, 99.6% of the cases selected to be the power-law distribution, and 0.4% of the cases selected to be the exponential distribution). If the evidence presented by the craters themselves outweighed the strong evidence against the exponential model, this treatment would allow that evidence to speak for itself.

Taking lunar craters alone, we still find a Bayes factor of 6:1 in favor of the broken impact rate over the uniform rate. Marginalizing over both the model of the age-$RA_{95/5}$ relationship and its parameters, we find evidence against a uni-

form impact rate. Including terrestrial craters has the same effect as before (*6*), producing a Bayes factor of 100:1 in favor of the broken impact rate over the uniform rate.

Finally, we examine whether a constant impact flux for large impactors is consistent with lunar data. Given that our age-$RA_{95/5}$ relationship has been questioned (*1*), here we only use the index crater Copernicus, whose age of 800 Ma was directly derived from Apollo 12 samples (*6*, *8*). We find that 20 lunar craters with $D \geq 20$ km have higher $RA_{95/5}$ values than Copernicus. Accordingly, within reasonable error, all should be < 800 Ma. The terrestrial production rate of $D \geq 20$ km craters over the past ~100 Ma is ~2.5 to ~3.0 × $10^{-15}$ $km^{-2}$ $year^{-1}$ [(*9*), used by (*1*, *2*); see also (*6*)]. When translated to the Moon (*6*) and kept constant for 800 Ma, it would yield 50 to 60 lunar craters—2.5 to 3 times the number observed. This high number of craters is also inconsistent with (i) the relative ages of large lunar craters as derived from optical maturity observations (*6*, *10*), and (ii) superposed crater counts on large lunar craters [i.e., many $D > 50$ km craters once considered Copernican-era are instead much older (*11*)].

## REFERENCES


1. S. Hergarten, G. Wulf, T. Kenkmann, Comment on "Earth and Moon impact flux increased at the end of the Paleozoic". *Science* **365**, eaaw7471 (2019).
2. S. Hergarten, T. Kenkmann, The number of impact craters on Earth: Any room for further discoveries? *Earth Planet. Sci. Lett.* **425**, 187–192 (2015). doi:10.1016/j.epsl.2015.06.009
3. T. J. Blackburn, S. A. Bowring, J. T. Perron, K. H. Mahan, F. O. Dudas, K. R. Barnhart, An exhumation history of continents over billion-year time scales. *Science* **335**, 73–76 (2012). doi:10.1126/science.1213496 Medline
4. M. A. Summerfield, N. J. Hulton, Natural controls of fluvial denudation rates in major world drainage basins. *J. Geophys. Res. Solid Earth* **99**, 13871–13883 (1994). doi:10.1029/94JB00715
5. C. B. Keller, J. M. Husson, R. N. Mitchell, W. F. Bottke, T. M. Gernon, P. Boehnke, E. A. Bell, N. L. Swanson-Hysell, S. E. Peters, Neoproterozoic glacial origin of the Great Unconformity. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1136–1145 (2019). doi:10.1073/pnas.1804350116 Medline
6. S. Mazrouei, R. R. Ghent, W. F. Bottke, A. H. Parker, T. M. Gernon, Earth and Moon impact flux increased at the end of the Paleozoic. *Science* **363**, 253–257 (2019). doi:10.1126/science.aar4058 Medline
7. R. M. Flowers, S. A. Bowring, P. W. Reiners, Low long-term erosion rates and extreme continental stability documented by ancient (U-Th)/He dates. *Geology* **34**, 925 (2006). doi:10.1130/G22670A.1
8. D. Stöffler, G. Ryder, Stratigraphy and isotope ages of lunar geologic units: Chronological standard for the inner solar system. *Space Sci. Rev.* **96**, 9–54 (2001). doi:10.1023/A:1011937020193
9. D. W. Hughes, A new approach to the calculation of the cratering rate of the Earth over the last 125 ± 20 Myr. *Mon. Not. R. Astron. Soc.* **317**, 429–437 (2000). doi:10.1046/j.1365-8711.2000.03568.x
10. J. A. Grier, A. S. McEwen, P. G. Lucey, M. Milazzo, R. G. Strom, Optical maturity of ejecta from large rayed lunar craters. *J. Geophys. Res. Planets* **106**, 32847–32862 (2001). doi:10.1029/1999JE001160
11. M. R. Kirchoff, C. R. Chapman, S. Marchi, K. M. Curtis, B. Enke, W. F. Bottke, Ages of large lunar impact craters and implications for bombardment during the Moon's middle age. *Icarus* **225**, 325–341 (2013). doi:10.1016/j.icarus.2013.03.018



4 March 2019; accepted 21 June 2019
Published online 19 July 2019
10.1126/science.aaw9895

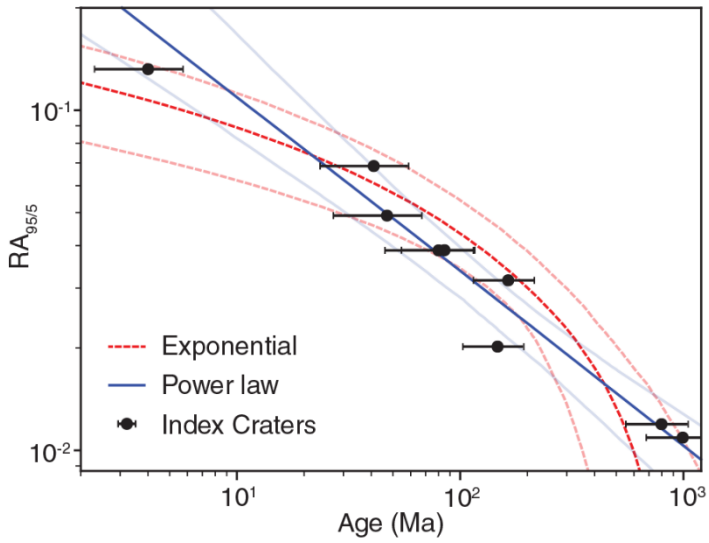
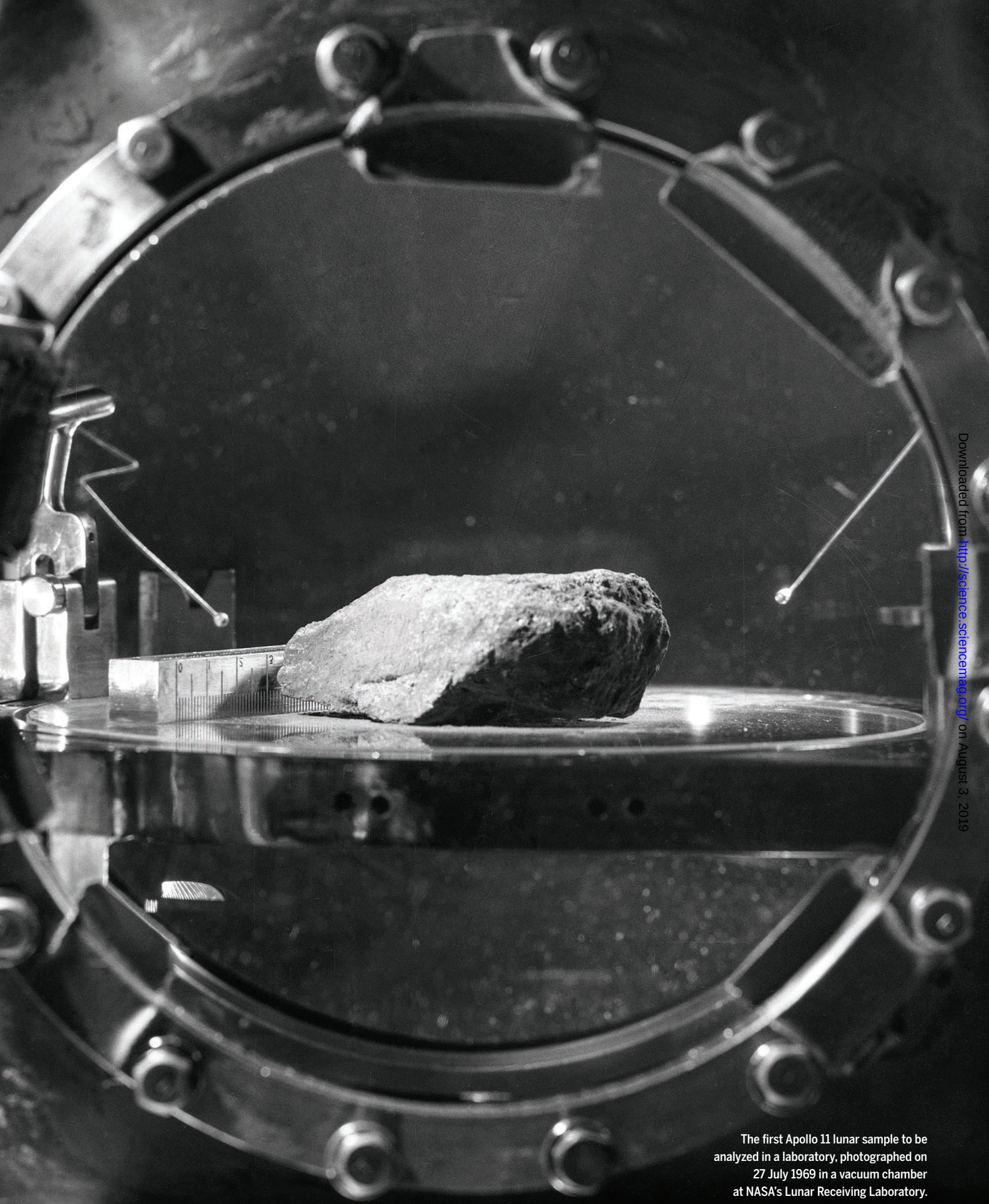



**Fig. 1. Comparison of power-law and exponential age-$RA_{95/5}$ relationships for lunar craters.** Posterior PDFs for model parameters are derived using the same approach as Mazrouei *et al*. (*6*). The median (dark lines) and 95% range (light lines) are illustrated. If the exponential model is adopted (dashed red lines), the uncertainty inflation term *c* prefers a higher value than if the power-law model is adopted (solid blue lines), with a mean uncertainty scaling factor $c \sim 0.5$ versus the mean of $c \sim 0.3$ found for the power-law model.



The first Apollo 11 lunar sample to be analyzed in a laboratory, photographed on 27 July 1969 in a vacuum chamber at NASA's Lunar Receiving Laboratory.

# SHOOTING FOR THE MOON

*By* **Keith T. Smith**

Fifty years ago, on 20 July 1969, the first humans landed on the surface of another world, when the Apollo 11 lunar module touched down on the Moon. A few hours later, astronauts Neil Armstrong and Buzz Aldrin took the first steps onto an alien surface while Michael Collins orbited overhead in the command module.

The motivation for this milestone of space exploration was geopolitical, not scientific. The Apollo program was intended to demonstrate technological supremacy during the Cold War space race. Yet Apollo 11 and the five additional crewed landings that followed have left a rich scientific legacy. They carried a variety of instruments, including seismometers, magnetometers, and retroreflectors for laser ranging observations. Perhaps their greatest scientific value lies in the 382 kg of rocks and regolith (lunar soil) that were brought back to Earth. Laboratory analysis of these samples revolutionized our understanding of how the Moon, Earth, and other planets formed. Improved technology and advances in planetary science mean that the samples are still actively studied today.

Funding for lunar exploration soon ran out. No humans have traveled beyond low Earth orbit since Apollo 17 in 1972, and even robotic missions to the Moon ceased between 1976 and 1990. The 21st century has seen a renewed interest in exploring our nearest neighbor, in part because a growing number of space agencies and private companies now have the capability to reach it. Some have plans to return humans to the lunar surface over the next decade. Even more ambitious are ideas for permanent lunar bases, such as astronomical observatories, which could reap a rich scientific harvest.

One day humans will return to the Moon. Until then, the Apollo landings stand alone at the pinnacle of crewed space exploration.

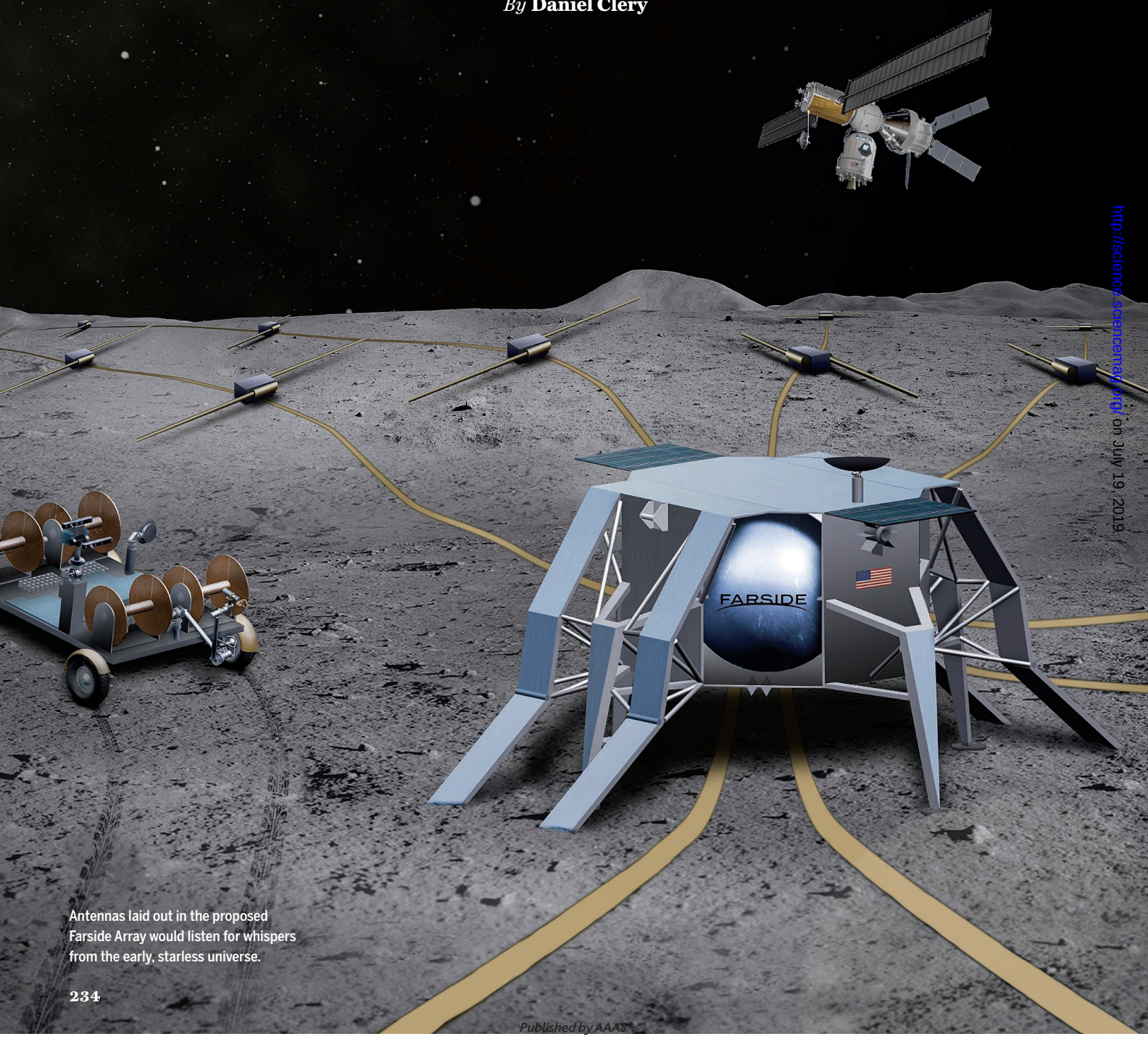


PHOTO: (OPPOSITE PAGE) NASA/NOVAPIX

# MOON GAZING

## Nearly 50 years ago, NASA put a telescope on the Moon. Astronomers have been trying to return ever since

*By* **Daniel Clery**




Antennas laid out in the proposed Farside Array would listen for whispers from the early, starless universe.

In the undulating, dust-covered Descartes Highlands, 380 kilometers southwest of Tranquility Base, where Apollo 11 landed half a century ago, a lonely gold-plated telescope has sat inert since 24 April 1972, when Apollo 16 astronauts John Young and Charles Duke blasted off the surface and left it behind. It was a small part of their 3-day mission, but a milestone for astronomers: the first observatory on another world.

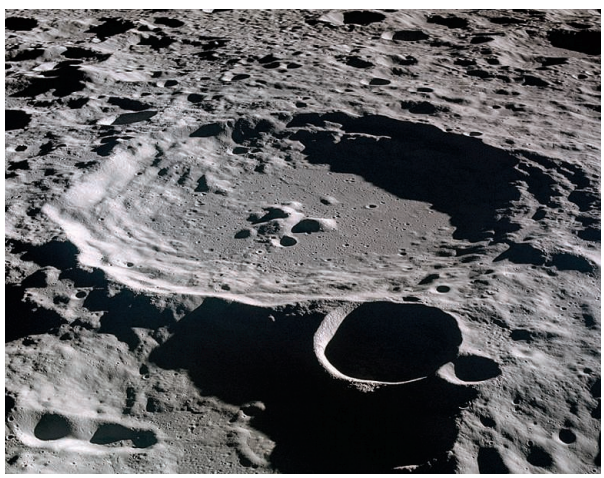
The designer of the telescope was George Carruthers, a young researcher at the Naval Research Laboratory in Washington, D.C., who had made his name building ultraviolet (UV) telescopes for sounding rockets, which make short flights above the UV-blocking atmosphere. A big question at the time was whether the hydrogen in interstellar gas clouds was made of individual atoms or molecules of hydrogen ($H_2$). The answer lay in its UV spectrum, which is difficult to capture from a sounding rocket.

Following the Apollo 11 landing, NASA called for science experiments to fly on future moonshots. Carruthers saw an opportunity to answer the hydrogen question—and gain a unique look at the gases surrounding Earth. Once his proposal was accepted, Carruthers had only 2 years to build the UV sensor—which converted photons to electrons and recorded that signal on film—and train Young and Duke to use the telescope. To block the sun's glare, the astronauts had to set it up in the shadow of the lunar landing module. Every so often, they would return to point the telescope at another target on Carruthers's list. At the end of the mission, the astronauts extracted a film cartridge and left the camera to its fate. That roll of film contained 178 frames, a meager haul, but enough for Carruthers to confirm the existence of interstellar molecular hydrogen—and reveal first-ever UV images of Earth's ionosphere and aurorae, the solar wind, and distant galaxy clusters.

His pioneering lunar observatory will not be the last. The Moon beckons astronomers because it is dry, airless, and seismically quiet. It has room for sprawling arrays of multiple instruments, and it turns slowly, allowing long exposures. For radio astronomers, its far side is a piece of heaven, entirely shielded from interference from terrestrial transmitters.

For many years, however, a lunar observatory has been a dream deferred. Just months after Carruthers's telescope was installed, Apollo 17's Eugene Cernan and Harrison Schmitt left the Moon in a cloud of dust, and the age of lunar exploration suddenly came to an end. That didn't stop astronomers from plotting and planning, but astrophysicist Peter Chen of NASA's Goddard Space Flight Center in Greenbelt, Maryland, says, "For a long time, [a lunar observatory] was not something you brought up in polite conversation."

The Moon is becoming a hot target again, and this time the players include not just NASA, but also other space agencies and commercial space companies. Astronomers are once again along for the ride. On 14 December 2013, China's Chang'e-3 touched down in the Mare Imbrium carrying a near-UV telescope. Although mostly a technology demonstration, the 15-centimeter scope has been collecting data on quasars, blazars, novae, and other hot bright objects and beaming them back to Earth in quantities Carruthers couldn't have imagined in the days of film.



The 100-kilometer-wide Daedalus crater, in the center of the lunar far side, would protect radio astronomy arrays from interference from Earth.

That's just the start of what Steve Durst, president of the Waimea, Hawaii–based International Lunar Observatory Association, calls "this Moon-rush we're experiencing now." His organization hopes to launch a tiny, 1-kilogram precursor telescope with Moon Express of Cape Canaveral, Florida, or one of the other startup companies vying for NASA contracts to take payloads to the Moon. Once missions start to go to the Moon's south pole, the association hopes to send a 2-meter radio dish and a small optical scope to the peak of the 5000-meter Malapert Mountain. The peak receives around-the-clock sunlight, useful for solar power, and its unobstructed view of Earth could aid data transmissions. And the Indian Institute of Astrophysics in Bengaluru is developing a small near-UV imager that the launch company OrbitBeyond would carry on its first mission in 2020.

Astronomers are arguing for far larger radio observatories. From the Moon's quiet far side, they hope to peer back to the universe's dark ages, before the first stars began to shine. Some even envisage building giant optical telescopes, larger than any on Earth, that could not just find Earth-like exoplanets, but see features on their surfaces.

"I've been thinking about this for 35 years and we're finally seeing some traction," says astronomer Jack Burns of the University of Colorado in Boulder. "The feeling is that the Moon isn't that hard to do anymore."

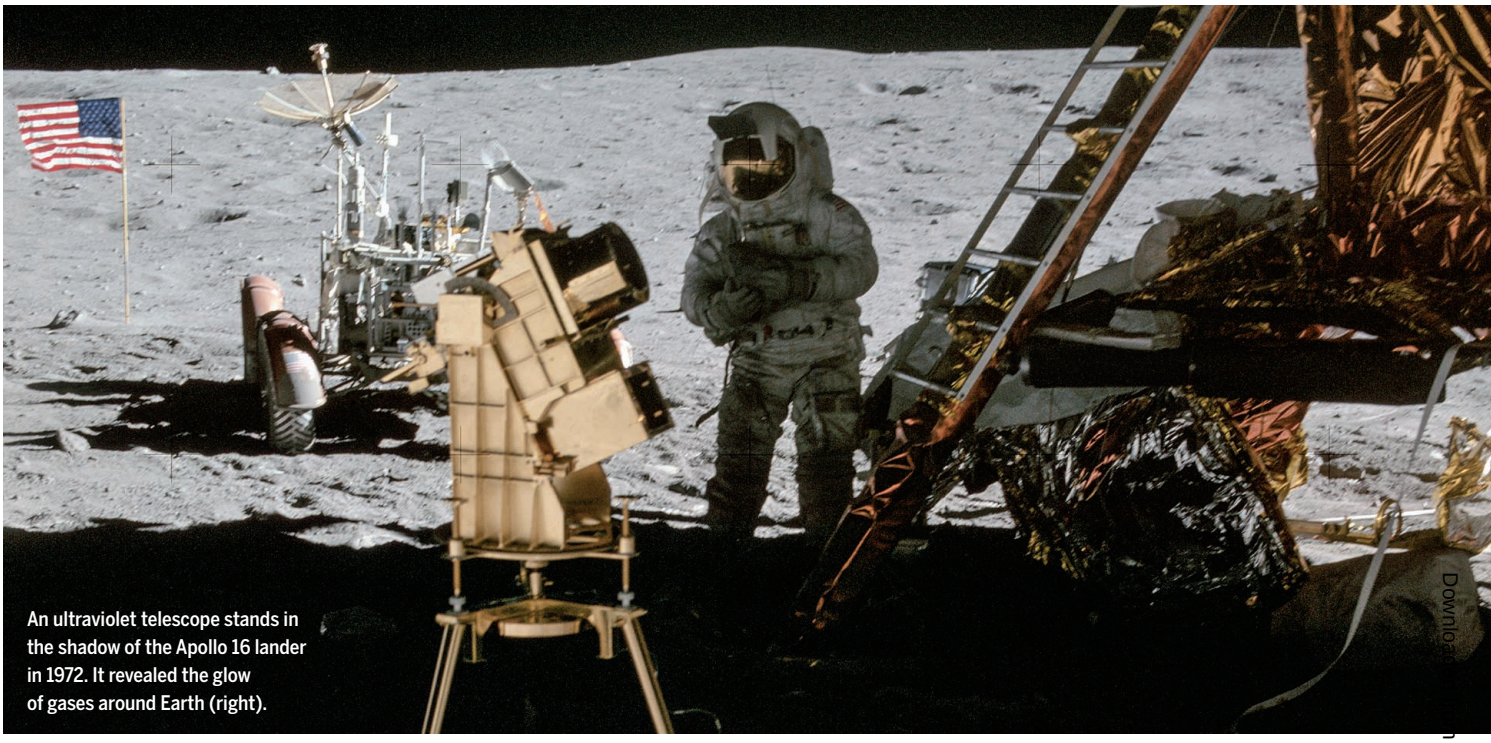
**THE IDEA** of an observatory on the Moon goes back at least to German astronomers Wilhelm Beer and Johann Heinrich von Mädler, who drew the first accurate lunar maps in the 1830s. They realized that, lacking an atmosphere, the Moon would offer a pristine view of the sky. When the space age suddenly made the idea seem possible, U.S. astronomers held multiple lunar astronomy workshops beginning in 1965. At a large gathering in Annapolis in 1990, astronomer Harlan Smith of the University of Texas in Austin, a champion of the Hubble Space Telescope that was then due for launch, charted the progress of astronomy toward ever-larger telescopes in ever-more-remote and high-altitude locations. "The steady unfolding of this tradition should soon lead us to the Moon—the best place in the solar system from which to do many if not even most kinds of astronomy."

Before long, however, orbiting space telescopes such as Hubble, Spitzer, and Chandra were scoring major successes across the electromagnetic spectrum. Few could see the gains of shipping a telescope all the way to the Moon for a risky landing when a cheaper trip to low-Earth orbit would work just as well.

One group has continued to hold out for the Moon: radio astronomers. On Earth, radio telescopes must tussle with signals from terrestrial TV and radio transmitters for clean looks at the sky. The problem is worst for observations at low frequencies, below 50 megahertz (MHz). At those frequencies, the ionosphere, a layer containing charged particles high in the atmosphere, can refract, disperse, or completely block radio waves. Yet such low-frequency waves hold clues to events in the deepest, darkest era of the universe's history.

Several hundred thousand years after the big bang, the roiling plasma of charged particles it created had cooled enough for

PHOTOS: NASA; (OPPOSITE PAGE) ROBIN CLARKE/JPL



An ultraviolet telescope stands in the shadow of the Apollo 16 lander in 1972. It revealed the glow of gases around Earth (right).



electrons to latch onto protons and make hydrogen atoms, releasing a burst of light that we now see as the cosmic microwave background (CMB). Then, the dark ages began. For several hundred million years there was nothing in the universe but dark clouds of neutral hydrogen and a few other atoms, slowly being pulled together by clumps of dark matter.

The starless universe wasn't completely dark. On rare occasions the spins of the protons and electrons in neutral hydrogen atoms flip from being parallel to antiparallel, emitting a photon at a frequency of 1.4 gigahertz—the so-called 21-centimter radiation. This dark ages signal is just a whisper. But, theorists calculate, a sufficiently large array of antennas should be able to detect it. They could also trace its disappearance later in cosmic history, as the first stars ended the dark ages by bursting into light and ionizing the hydrogen around them with their UV radiation. The details of the process could reveal whether the first generation of massive stars did most of that ionization, or whether much more powerful quasars, extremely bright galactic cores centered on voracious black holes, also played a role.

The epoch contains far more data than the CMB, which only records a brief moment in cosmic time. "This is a 3D volume with primordial fluctuations in there, just waiting to be measured," says astronomer Jacqueline Hewitt of the Massachusetts Institute of Technology in Cambridge.

Charting the dark ages' 21-centimeter radiation "is essential. It has to be done at some point," says radio astronomer Heino Falcke of Radboud University in Nijmegen, the Netherlands, who works on the Low Frequency Array (LOFAR), a vast radio telescope spread across northern Europe. The problem is that after 13 billion years of cosmic expansion, photons from the dark ages arrive at Earth stretched to frequencies below 50 MHz, where they not only have to contend with the ionosphere, but also with confounding signals from ham radio, aviation, and long-distance communications. LOFAR, for example, has struggled to detect the 21-centimeter signal from that epoch (*Science*, 7 November 2014, p. 688). The Hydrogen Epoch of Reionization Array, which Hewitt and others are building in South Africa, may have a better shot. But Falcke and others are looking to the Moon.

Burns's 35-year campaign for a lunar radio telescope won him NASA funding in 2008 to form a team called LUNAR to work out how to build it. The researchers designed a lunar telescope array that would include hundreds of simple dipole antennas laid flat on the ground. They demonstrated how autonomous rovers could lay out strips of conducting film to act as antennas. In a 2013 experiment, astronauts on the International Space Station remotely guided a rover that laid out antenna strips on a simulated moonscape at the NASA Ames Research Center in Mountain View, California, to show how a future crew onboard NASA's proposed Moon-orbiting Lunar Gateway station could supervise construction.

A more recent NASA-funded project, the Network for Exploration and Space Science (NESS), is drawing up plans for lunar radio telescopes that could probe exoplanets as well as the early universe. Planets that have a magnetic field emit low-frequency radio waves when electrons spiral around the field lines. Because planetary magnetic fields require an internal dynamo, like a spinning liquid metal core, detecting a field could offer clues to an exoplanet's interior—and, Burns says, "could play a key role in assessing [its] habitability." Earth's magnetic field, for example, not only protects life from cancer-causing high energy particles, but is also thought to have shielded water in the atmosphere from being lost to space.

Earth-based radio arrays might be able to detect the radio emissions from a Jupiter-size exoplanet, with a mighty magnetic field. But a lunar observatory might pick up the fainter—and lower frequency—emission of a rocky exoplanet.

NESS is working on two proposals. The first is a $60 million suitcase-size satellite with antennas on four sides that will tune in to a range of wavelengths by unfurling like frog tongues to lengths of up to several meters. It will orbit the Moon and, when above the far side and shielded from Earth, will try to detect the dark age 21-centimeter signal. Known as DAPPER, the Dark Ages Polarimeter Pathfinder satellite should be

PHOTO: NASA

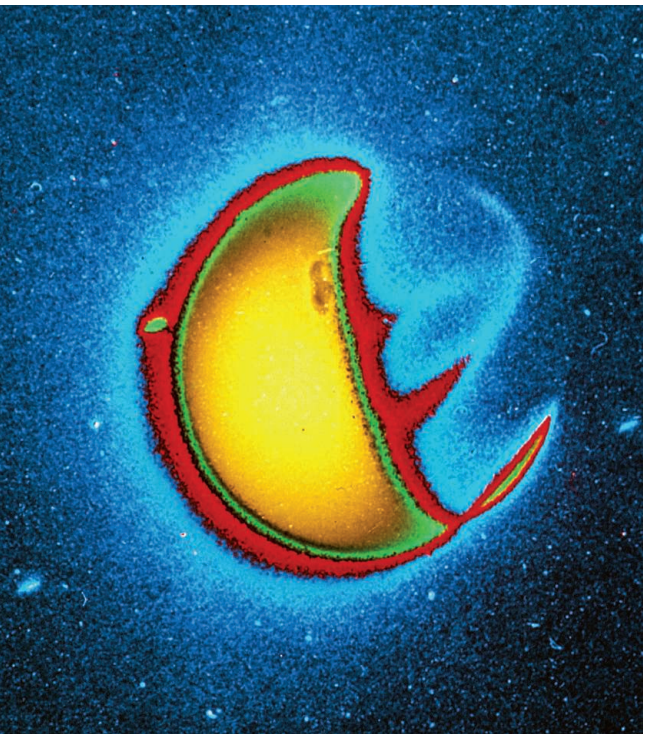


IMAGE: NASA

able to map the radiation in enough detail to determine how dark matter tugged the primordial hydrogen clouds into clumps. That could help distinguish between rival models of cosmic structure formation.

The second is a billion-dollar class mission to put a basic radio telescope on the surface of the Moon. The Farside Array would use the techniques pioneered by the LUNAR consortium: small rovers laying out a total of 128 dipole antennas across an area 10 kilometers wide, supervised by astronauts in lunar orbit. This month, NESS submitted a proposal to Astro2020, the U.S. astronomy community's once-a-decade priority-setting exercise, in the hope of winning wider support.

European and Chinese researchers are thinking along similar lines. Last year, a team led by Falcke added a radio astronomy experiment to China's Queqiao probe, which was parked near the Moon to relay signals to and from the Chang'e-4 lander on the lunar far side (*Science*, 18 May 2018, p. 698). Now that Queqiao has completed its communications duties, Falcke says, the Netherlands-China Low-Frequency Explorer (NCLE) can start observations later this year. The NCLE will not get a clean 21-centimeter signal because it is not completely shielded from earthly interference, but it will be a testbed for observing in the lunar neighborhood in preparation for a bigger goal: a lunar LOFAR. "We're discussing with China about follow-up projects," Falcke says.

The European Space Agency (ESA) is also considering options. In May, the agency published a 10-year strategy for science on the Moon; its goals include testing low-frequency radio receivers on the far side. "To make the first observations [from the far side] would be fantastic," says James Carpenter, strategy officer in ESA's human and robotic exploration division in Noordwijk, the Netherlands. A larger array of antennas is "a long term aspiration," he adds.

**OPTICAL ASTRONOMERS** don't have the same all-or-nothing need for a lunar observatory. But that hasn't stopped them from imagining what they could see with a telescope as clear-eyed as Hubble, but with a mirror larger than could ever be launched into Earth orbit. The Moon's low gravity and rock-solid stability mean support structures for a giant telescope could be simple and cheap. But here's the catch: The mirror would have to be made in situ.

One solution, proposed by a NASA-funded team last decade, is to build a liquid mirror. The idea is simple: Construct a large shallow circular dish, fill it with a liquid, and set it gently spinning. Centrifugal force will pull the surface into a parabolic shape. Liquid mirror telescopes on Earth use mercury, which is naturally reflective. The biggest so far was the 6-meter Large Zenith Telescope in British Columbia in Canada. That testbed telescope, built in 2003 and now decommissioned, paved the way for the 4-meter International Liquid Mirror Telescope, which will take its first view of the sky from the Devasthal Observatory in India later this year. Although a liquid mirror is limited to looking straight up, the rotation of Earth—or the Moon—scans it across the sky.

Mercury won't work on the Moon—it would evaporate in the lunar vacuum and is too heavy to transport from Earth. In its place, the NASA team came up with a class of organic compounds called ionic liquids, essentially molten salts, that would remain liquid in the frigid lunar night. Ionic liquids are not reflective, but could be glazed with silver to make an ideal reflecting surface. Superconducting bearings could levitate the dish and keep it spinning frictionlessly. "In principle there is no limit on the size," says team member Ermanno Borra of Laval University in Quebec City, Canada. "This would be totally impossible in space, but not that expensive on the Moon."

The Canadian Space Agency (CSA) followed up the NASA study by looking at the practicalities of building such a scope. "There were no showstoppers and the mechanical tolerances were more relaxed than for a space telescope," says Paul Hickson of the University of British Columbia in Vancouver. A 20-meter telescope, CSA concluded, would require no more than 3.5 tons of material to be transported to the surface. An even larger instrument, as big as 100 meters across, "would be in a class of its own," Hickson says, able to study the very first stars that formed and coalesced into galaxies at the end of the dark ages.

Chen wants to make giant lunar mirrors from moondust, minimizing the mass needed to be brought from Earth. In a small laboratory test last decade, he mixed a simulated lunar soil, or regolith, with epoxy to produce a paste the consistency of melted chocolate. He spun the mixture in a dish, like a liquid mirror, then cured the epoxy with UV light, which is plentiful on the Moon. The result was a 30-centimeter parabolic dish that could be coated with aluminum to produce a mirror. Other parts of the telescope structure could also be made from regolith paste using a 3D printing machine, he says. More recently, Chen has made thin moondust mirrors that can be reshaped by actuators on the back surface to compensate for temperature changes and for gravity as the mirror moves.

Such mirrors could be made in batches, Chen says, and combined into optical interferometers, which could achieve extraordinary resolution when viewing distant objects. An interferometer harnessing mirrors 1 kilometer apart would have the resolution of a 1-kilometer mirror, even if not its light-gathering power, enabling astronomers to scrutinize the surface of a nearby Earth-like exoplanet. But combining the light of widely spaced telescopes requires monumental precision, which small earthquakes can disrupt. That's one reason why Earth-based optical interferometers remain experimental and span no more than tens of meters. Much larger interferometers might be feasible on the Moon, with its seismic quiet. "The Moon is the only place you can do it," Chen says.

Although many of these schemes may sound like science fiction, would-be lunar astronomers feel a sense of urgency. Other parties also hope to exploit the Moon, for purposes that could disrupt astronomy. Mining could loft dust and cause tremors; Moon bases would generate radio interference that could affect the far side.

For radio astronomers, their sensitive instruments increasingly deafened on Earth, that is a bleak prospect. Falcke says: "If the far side gets spoiled, I don't know where we would go." ■

POLICY FORUM

# China's present and future lunar exploration program

Chunlai Li[1], Chi Wang[2]*, Yong Wei[3], Yangting Lin[3]

**Since the beginning of the 21st century, the pace of lunar exploration has accelerated, with more than a dozen probes having undertaken scientific exploration of the Moon. Prominent among these have been the robotic "Chang'E" (CE) missions of the China Lunar Exploration Program (CLEP). We discuss technological and scientific goals and achievements for the four completed, and four planned, CE missions, and longer-term goals and plans of the CLEP beyond the CE missions. The exploration plan is flexible and iterative, with an emphasis on international cooperation.**

On 3 January 2019, the first lunar farside in situ exploration mission, China's Chang'E-4 (CE-4), successfully landed in Von Kármán crater within the South Pole–Aitken (SPA) basin. Fifty years after U.S. astronaut Neil Armstrong took "one small step for [a] man, one giant leap for mankind" as the first human to set foot on the Moon, China's CE-4 lander and Yutu-2 rover left the footprints of humanity's first robotic visit to the surface of the far side of the Moon, opening a new chapter in lunar exploration. CE-4 is one of the missions of the China Lunar Exploration Program (CLEP). With China still in its infancy in this field, we discuss how the overall goal of CLEP is to undertake substantial scientific exploration by laying a robust engineering and technical foundation, with an emphasis on international cooperation.

The mysteries of the origin and evolution of the Moon are one of the most basic problems in natural science, continuing to attract the interest and excitement of scientists and engineers worldwide. Humanity has carried out more than 100 lunar exploration missions. Since the beginning of the 21st century, the pace has accelerated, with more than a dozen probes having undertaken scientific exploration of the Moon, including SMART-1 (ESA); SELENE Kaguya (Japan); Chandrayaan-1 (India); LCROSS; LRO; GRAIL; LADEE (United States); and CE-1, CE-2, CE-3, and CE-4 (China), resulting in many new discoveries and scientific achievements.

## Science goals, technology development

In 2004, China formulated a robotic lunar exploration program, consisting of three phases—orbiting, landing, and returning—and named the program the "Chang'E Project," after the Chinese goddess of the Moon. In 2005, Ziyuan Ouyang, the first chief scientist of CLEP, raised 14 key questions and issues that have become the blueprint of China's developing strategy and planning for lunar exploration (Fig. 1), providing direction for China's scientific goals and payload allocation (*1*): (i) the composition, source, and evolution of the lunar atmosphere; (ii) the nature of the ionosphere of the Moon; (iii) the nature and origin of lunar topography and geomorphology; (iv) the nature of the space environment and the lunar surface environment; (v) the origin, material sources, distribution characteristics, and formation process of lunar soil (regolith); (vi) the main lunar rock types and their distribution and origin; (vii) the mystery of water ice on the Moon; (viii) the nature and history of tectonic deformation on the Moon; (ix) the internal structure of the Moon and the formation processes; (x) the heterogeneity of the distribution of material within the Moon; (xi) the origin and evolution of the lunar global intrinsic dipole magnetic field; (xii) the thermal evolution of the Moon; (xiii) the nature and timing of the major events in the evolution of the Moon; and (xiv) the origin of the Moon and the Earth-Moon system. In 2018, the NASA Advisory Council Lunar Exploration Analysis Group (LEAG) updated and improved 11 key scientific concepts for lunar exploration, as well as specific research proposals to address these 11 scientific concepts (*2*). These science questions are fundamental in lunar exploration missions, with individual missions focusing on addressing some of them.

The goals of CLEP from 2004 to 2020 are to gain a global and comprehensive understanding of the Moon through orbital spacecraft exploration; to conduct exploration and surveying of the lunar surface, through Earth-based monitoring, sky mapping, and lunar soft landing with landers and rovers; and to develop a more in-depth understanding of the Moon and its history through the sampling of lunar rocks and soils, and returning them to Earth. The missions are iterative and intertwined from the perspectives of science, engineering, and technology. The exploration plan is very flexible, depending on what is learned from these areas during each step in the program.

In the course of successfully executing the CE-1 to CE-4 missions (2007–2019), China has made considerable progress and development in many areas of space technology, such as orbit design, flight control, high-precision telemetry and telecommunication, lunar soft landing, and traverse exploration. For example, the CE-4 mission positioned a relay satellite in the Earth-Moon L2 halo orbit to establish communication between Earth and the far side of the Moon. In addition to laying solid technical foundations for subsequent lunar exploration missions, the CE missions have achieved important scientific results. For example, orbital exploration of the Moon from CE-1 and CE-2 provided a 7-m-resolution lunar global image and topographic maps (*3*), the first global analysis of the Moon's microwave radiation (*4*), and discovery of the acceleration of protons at the lunar day–night interface (*5*). The CE-3 lander and rover revealed a new type of lunar basaltic rock (*6*), and the characteristics of the subsurface structure of the landing area through radar probes (*7*, *8*). Using the lunar surface as a unique astronomical observation platform, CE-3 studied brightness changes of stars, as well as changes in Earth's plasmasphere in the ultraviolet band (*9*). The CE-4 spacecraft provided the first in situ measurements from the far side of the Moon. Visible and near-infrared spectra of impact ejecta excavated and delivered from the adjacent Finsen crater exhibit the petrographic features of the lunar deep interior materials (*10*). CE-4 is analyzing the electromagnetic environment of the far side of the Moon, assessing its potential as a site for low-frequency radio astronomy observations and research. CE-4 is also seeking to detect low-frequency radio waves (0.1 to 40 MHz) generated by solar activity and other sources (*11*).

The CE-5 mission, scheduled for early 2020, is designed to return samples to Earth from Mons Rümker, the northern part of Oceanus Procellarum. This site was chosen for having some of the youngest of the lunar volcanic mare basalt rocks, as yet unsampled. The scientific goals of CE-5 are to obtain a firm date for the end of lunar volcanism, to understand how the mineralogy and petrology of young volcanism differs from those of earlier times, and to provide a comprehensive view of lunar thermal and interior evolution.

In 2015, after the launch and successful execution of early phases of the CLEP orbiting, landing, and returning project, China proposed a follow-up plan to take place before 2030. This program consists of surveying the lunar environment and resources, constructing a long-term fundamental research platform, and verifying technology for exploiting available resources. The overall scientific objectives include (i) investigating the global distribution, content, and source of water and volatile components; (ii) investigating the composition and structure of the Moon's interior; (iii) measuring the age of the South Pole–Aitken Basin; and (iv) investigating the space physical environment above the lunar South Pole. The application objectives include (i) in situ lunar resource (water, helium, etc.) utilization experiments; (ii) Moon-based observation and research of Earth; and (iii) scientific experiments related to the lunar surface ecosystem, etc. On the basis of the overall scientific objectives, three missions were initially planned to be implemented by

[1]National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China. [2]National Space Science Center, Chinese Academy of Sciences, Beijing, China. [3]Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China.
*Corresponding author. Email: cw@spaceweather.ac.cn

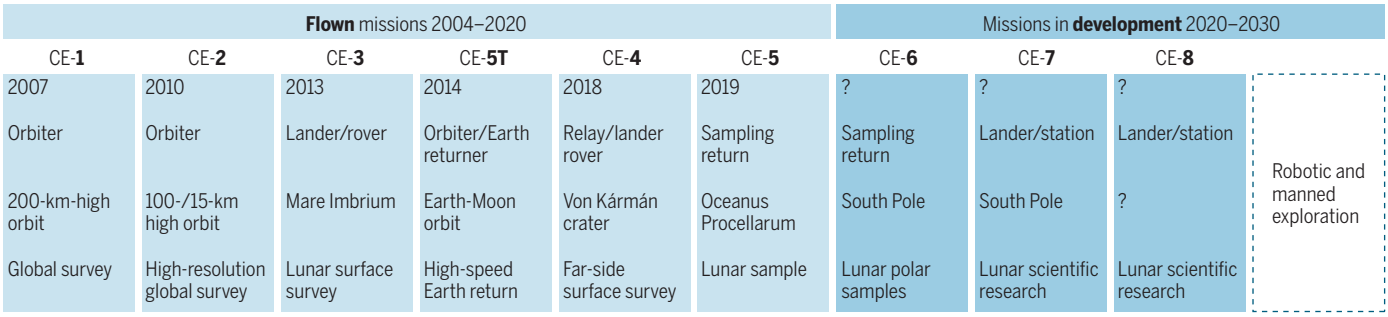



**Fig. 1. China's lunar exploration development strategy.** Missions CE-1 to CE-5 before 2020 aim at successfully implementing the CLEP early three-phase "orbiting, landing, and sampling return" project. Follow-up missions CE-6 to CE-8 and possible further robotic and manned missions aim at "surveying, constructing, and exploiting" the Moon from 2020 to 2030.

2030, including CE-6, to sample return from the south polar area; CE-7, to survey the environment and resources in the south polar area; and CE-8, to verify key technologies such as "3D printing" construction on the Moon (*12*). Through these missions, a robotic scientific research station prototype will be built on the Moon. Exploration targets will shift focus from development of space technology, to space science and space applications. To achieve these goals, it is necessary to develop several key technologies, such as high-precision fixed-point landing, landing of large payloads, developing exploration capabilities in the permanently shadowed areas of impact craters, developing intelligent robot adaptation to the harsh environment of the polar regions, coordinated operation of intelligent robots, comprehensive operational control of a Lunar Scientific Research Station, and separation and extraction of rare gases. The Lunar Scientific Research Station, with the capability of long-duration operations and intelligent operational control, will be designed to carry out technical verification and validation of resource development and utilization technology, explore prospects for applications, enhance the ability of lunar science and resource application, and lay the foundation for the construction and operation of future Lunar Research Stations, as well as exploration of the Moon by humans.

After 2030, China's lunar exploration program will continue to develop capabilities in both robotic and human exploration. On the basis of factors such as technology development, launch vehicle selection, economic affordability, and cost-effectiveness ratios, robot exploration will remain the primary development direction of subsequent lunar exploration missions. By deepening lunar scientific exploration and verifying technology, such as resource development and utilization, bio-regeneration, and life support, the Lunar Research Station could be built into a long-duration lunar base that astronauts can visit for a short time, with the eventual goal of long-term stay of astronauts on the Moon.

## Foundation of international cooperation

International cooperation is an important element in China's strategy of lunar and deep space exploration. For example, the CE-4 mission provided a host platform and opened its payload resources to the international community. The CE-4 lander, rover, and relay satellite carried experiments from Germany [Lunar Lander Neutrons & Dosimetry (LND)], Sweden [Advanced Small Analyzer for Neutrals (ASAN)], and the Netherlands [Netherlands-China Low Frequency Wavelength Explorer (NCLF)]. China cooperated with these countries in scientific instrument development, scientific exploration planning, exploration data research, and other aspects, and jointly completed the first soft landing mission on the far side of the Moon. In March 2019, the China National Space Administration (CNSA) and the Russian State Aerospace Corporation reached an agreement to jointly implement China's lunar polar exploration mission and Russia's Luna mission and collaborate on landing site survey capabilities, relay communication, scientific payload, and space experiments. In April 2019, the CNSA announced cooperation opportunities for China's CE-6 and its asteroid exploration mission (*13*) and signed lunar exploration cooperation agreements with the United Nations Office for Outer Space Affairs (UNOOSA), Turkey, Ethiopia, and Pakistan (*14*). China is also open to cooperation with NASA on lunar exploration; both sides can start cooperating on aspects such as exchange of scientific data and space situational awareness information (*15*). China also looks forward to exploring more opportunities to cooperate with NASA to preserve the space environment for generations to come. China has made policy adjustments to the sources of funding for future lunar and deep space exploration missions and has widely encouraged the participation of commercial and private enterprises in addition to governments. The CNSA is setting up general procedures for international collaboration and the mechanism to organize international teams. Extensive international cooperation and diverse sources of funding will inject additional vitality into the ongoing planning and development of lunar and deep space exploration.

Just as the Apollo program played a positive role in promoting the development of human society, China will work with countries around the world in its forward-looking lunar and deep space exploration projects to build a community and a shared future for humankind, promote the development of science, and jointly create a better future for humanity through space exploration and achievements in space science and technology.

## REFERENCES AND NOTES


1. Z. Y. Ouyang *et al.*, *Introduction to Lunar Science* (in Chinese) (China Astronautic Publishing House, 2005).
2. Lunar and Planetary Institute, "Advancing Science of the Moon: Report of the Specific Action Team" (2018); www.lpi.usra.edu/leag/reports/ASM-SAT-Report-final.pdf.
3. X. Ren *et al.*, *IEEE Trans. Geosci. Remote Sens.* 10.1109/TGRS.2019.2908813 (2019).
4. Y. C. Zheng *et al.*, *Icarus* **219**, 194–210 (2012).
5. X. D. Wang *et al.*, *Geophys. Res. Lett.* **37**, L07203 (2010).
6. Z. Ling *et al.*, *Nat. Commun.* **6**, 8880 (2015).
7. J. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5342–5347 (2015).
8. L. Xiao *et al.*, *Science* **347**, 1226–1229 (2015).
9. H. He *et al.*, *Sci. Rep.* **6**, 32362 (2016).
10. C. Li *et al.*, *Nature* **569**, 378–382 (2019).
11. W. R. Wu *et al.*, *Nat. Geosci.* **12**, 222–223 (2019).
12. CNSA, "Chang'e 4 press conference" (2019); https://www.youtube.com/watch?v=v7FiaHwv-BI.
13. CNSA, "Announcement of Opportunities for Scientific Payloads onboard Chang'E-6 Mission", "Announcement of Opportunities for Scientific Payloads and Projects onboard Asteroid Exploration Mission" (2019); www.cnsa.gov.cn/english/n6465645/n6465648/c6806207/part/6781080.pdf, www.cnsa.gov.cn/english/n6465652/n6465653/c6805891/part/6780424.pdf.
14. CNSA, "United Nations/China Forum on Space Solutions: Realizing the Sustainable Development Goals " (2019); www.cnsa.gov.cn/english/n6465684/n6760084/index.html, www.cnsa.gov.cn/n6758823/n6758838/c6806136/content.htm.
15. SpaceNews, "NASA opens door to additional cooperation with China " (2018); https://spacenews.com/nasa-opens-door-to-additional-cooperation-with-china/.


## ACKNOWLEDGMENTS


C.L. is Deputy Chief Engineering Designer of the CE5 mission, Director of the Ground Research and Application System of the Chinese Lunar Exploration Project. C. W. is Deputy Chief Engineering Designer of the CE4 mission, Director of the Scientific Payload Subsystem of the Chinese Lunar Exploration Project. The authors thank R. X. Zhu, Z. Y. Ouyang, W. R. Wu, Z. Y. Pei, J. J. Liu, W. Zuo, Y. L. Zou, L. Xu, H. Y. He, and Z. Zhang for their contributions to this paper.

10.1126/science.aax9908

REVIEW

# Analysis of lunar samples: Implications for planet formation and evolution

Richard W. Carlson

**The analysis of lunar samples returned to Earth by the Apollo and Luna missions changed our view of the processes involved in planet formation. The data obtained on lunar samples brought to light the importance during planet growth of highly energetic collisions that lead to global-scale melting. This violent birth determines the initial structure and long-term evolution of planets. Once past its formative era, the lunar surface has served as a recorder of more than 4 billion years of interaction with the space environment. The chronologic record of lunar cratering determined from the returned samples underpins age estimates for planetary surfaces throughout the inner Solar System and provides evidence of the dynamic nature of the Solar System during the planet-forming era.**

Planetary scientists seek to understand how planets formed and evolved to their present states. Because Earth is geologically active, much of the terrestrial record from the time of planet formation has been overwritten so many times that it is now hard to separate the role of ancient events from more recent ones. In contrast, many features of the Moon have been preserved from this formative era. Some can be studied remotely, but others were revealed only after the lunar samples returned during the Apollo and Luna missions were analyzed in terrestrial laboratories.

## Magma oceanography

Prior to the Apollo landings, the following statement represented the prevailing view of terrestrial planet formation: "It seems possible and indeed probable that the earth could and did accumulate below the melting point of silicates throughout its entire growth from a small size to its present one" (*1*). This reflected the opinion that planets grew by the gentle accumulation of asteroid-sized planetesimals, so that high temperatures only occurred locally in some of the larger impacts. The samples collected during the Apollo 11 mission in 1969 quickly disproved the idea that planetary bodies, even ones the size of our Moon, started out cold. Bulk samples from the Apollo 11 landing site contained a minor component (4%) of anorthosite, an igneous rock type that is known, but uncommon, on Earth. Anorthosite consists predominantly of a single mineral, plagioclase feldspar. A handful of small mineral grains led to the suggestion that the entire lunar highlands were dominated by anorthosite (*2*, *3*). Based on the elevation of the highlands relative to the flat plains of basaltic lavas that infilled the major impact basins on the Moon's Earth-facing side (originally misidentified by early astronomers as seas, leading them to be called mare basins), the assumption of an anorthositic composition for the highlands implied a highlands crust some 25 km thick (*2*). The current best estimate for lunar nearside crustal thickness is ~30 km based on lunar gravity data measured more than 40 years later (*4*).

How could essentially all of the lunar highlands crust be composed of anorthosite? In molten rock (magma) of composition similar to that of the bulk Moon, anorthite plagioclase, a calcium-aluminum silicate, has a lower density than the magma, and so would float. An obvious way to form an anorthositic crust is to assume that a large portion of the Moon was at one time molten, with the anorthositic crust forming by flotation above the crystallizing magma in much the same way that an iceberg floats on the ocean. This idea led to what became known as the "magma ocean" model for early lunar differentiation (Fig. 1).

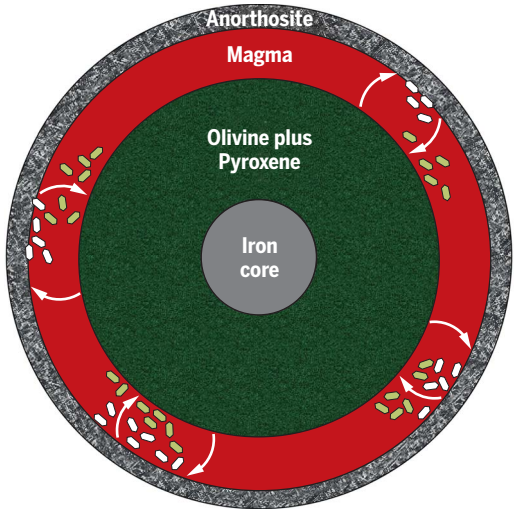



**Fig. 1. Diagram of the lunar magma ocean model.** Convection (arrows) within the magma ocean circulates the magma. Crystallization of dense olivine and pyroxene (green) and buoyant plagioclase (gray) separates the magnesium-rich minerals of the interior from the anorthositic crust. The thickness of the lunar crust is exaggerated by about a factor of 10 for display.

From a magma with the composition expected for the bulk Moon, the dense magnesium silicates olivine and pyroxene would have been the first minerals to crystallize (*5*–*9*). Settling of these crystals to the bottom of the magma ocean left the remaining magma richer in calcium and aluminum, leading to crystallization of plagioclase. Olivine and pyroxene crystallization also drove the remaining magma to higher concentrations of iron, which served to further increase its density. By the time the magma became saturated in plagioclase, as it crystallized, its buoyancy with respect to the magma caused the plagioclase to float to the surface of the magma ocean to form the highlands crust. As the crust formed, it trapped beneath it the remaining magma that was enriched in all elements that were not incorporated into the minerals that had crystallized by that time. These elements include potassium (K), the rare earth elements (REE), and phosphorus (P), leading to the acronym KREEP for the rocks formed from this residual magma (*10*). Because KREEP is rich in radioactive elements such as uranium, thorium, and potassium, whose concentrations can be measured from orbit because they emit gamma rays (*11*), we now have maps of the lunar surface that show that KREEP is concentrated dominantly on the Earth-facing side of the Moon, in the region of the major mare basalt-filled impact basins (*11*–*13*) (Fig. 2).

The restricted areal distribution of KREEP could reflect the presence of a KREEP layer beneath the whole crust that was only excavated to the surface by impacts large enough to penetrate through the crust. However, the largest impact basin on the Moon, the South Pole–Aitken basin, shows only slightly elevated abundances of thorium (Fig. 2). Alternatively, both the apparently asymmetric distribution of KREEP and the thicker lunar crust on the farside than on the nearside (*4*) have led to proposals that the asymmetry originated in the magma ocean era by either gravitational or thermal interaction with Earth (*14*, *15*). The concentration of radioactive KREEP into a smaller portion of the Moon also could have provided a heat source to explain both mare basalt volcanism that continued for more than a billion years after the Moon solidified (*16*) and the higher number of large impact basins facing Earth than on the lunar farside (*17*).

The simple magma ocean model described above almost certainly underestimates the complexity of the process. Several investigations have explored more complicated evolution scenarios, including the consequences of density-driven overturn of different layers in the lunar interior after magma ocean crystallization (*18*–*20*). Nonetheless, the basic magma ocean model explains many characteristics of the lunar samples. Later Apollo missions that landed on, or close to, the highlands returned substantial amounts of anorthosite. The global importance of anorthosite has been confirmed through

Department of Terrestrial Magnetism, Carnegie Institution for Science, Washington, DC 20015, USA.
Email: rcarlson@carnegiescience.edu

compositional analysis of the lunar surface from orbit (*13*), as well as by meteorites from the Moon that provide samples of the lunar crust likely far removed from the Apollo landing sites (*21*). Some mare basalts have compositions similar to terrestrial basalts, but others have distinctively high contents of titanium (*5*, *22*). These unusual lunar lava compositions most likely reflect a wide range in the composition of the rocks in the lunar interior that melted to create these lava flows (*5*, *7*, *9*). This type of compositional diversity in the lunar interior is predicted by the magma ocean model, as it fits the sequence of dense mineral accumulation that would occur as the magma ocean cooled and crystallized (*6*).

Further evidence in support of a lunar magma ocean comes from measurements of the isotopic composition of elements that receive the decay products of naturally occurring radioactive isotopes. Unless a magma ocean has a thick insulating crust, or atmosphere, above it, heat loss to space causes cooling and crystallization of the magma in a few million years or less (*23*, *24*)—a time that is short relative to the chronological resolution of many radioactive clocks used to determine the age of rocks. If they formed by direct crystallization of the lunar magma ocean, the rocks of the highlands crust should show little variation in age, corresponding to the date of magma ocean crystallization. The data, however, show a dispersion in ages for different highland rocks (*25*). Whether this reflects the range in actual crystallization ages, or resetting of the radiometric clocks in the samples by slow cooling deep in the crust or by impact-related metamorphism, is unclear. Modern applications of radiometric dating techniques are narrowing the age range of the crustal rocks to 4.36 to 4.40 billion years (*25*), 150 to 200 million years after the start of Solar System formation. Resolving the age range of the lunar crust is made difficult by the small number of returned crustal samples appropriate for dating and the fact that the oldest crust on the Moon has been repeatedly pummeled by impacting objects. Obtaining a clear answer to the age of the magma ocean, and the lunar crust in general, likely will require a broader collection of crustal rocks to be returned from the Moon.

Another prediction of the magma ocean model is that the anorthositic crust, KREEP, and the accumulated crystals in the lunar interior that would later be melted to produce mare basalt lavas all formed over a relatively short time interval. Radioactive dating of mare and KREEP basalts indicates eruption ages from 4.25 to 2.9 billion years ago (*8*). At the time of their eruption, however, the basalts were characterized by a wide range in the relative abundances of the isotopes that are produced by the radioactive decay of naturally occurring radioactive elements (*26*). This is true also for one radiometric system, the 103 million–year half-life decay of $^{146}$Sm to $^{142}$Nd. In this system, the parent isotope, $^{146}$Sm, effectively became extinct about 4 billion years ago. Nonetheless, lunar lavas younger than 4 billion years show a range in $^{142}$Nd/$^{144}$Nd (*27*–*29*), where $^{144}$Nd is a stable isotope of neodymium (Fig. 3). The inherited variability in the isotopic composition of these elements at the time of mare basalt eruption shows that the compositional variability of the basalts can be traced back to the characteristics of the rocks in the lunar interior that were melted to generate these lavas. Combining the radiogenic isotope data for all analyzed lunar igneous rocks (e.g., Fig. 3) suggests that their source regions formed in an initial event that occurred between 4.37 and 4.40 billion years ago (*27*–*31*), similar to the ages of crustal anorthosites. Such a result is consistent with the idea that the main compositional structure of the lunar crust and mantle was determined by an event of duration less than a few millions of years, as is predicted by the magma ocean model.

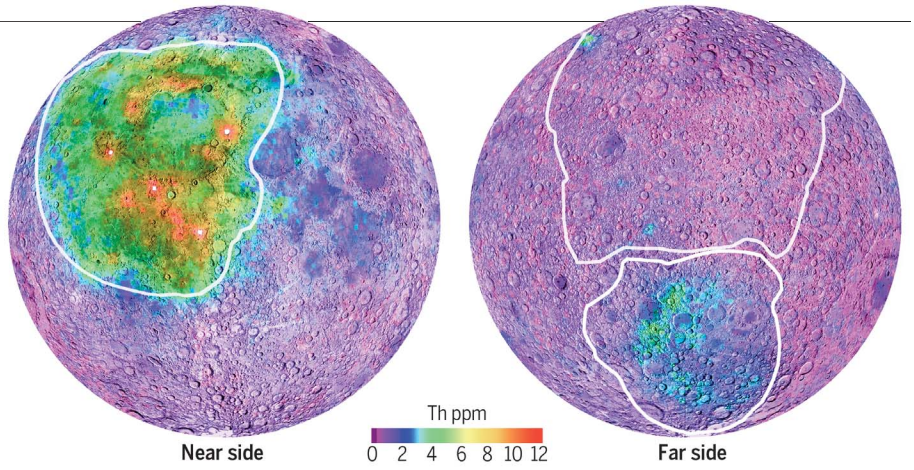



**Fig. 2. Map of the distribution of thorium (Th) on the lunar surface.** Data from the Lunar Prospector and Clementine missions (*82*) allow division of the lunar crust into three compositionally distinct terranes, outlined here in white. On the near side is the Th-rich Procellarum-KREEP Terrane (PKT); the far side is dominated by the Th-poor Feldspathic Highlands Terrane (top) penetrated by the impact that formed the South Pole–Aitken basin (bottom) that exposed slightly more Th-rich material, but nowhere near as Th-rich as in the PKT. [Figure provided by Bradley Jolliff updated and modified from the figure shown in (*12*)]

## An energetic birth

The magma ocean model implies that some energetic process involving high temperatures was involved in Moon formation. The nature of that process became clearer from comparison of the composition of lunar and terrestrial rocks. Early analyses of mare basalts were used to argue that the composition of the bulk Moon was similar to that of Earth's mantle (*32*). Earth, Mars, and their presumed meteoritic building blocks have subtle stable isotopic differences that reflect imperfect mixing of the contributions of the many stellar nucleosynthetic events in the galaxy that created the elements in our Solar System before the planets began to assemble (*33*–*35*). Earth and the Moon, however, appear to share identical isotopic composition (*36*, *37*). This holds even for tungsten (*38*), which tracks the timing of core formation due to the radioactive decay of $^{182}$Hf to $^{182}$W (*39*). The timing of core formation is unlikely to be the same on two bodies of such different size as Earth and the Moon. The Moon thus appears to share a very close genetic relationship with Earth.

The idea that the Moon formed from Earth dates back to at least 1879 (*40*, *41*) with the suggestion that the Moon was flung out by centrifugal force from a fast-spinning Earth. Planetary-scale fission of this nature likely requires much more energy than can be provided by spin alone. The compositional similarities of lunar and terrestrial rocks, however, revived the discussion of whether and how the Moon could have been derived from Earth. The focus turned rapidly to the question of whether a large impact into Earth could put enough material into stable Earth orbit to form the Moon (*42*). The answer appears to be "yes" (*42*–*44*), but exactly how is still being debated. Early models of the Moon-forming impact involved a glancing collision of the proto-Earth with an object roughly the size of Mars (*45*). Such an impact could place enough material into Earth orbit to form the Moon. The rapid reassembly of this material into a single body would result in a largely molten Moon, consistent with the evidence for a lunar magma ocean. Another attractive feature of the impact model is that it provides an explanation for the low bulk density of the Moon that is related to its small core, and hence its low iron/silicon ratio. Impact simulations show that the dense iron metal component is preferentially captured by Earth as the iron cores of both impactor and target merge (*46*, *47*). In a glancing-blow impact, however, the major fraction of the Moon is made up of material from the impactor's mantle, not Earth, thus requiring that the impactor was compositionally similar to Earth. This is not impossible, but neither is it likely given the randomness of planetary accretion and the timing of core formation on different planetary bodies. To address this issue, more recent giant-impact models explore still more energetic collisions (*46*, *47*). The most energetic of these (*48*) suggests that the outer layers of Earth expanded as a silicate vapor cloud of sufficient diameter to allow Moon formation from the condensed vapor as it cooled.

These more energetic impacts mix Earth's mantle and the impactor to the point where any isotopic distinctions between target and impactor before the impact would be no longer resolvable.

Although the role of giant impacts in more general aspects of planet formation remains unclear, the evidence for a lunar magma ocean and the Earth-Moon compositional similarity—both derived primarily from analysis of returned lunar samples—displaced the idea of cold, gentle planet assembly with one dominated by highly energetic impacts. As a result, the role of giant impacts is now an intrinsic part of models for planet formation (*49*), as are magma oceans as the initial stage of planet differentiation (*23*).

## Keeping hydrated

Another early indication that the Moon formed hot came from the depletion of moderately volatile elements—such as the alkali metals, halogens, and lead—in lunar samples relative to Earth (*32*), which is itself depleted in these elements relative to the Sun and primitive meteorites (*50*). Given the very high temperatures (several thousand kelvin) expected in a giant impact, more highly volatile compounds such as water were presumed to have been completely lost from the Moon (*51*). Hydrous minerals in lunar basalts are extremely rare, but whether this resulted from completely dry parental magmas or the loss of water from the magma during its eruption into the vacuum of the lunar surface was not clear. Deposits of volatile-driven pyroclastic volcanism—for example, the orange soils noticed by Apollo 17 astronaut Harrison Schmitt—were found at most Apollo sites and now have been mapped more widely from orbit (*52*). The identity of the gas fueling their explosive eruption is not clear. The water and $CO_2$ contents of samples of the pyroclastic glasses were too low to be quantified by the analytical tools available during the Apollo era. Improvements in the detection limits for hydrogen provided by secondary ion mass spectrometry allowed hydrogen to be detected in these pyroclastic materials (*53*). The moderate water contents so calculated were later supported by direct measurements on inclusions of glass formed when small amounts of melt were trapped inside minerals that had crystallized in the magma chamber prior to eruption (*54*). The melt inclusions were kept from outgassing by the crystal that encapsulates them. Related work on the mineral apatite, a calcium phosphate that can incorporate water into its crystal structure, also has provided evidence for moderate water contents in lunar magmas (*55*). The presence of water in lunar magmas has prompted reexamination of the fate of volatile elements during the giant impact (*56*). More generally, results from lunar sample analysis have driven a reinvestigation of the mechanism by which the terrestrial planets obtained water and carbon, with the implication that these volatile components may have been present when the terrestrial planets formed, rather than having been added later by accretion of icy bodies such as comets or icy asteroids (*57*).

## Sampling the space environment

The Moon-forming giant impact was not the last collision to be experienced by either Earth or the Moon. Only a few parts per million of the surface area of Earth dates to before 3.8 billion years ago, but most of the lunar surface is that old or older. The Moon's surface thus provides the record of bombardment from the era when impacts were a common force influencing planetary surfaces.

At the smallest scale, the Moon's surface is continually bombarded by energetic atoms from cosmic rays and the solar wind. The surface traps some of these atoms and thus serves as a long-term record of the composition of the solar wind (*51*, *58*–*60*). Examination of lunar samples showed the presence of surface coatings of iron and titanium, caused by irradiation with solar wind protons, which, coupled with the small grain sizes of the lunar regolith, contributed to the Moon's low surface reflectance (albedo) (*61*). This "space weathering" occurs on the surface of any Solar System body that is not protected by either an atmosphere or a magnetic field, with consequences for the interpretation of spectroscopic data for other airless bodies in the Solar System, particularly asteroids (*62*). Lunar "swirls," patterned color differences on the lunar surface that reflect focused particle irradiation directed by lunar magnetic fields, are another expression of this phenomenon (*51*, *63*). Recent measurements of the remanent magnetism in lunar samples suggest that the Moon had a global magnetic field of a few tens of microteslas (Earth's magnetic field ranges from about 25 to 65 microteslas at the surface) from at least 4.25 to 2.5 billion years ago (*64*).

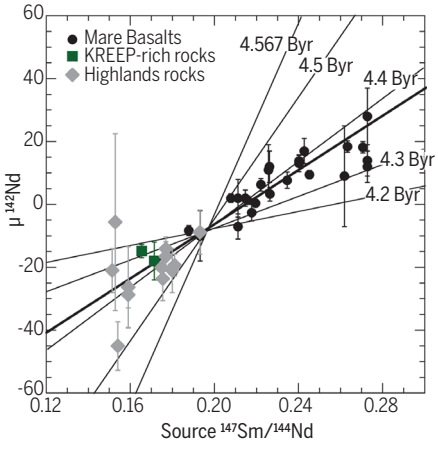



**Fig. 3. Illustration of a common source history for lunar igneous rocks.** The neutron fluence–corrected $^{142}Nd/^{144}Nd$ ($\mu^{142}Nd$ is the deviation in the sample's $^{142}Nd/^{144}Nd$ relative to the laboratory standard in parts per million) is plotted against the $^{147}Sm/^{144}Nd$ ratio of the sample (highlands rocks) or calculated for the source of the basalts according to their initial $^{143}Nd/^{144}Nd$ ratio. Both Nd isotope ratios are modified by the radioactive decay of $^{146}Sm$ (half-life 103 million years) to $^{142}Nd$ and that of $^{147}Sm$ (half-life 106 billion years) to $^{143}Nd$. The lines on the diagram show the slopes expected for different ages (Byr, billion years) for the process that created the range in Sm/Nd ratios that resulted in the initial isotopic range in these rocks. The alignment of the data for mare basalts, KREEP-rich rocks, and various highland samples along a single line of slope corresponding to an age of 4.38 billion years indicates that the source materials of these lavas formed during a single short-duration event. [Data from (*27*, *29*, *83*, *84*)]

Kaguya (*65*) allowed calibration of the lunar impact flux through time (Fig. 4). This flux estimate can be extrapolated throughout the inner Solar System, allowing the cratering record for other planets and moons to be turned from relative to absolute chronologies for their surfaces (*66*). Such data provide basic information on the rate of planetary resurfacing either destructively by erosion, or constructively by volcanism, with the latter providing key information on the dynamics of planetary interiors.

Individual lunar craters have had their ages determined via geochronologic studies of lunar breccias and particularly impact melts. Unexpectedly, ages near 3.9 billion years occurred repeatedly (*67*). The same age was also seen in a number of lunar highlands rocks (*68*). This result was used to suggest that several of the large lunar basins on the Moon's nearside formed at roughly the same time because of a large increase in the influx rate of bombarding planetesimals. The idea of a "terminal lunar cataclysm" or "late heavy bombardment," as this event has been called (*69*), led to a wide variety of hypothesized causes and consequences including orbital migration of the gas giant planets (*70*), an explanation for the limited amount of preserved terrestrial crust older than 3.9 billion years (*71*), and the delayed start to life on Earth (*72*, *73*).

Recent analyses of lunar and meteoritic samples have questioned whether the late heavy bombardment is required by the lunar impact data (*74*). The debate is informed by two approaches. One is the measurement in lunar impact rocks of the abundance of a group of elements, called the highly siderophile elements (siderophile indicates a solubility preference for iron metal rather than silicate), which are present at $10^{-6}$ g/g or higher levels in meteorites, but at $10^{-9}$ g/g or lower abundances in planetary mantles that have lost most of these elements to the core. The goal of this approach is to use the abundance patterns of these elements in lunar rocks to discriminate the types of impactors involved and thus distinguish one impact from another (*75*–*78*). An impact the size of a major basin-forming event (e.g., the Imbrium

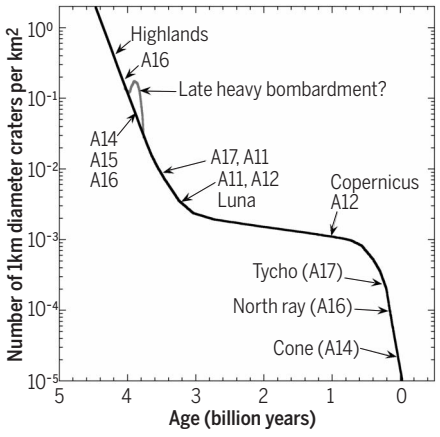



**Fig. 4. Impact rate on the lunar crust through time.** Calibration points are based on crater densities from different age terranes sampled at the Apollo (designated by A) and Luna 16 landing sites. The gray peak at 3.9 billion years represents the hypothesized spike in impacts associated with the late heavy bombardment. [Figure adapted from (*85*)]

basin) will scatter ejecta over a good portion of the lunar surface, so associating any particular surface-collected rock formed by impact with the crater that ejected it is not straightforward. The other approach examines the probability that the record of older impacts is overprinted by younger impacts until the rate of impact falls below some critical level (*69*), in which case the "late heavy bombardment" may simply reflect the end of a rapidly declining rate of impact (Fig. 4). Humans conducted field work on the Moon for less than 12 days, covering a very limited portion of the Earth-facing side of the Moon. Refining the chronologic sequence of large impact events on the Moon requires more detailed field studies conducted in future visits that cover a larger portion of the lunar surface.

## Lunar samples never stop giving

Many of the discoveries discussed above were made shortly after the lunar samples became available for analysis. Others, however, were delayed until analytical techniques, and the knowledge base needed to ask new questions, advanced. Examples of the long-term yield from lunar samples are the determination of their water content, continuing refinements in the ages of lunar samples and particularly the evidence for an early lunar differentiation event, and measurements that show a long-lived lunar magnetic field. Some outstanding problems likely will require new sample collection efforts—for example, determining the true age range of the lunar crust and hence the age of the Moon, and resolving whether the late heavy bombardment reflected a restricted interval of enhanced impacts or the tail of a rapidly declining impact flux. With the still improving ability to make measurements on smaller samples, another opportunity lies in exploring the archived quantity of lunar samples. Interior images of breccias using x-ray tomography (*79*) can search for rock clasts within the breccias that might be of use for additional age determinations, particularly of the ancient lunar crust. Determining whether the data obtained from Apollo and Luna samples can be extrapolated to the whole Moon, or whether they constitute a biased record heavily influenced by the nearside basin-forming events, can only be addressed by sample returns from new locations well removed from the Apollo and Luna sites. The recent Chang'e 4 landing in the South Pole–Aitken basin on the far side of the Moon has returned data suggesting that this impact penetrated into the lunar mantle (*80*). Much can be learned by remote analysis, and indeed the many lunar orbital missions conducted since Apollo inform the choices for where to sample next, but many of the results discussed above required the most advanced analysis possible in terrestrial laboratories, and hence the return of lunar samples to Earth (*81*).

## REFERENCES AND NOTES


1. H. C. Urey, *Geochim. Cosmochim. Acta* **1**, 209–277 (1951).
2. J. A. Wood, J. S. Dickey Jr., U. B. Marvin, B. N. Powell, in *Proceedings of the Apollo 11 Lunar Science Conference* (Pergamon, 1970), pp. 965–988.
3. J. V. Smith, A. T. Anderson, R. C. Newton, E. J. Olsen, P. J. Wyllie, in *Proceedings of the Apollo 11 Lunar Science Conference* (Pergamon, 1970), pp. 897–925.
4. M. A. Wieczorek *et al*., *Science* **339**, 671–675 (2013).
5. C. K. Shearer *et al*., *Rev. Mineral. Geochem.* **60**, 365–518 (2006).
6. G. A. Snyder, L. A. Taylor, C. R. Neal, *Geochim. Cosmochim. Acta* **56**, 3809–3823 (1992).
7. D. Walker, J. Longhi, J. F. Hays, in *Proceedings of the 6th Lunar Science Conference* (Pergamon, 1975), pp. 1103–1120.
8. M. A. Wieczorek *et al*., *Rev. Mineral. Geochem.* **60**, 221–364 (2006).
9. S. M. Elardo, D. S. Draper, C. K. Shearer Jr., *Geochim. Cosmochim. Acta* **75**, 3024–3045 (2011).
10. P. H. Warren, J. T. Wasson, *Rev. Geophys. Space Phys.* **17**, 73–88 (1979).
11. D. J. Lawrence *et al*., *Science* **281**, 1484–1489 (1998).
12. B. L. Jolliff, J. Gillis, L. Haskin, R. Korotev, M. Wieczorek, *J. Geophys. Res.* **105**, 4197–4216 (2000).
13. T. H. Prettyman *et al*., *J. Geophys. Res. Planets* **111**, E12007 (2006).
14. A. C. Quillen, L. Martini, M. Nakajima, *Icarus* **329**, 182–196 (2019).
15. A. Roy, J. T. Wright, S. Sigurðsson, *Astrophys. J.* **788**, L42 (2014).
16. M. Laneuville, M. A. Wieczorek, D. Breuer, N. Tosi, *J. Geophys. Res. Planets* **118**, 1435–1452 (2013).
17. K. Miljković *et al*., *Science* **342**, 724–726 (2013).
18. A. E. Ringwood, S. E. Kesson, in *Proceedings of the 7th Lunar Science Conference* (Pergamon, 1976), pp. 1697–1722.
19. P. C. Hess, E. M. Parmentier, *Earth Planet. Sci. Lett.* **134**, 501–514 (1995).
20. L. T. Elkins Tanton, J. A. V. Orman, B. H. Hager, T. L. Grove, *Earth Planet. Sci. Lett.* **196**, 239–249 (2002).
21. J. Gross, A. H. Treiman, C. N. Mercer, *Earth Planet. Sci. Lett.* **388**, 318–328 (2014).
22. C. R. Neal, L. A. Taylor, *Geochim. Cosmochim. Acta* **56**, 2177–2211 (1992).
23. L. T. Elkins-Tanton, *Annu. Rev. Earth Planet. Sci.* **40**, 113–139 (2012).
24. V. S. Solomatov, in *Treatise on Geophysics, vol. 9*, D. J. Stevenson, Ed. (Elsevier, Amsterdam, 2007), pp. 91–119.
25. L. E. Borg, A. M. Gaffney, C. K. Shearer, *Meteorit. Planet. Sci.* **50**, 715–732 (2015).
26. L. E. Nyquist, C. Y. Shih, *Geochim. Cosmochim. Acta* **56**, 2213–2234 (1992).
27. M. Boyet, R. W. Carlson, *Earth Planet. Sci. Lett.* **262**, 505–516 (2007).
28. L. E. Nyquist *et al*., *Geochim. Cosmochim. Acta* **59**, 2817–2837 (1995).
29. C. L. McLeod, A. D. Brandon, R. M. G. Armytage, *Earth Planet. Sci. Lett.* **396**, 179–189 (2014).
30. A. M. Gaffney, L. E. Borg, *Geochim. Cosmochim. Acta* **140**, 227–240 (2014).
31. P. Sprung, T. Kleine, E. E. Scherer, *Earth Planet. Sci. Lett.* **380**, 77–87 (2013).
32. A. E. Ringwood, S. E. Kesson, in *Proceedings of the 8th Lunar Science Conference* (Pergamon, 1977), pp. 371–398.
33. R. N. Clayton, N. Onuma, T. K. Mayeda, *Earth Planet. Sci. Lett.* **30**, 10–18 (1976).
34. M. Fischer-Gödde, C. Burkhardt, T. S. Kruijer, T. Kleine, *Geochim. Cosmochim. Acta* **168**, 151–171 (2015).
35. P. H. Warren, *Earth Planet. Sci. Lett.* **311**, 93–100 (2011).
36. L. Qin, C. M. O. D. Alexander, R. W. Carlson, M. F. Horan, T. Yokoyama, *Geochim. Cosmochim. Acta* **74**, 1122–1145 (2010).
37. E. D. Young *et al*., *Science* **351**, 493–496 (2016).
38. M. Touboul, I. S. Puchtel, R. J. Walker, *Nature* **520**, 530–533 (2015).
39. A. N. Halliday, T. Kleine, in *Meteorites and the Early Solar System II* (Univ. of Arizona Press, 2005), pp. 775–801.
40. G. H. Darwin, *Philos. Trans. R. Soc. London* **170**, 447–538 (1879).
41. S. G. Brush, in *Origin of the Moon* (Lunar and Planetary Institute, 1986), pp. 3–15.
42. W. K. Hartmann, D. R. Davis, *Icarus* **24**, 504–515 (1975).
43. D. J. Stevenson, *Annu. Rev. Earth Planet. Sci.* **15**, 271–315 (1987).
44. W. Benz, W. L. Slattery, A. G. W. Cameron, *Icarus* **71**, 30–45 (1987).
45. R. M. Canup, E. Asphaug, *Nature* **412**, 708–712 (2001).
46. R. M. Canup, *Science* **338**, 1052–1055 (2012).
47. M. Ćuk, S. T. Stewart, *Science* **338**, 1047–1052 (2012).
48. S. J. Lock *et al*., *J. Geophys. Res. Planets* **123**, 910–951 (2018).
49. H. Genda, Y. Abe, *Science* **164**, 149–162 (2003).
50. H. Palme, H. S. C. O'Neill, in *Treatise on Geochemistry, vol. 3*, R. W. Carlson, Ed. (Elsevier, 2014), pp. 1–39.
51. P. Lucey *et al*., *Rev. Mineral. Geochem.* **60**, 83–219 (2006).
52. J. O. Gustafson, J. F. Bell III, L. R. Gaddis, B. R. Hawke, T. A. Giguere, *J. Geophys. Res. Planets* **117**, E00H25 (2012).
53. A. E. Saal *et al*., *Nature* **454**, 192–195 (2008).
54. E. H. Hauri, T. Weinreich, A. E. Saal, M. C. Rutherford, J. A. Van Orman, *Science* **333**, 213–215 (2011).
55. F. M. McCubbin *et al*., *Am. Mineral.* **100**, 1668–1707 (2015).
56. E. H. Hauri *et al*., *Annu. Rev. Earth Planet. Sci.* **45**, 89–111 (2017).
57. C. M. O. Alexander, *Philos. Trans. R. Soc. London Ser. A* **375**, 20150384 (2017).
58. I. A. Crawford, S. A. Fagents, K. H. Joy, M. E. Rumpf, *Earth Moon Planets* **107**, 75–85 (2010).
59. S. A. Fagents, M. E. Rumpf, I. A. Crawford, K. H. Joy, *Icarus* **207**, 595–604 (2010).
60. R. Wieler, *Chem. Erde* **76**, 463–480 (2016).
61. T. Gold, E. Bilson, R. L. Baron, *J. Geophys. Res.* **82**, 4899–4908 (1977).
62. C. R. Chapman, *Meteoritics* **31**, 699–725 (1996).
63. D. J. Hemingway, S. Tikoo, *J. Geophys. Res. Planets* **123**, 2223–2241 (2018).
64. S. M. Tikoo *et al*., *Sci. Adv.* **3**, e1700207 (2017).
65. D. Stöffler *et al*., *Rev. Mineral. Geochem.* **60**, 519–596 (2006).
66. D. Stöffler, G. Ryder, in *Chronology and Evolution of Mars* (Springer, 2001), pp. 9–54.
67. G. Turner, P. H. Cadogan, C. J. Yonge, in *Proceedings of the 4th Lunar Science Conference* (Pergamon, 1973), pp. 1889–1914.
68. F. Tera, D. A. Papanastassiou, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **22**, 1–21 (1974).
69. W. F. Bottke, M. D. Norman, *Annu. Rev. Earth Planet. Sci.* **45**, 619–647 (2017).
70. R. Gomes, H. F. Levison, K. Tsiganis, A. Morbidelli, *Nature* **435**, 466–469 (2005).
71. O. Abramov, D. A. Kring, S. J. Mojzsis, *Chem. Erde* **73**, 227–248 (2013).
72. K. A. Maher, D. J. Stevenson, *Nature* **331**, 612–614 (1988).
73. K. Zahnle *et al*., *Space Sci. Rev.* **129**, 35–78 (2007).
74. N. E. B. Zellner, *Orig. Life Evol. Biosph.* **47**, 261–280 (2017).
75. J. Gros, H. Takahashi, J. Hertogen, J. W. Morgan, E. Anders, in *Proceedings of the 7th Lunar Science Conference* (Pergamon, 1976), pp. 2403–2425.
76. M. D. Norman, V. C. Bennett, G. Ryder, *Earth Planet. Sci. Lett.* **202**, 217–228 (2002).
77. I. S. Puchtel, R. J. Walker, O. B. James, D. A. Kring, *Geochim. Cosmochim. Acta* **72**, 3022–3042 (2008).
78. M. Fischer-Gödde, H. Becker, *Geochim. Cosmochim. Acta* **77**, 135–156 (2012).
79. R. D. Hanna, R. A. Ketcham, *Chem. Erde* **77**, 547–572 (2017).
80. C. Li *et al*., *Nature* **569**, 378–382 (2019).
81. I. A. Crawford *et al*., *Planet. Space Sci.* **74**, 3–14 (2012).
82. J. J. Gillis, R. L. Korotev, B. L. Jolliff, *Geochim. Cosmochim. Acta* **68**, 3791–3805 (2004).
83. A. D. Brandon *et al*., *Geochim. Cosmochim. Acta* **73**, 6421–6445 (2009).
84. M. Boyet, R. W. Carlson, L. E. Borg, M. Horan, *Geochim. Cosmochim. Acta* **148**, 203–218 (2015).
85. H. Hiesinger, J. W. Head III, U. Wolf, R. Jaumann, G. Neukum, *J. Geophys. Res.* **115**, E03003 (2010).


## ACKNOWLEDGMENTS

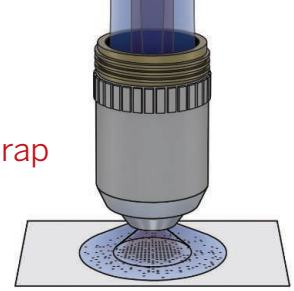

I thank two anonymous reviewers for suggestions that substantially improved the paper. **Funding:** Writing of the paper was supported by the Carnegie Institution for Science. **Competing interests:** None.

10.1126/science.aaw7580

# RESEARCH

Optical lattices trap quantum gases

*Chiu* et al., *p. 251*



## IN *SCIENCE* JOURNALS

*Edited by* **Stella Hurtley**

Bárðarbunga Volcano in the Holuhraun lava field in Iceland

VOLCANOLOGY

### Millennial magma reserves

Lava erupts after being stored as magma deeper underground. Silica-rich magmas characteristic of volcanoes like Pinatubo are stored for thousands to hundreds of thousands of years. However, we have little understanding of the time scales for deeper basaltic magma like that supplying Icelandic volcanoes. Mutch *et al*. used diffusion of chromium and aluminum atoms to show that these magmas are stored for hundreds to thousands of years at the crust-mantle boundary. This discovery elucidates the time scale for understanding how magma is created and stored and how it erupts. —BG

*Science*, this issue p. 260

METASURFACES

### A dynamic metasurface in the mix

Metasurfaces have been designed and fabricated to function as many passive optical elements. A time-dependent active element can now be obtained by marrying concepts in microelectromechanics with an array of one-dimensional dielectric nanowires. Although such a static array has been used to select for desired wavelengths, the wavelength is dependent on the nanowire-to-substrate separation. Holsteen *et al.* added a voltage-controlled actuation element to the design that could steer light beams as well as select and mix certain colors. The fast actuation capability could affect applications ranging from autonomous vehicles to augmented and virtual reality systems. —ISO

*Science*, this issue p. 257

IMMUNOTHERAPY

### Targeted tumor immunotherapy

Although immunotherapy has transformed the cancer therapeutics landscape, a number of problems remain to be solved, from improving efficacy to limiting side effects. Wang *et al.* engineered nanoparticles that could be activated specifically in tumors. They conjugated antibodies against programmed death ligand 1 (PDL1) with matrix metalloproteinase protein 2 (MMP-2)–sensitive nanoparticles carrying a photosensitizer. MMP-2 is highly expressed in tumors, and delivery of the nanoparticle to tumors elicited release of the antibody from the nanoparticle. When used in conjunction with localized near-infrared radiation that activated the photosensitizer to produce reactive oxygen species, the nanoparticles outperformed systemic anti-PDL1 in limiting growth and metastasis of murine tumors. —AB

*Sci. Immunol.* **4**, eaau6584 (2019).

QUANTUM GASES

### A quantum breakdown

At low temperatures, two-dimensional (2D) systems with contact interactions are expected to exhibit quantum anomalies—a breakdown of scaling laws that characterize such systems in the classical regime. Signatures of these anomalies have been observed in the real-space properties of 2D Fermi gases, but the effect is much less pronounced than expected on theoretical grounds. Murthy *et al.* studied the momentum-space profiles of 2D superfluids of fermionic atoms. They initially perturbed the gas and then monitored the momentum distribution of its atoms. In the regime of strong interactions between the atoms, the momentum profiles deviated markedly from the classical scaling. —JS

*Science*, this issue p. 268

CREDITS: (TOP TO BOTTOM) CHIU *ET AL.*; ARCTIC IMAGES/ALAMY STOCK PHOTO

MARINE MANAGEMENT

## Predicting marine futures

The ability to predict how climate variations will affect marine ecosystems would allow better economic and ecosystem planning and management. Park *et al.* found that a global Earth system model skillfully predicted seasonal to multiannual ocean chlorophyll fluctuations in many regions. This could allow annual fish catches in some regions to be forecast 2 to 3 years in advance. —HJS

*Science*, this issue p. 284

NATURAL PRODUCTS

## Growing independently

Bacteria produce natural products with a range of functions. A major family comprises ribosomally synthesized and posttranslationally modified peptides. Ting *et al.* identify a biosynthetic pathway in which a natural product is derived from an amino acid that is added to a ribosomally synthesized peptide independent of the ribosome. This biosynthetic paradigm is used in the synthesis of thiaglutamate and ammosamides, and the finding of related gene clusters suggests that the strategy may be used more widely. —VV

*Science*, this issue p. 280

VACCINES

## Building a safer CMV vector

Vaccine vectors based on cytomegalovirus (CMV) show strong T cell induction and protection against a multitude of pathogens. However, such vaccines may not be safe for people who are immunodeficient or immunosuppressed. Marshall *et al.* genetically modified rhesus CMV to allow engagement of host intrinsic immunity. The modified ΔRh110 vector did not spread once administered to nonhuman primates but still induced robust T cell immunity. Hansen *et al.* showed in a simian immunodeficiency virus challenge model that the ΔRh110 vector provided durable protection that was comparable to the parental vector. Mutations in the human CMV vector could thus yield a potent but restrained CMV for human vaccination. —LP

*Sci. Transl. Med.* **11**, eaaw2603, eaaw2607 (2019).

MOLECULAR KNOTS

## Carbon catenation

Preparing interlocked rings and knots at the molecular scale traditionally relies on preorientation of the building blocks by nitrogen or oxygen substituents. Segawa *et al.* devised a distinct strategy to synthesize catenane and trefoil structures composed exclusively of carbon and hydrogen (see the Perspective by Van Raden and Jasti). They linked phenyl rings end to end into macrocycles that met in the middle at a silicon center. Excision of the silicon with fluoride then yielded the interlocked products. —JSY

*Science*, this issue p. 272; see also p. 216

PALEONTOLOGY

## Early suckler?

One trait that is unique to mammals is milk suckling. Suckling requires the presence of stability and motion in the throat, both of which require a complex hyoid apparatus. Zhou *et al.* describe a mammaliform docodontan fossil from the Jurassic that was preserved with a nearly intact hyoid (see the Perspective by Hoffmann and Krause). The structure is complex and saddle shaped, like that seen in modern mammals, suggesting that a muscularized throat was present before the development of mammals. —SNV

*Science*, this issue p. 276; see also p. 222

# IN OTHER JOURNALS

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

NANOMATERIALS

## Glimpsing growing nanotubes

One-dimensional metal-organic frameworks, also called metal–organic nanotubes (MONTs), are similar to carbon nanotubes but have a wider design space of size and chemical functionality. However, there have been few examples of MONTs, in part because little is known about their growth mechanism. Vailonis *et al.* synthesized a ligand, 1,4-bis((4H-1,2,4-triazol-4-yl)methyl)naphthalene, that formed a MONT after 5 days of reaction with copper(II) bromide. They also followed this process with liquid-cell transmission electron microscopy and could follow the formation of fiber bundles after initial nucleation. Fitting of this anisotropic growth process to a Lifshitz-Slyozov-Wagner model was consistent with the growth being under thermodynamic control and occurring through a surface-specific monomer–monomer attachment process. Further structural analysis of the material showed that the cell material was the same as the bulk. —PDS

*J. Am. Chem. Soc.* **141**, 10177 (2019).

**Transmission electron micrograph taken near the start of formation of copper metal–organic nanotubes**

CREDIT: K. VAILONIS *ET AL.*, *J. AM. CHEM. SOC.* **141**, 10177 (2019)

EDUCATION

## Confronting bias before it happens

Student evaluations of teaching (SETs) remain the primary way to evaluate teaching effectiveness for tenure packages, even as evidence documenting bias, i.e., that female instructors tend to be evaluated more critically than males, in this type of evaluation grows. Peterson *et al.* added specific language, designed to mitigate gender bias, to certain SETs in a randomized experiment of four classes with large enrollment. Students in the antibias group ranked female instructors significantly higher than did

students in the standard treatment, with no differences seen for male instructors. Although these effects may be magnified by the unusual nature of the situation for the students, they do suggest that universities should consider adopting some form of antibias language. If students begin to see this language across all SETs, effects may be minimized. —MMc

*PLOS ONE* **14**, e0216241 (2019).

IMMUNOLOGY

## A is for antimicrobial

Vitamin A is an essential fat-soluble micronutrient that regulates immune function through its derivative, retinoic acid. One role of retinoic acid is to control skin infection and inflammation. Harris *et al.* report that dietary vitamin A regulates the expression of the antimicrobial protein RELMα in the skin. RELMα, which is induced by the skin microbiota, can kill bacteria by disrupting their membranes. Mice lacking RELMα have an altered skin microbiota composition and an enhanced susceptibility to certain bacterial infections. The work shows how vitamin A analogs so effectively treat skin conditions such as acne and psoriasis. —STS

*Cell Host Microbe* **25**, 777 (2019).

CELL BIOLOGY

## Inhibition by misdirection

The endoplasmic reticulum (ER) is the main entry point to the cellular secretory pathway. It is a major site of folding and quality control of newly synthesized proteins. When protein folding goes awry, the unfolded protein response (UPR) is activated to protect the cell from harmful effects of aberrant proteins. The UPR is triggered by an ER-tethered transcription factor, ATF6α. Small-molecule inhibitors known as Ceapins specifically inhibit ATF6α by retaining it at the ER. Torres *et al.* performed a genome-wide CRISPR interference screen to elucidate how Ceapins work. They found that Ceapins act via the ABCD3 peroxisomal transporter. In the presence of Ceapins, ABCD3 binds to ATF6α, which means the ER becomes tethered to peroxisomes and prevents ATF6α from leaving the ER and carrying out its function. —SMH

*eLife* **8**, e46595 (2019).

CRIMINOLOGY

## Body cameras and police misconduct

Following high-profile cases of police misconduct, there has been increasing public support for the use of police body-worn cameras but limited evidence about their effectiveness. Yokum *et al.* conducted a randomized control trial of body camera use involving more than 2000 officers in the Washington, D.C., police department. They found no significant effects of wearing body cameras on use of force, civilian complaints, police activity, or judicial outcomes. These findings suggest that large-scale employment of police body cameras may not solve the challenge of police misconduct on its own. —TSR

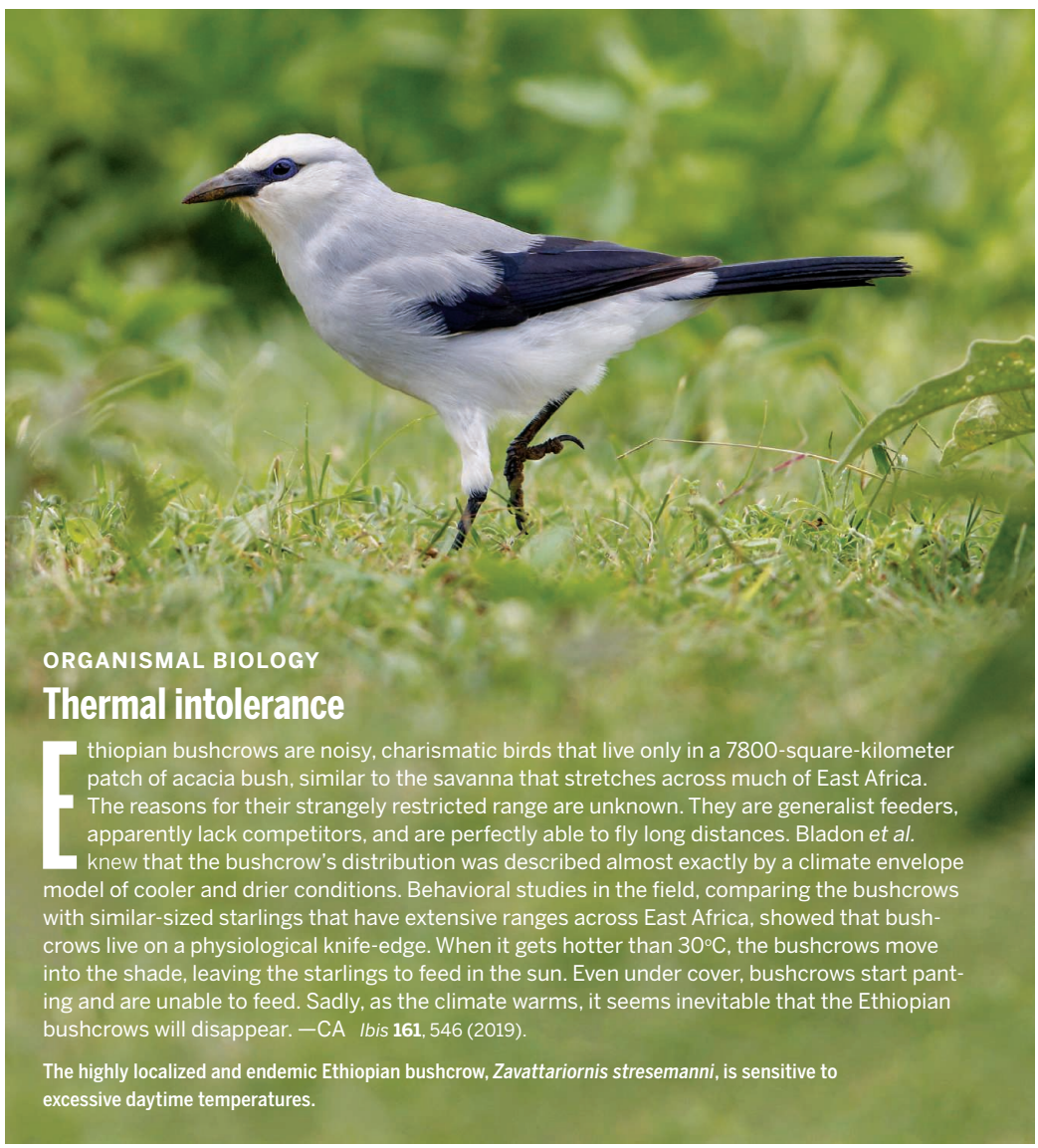
*Proc. Natl. Acad. Sci. U.S.A.* **116**, 10329 (2019).

NEUROSCIENCE

## Re-skilling the brain

We still do not understand how changes in the brain during learning lead to the acquisition of new skills. Brain-computer interfaces (BCIs) can be used to link neural activity to a response visible on a computer monitor. Oby *et al.* used a BCI with monkeys and found that new neural activity patterns emerged in the motor cortex of monkeys that learned to control the movement of a cursor. Learning to use a BCI is thus associated with new neural activity patterns in the same way as learning to use handheld tools or learning to play the piano. This work shows a direct link between neural learning and behavioral learning, which previously has only been surmised from neural correlates. —PRS

*Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1820296116 (2019).



ORGANISMAL BIOLOGY

## Thermal intolerance

Ethiopian bushcrows are noisy, charismatic birds that live only in a 7800-square-kilometer patch of acacia bush, similar to the savanna that stretches across much of East Africa. The reasons for their strangely restricted range are unknown. They are generalist feeders, apparently lack competitors, and are perfectly able to fly long distances. Bladon *et al.* knew that the bushcrow's distribution was described almost exactly by a climate envelope model of cooler and drier conditions. Behavioral studies in the field, comparing the bushcrows with similar-sized starlings that have extensive ranges across East Africa, showed that bushcrows live on a physiological knife-edge. When it gets hotter than 30°C, the bushcrows move into the shade, leaving the starlings to feed in the sun. Even under cover, bushcrows start panting and are unable to feed. Sadly, as the climate warms, it seems inevitable that the Ethiopian bushcrows will disappear. —CA *Ibis* **161**, 546 (2019).

**The highly localized and endemic Ethiopian bushcrow, *Zavattariornis stresemanni*, is sensitive to excessive daytime temperatures.**

PHOTO: FLPA/ALAMY STOCK PHOTO

## ALSO IN *SCIENCE* JOURNALS

*Edited by* **Stella Hurtley**

INFLAMMATORY PAIN

### A "sUPR" target for pain management?

The unfolded protein response (UPR) is initiated when unfolded or misfolded proteins accumulate in the endoplasmic reticulum. One highly conserved arm of the UPR, the IRE1α–XBP1 signaling pathway, also plays a role in various other UPR-independent processes, including hypoxia, angiogenesis, and inflammation. Chopra *et al.* report that this pathway additionally regulates the production of two molecules, cyclooxygenase 2 and microsomal prostaglandin E synthase 1, that help mediate inflammation-induced pain (see the Perspective by Avril and Chevet). When elements of the IRE1α–XBP1 signaling pathway were knocked out, pain behaviors were reduced in two different mouse models of pain. Targeting this pathway may result in improved pain management therapies. —STS

*Science*, this issue p. 248; see also p. 224

COMPARATIVE GENETICS

### The genetics of sexual dimorphism

In mammals, many species exhibit sex-specific phenotypes that differ between males and females. Although attention has been directed to the effects of the X and Y sex chromosomes, we do not understand how sex affects the rest of the genome. Naqvi *et al.* examined gene expression in 12 tissues in male and female humans, mice, rats, dogs, and cynomolgus macaques and identified diversity in gene expression between the sexes. Examining sex-biased gene expression in human height identified opposing male or female bias. Although conservation of differential sex-specific gene expression among species was observed, specific genes differed in the sexes among species and lineages suggesting the evolution of species- or lineage-specific sex-biased expression. —LMZ

*Science*, this issue p. 249

ALZHEIMER'S DISEASE

### Pericytes put the squeeze on cognition

Like a computer, the brain needs a reliable source of power, which is provided as oxygen and glucose in the blood. However, in many neurological disorders this energy supply is disrupted. Brain blood flow is controlled by adjustment of the diameters of the vessels supplying the blood. Nortley *et al.* found that, both in humans developing Alzheimer's disease (AD) and in a mouse model of AD, brain capillaries become squeezed by pericytes (see the Perspective by Liesz). By defining the underlying mechanism, they suggest potential targets for therapy in early AD. —SMH

*Science*, this issue p. 250; see also p. 223

MAGNETIC MATERIALS

### Liquid reconfigurable ferromagnetic materials

Ferromagnetic materials show a permanent magnetic dipole, whereas superparamagnetic ones only show magnetic properties under an applied field. Some materials, like ferrofluids, show liquid-like behavior but do not retain their magnetization in the absence of an applied field. Liu *et al.* show remnant magnetization of otherwise superparamagnetic magnetite nanoparticles at an oil-water interface of emulsion droplets (see the Perspective by Dreyfus). The permanent magnetization could be controlled by coupling and uncoupling the magnetization of individual nanoparticles, making it possible to "write and erase" shapes of the droplets or to elongate them into cylinders. —MSL

*Science*, this issue p. 264; see also p. 219

QUANTUM SIMULATION

### Looking for patterns in an optical lattice

One of the simplest models of interacting fermions on a two-dimensional (2D) lattice—the Hubbard model—becomes too tricky to simulate on classical computers as the density of empty lattice sites (holes) increases. Chiu *et al.* used a quantum microscope to take snapshots of thousands of realizations of the 2D Hubbard model in an optical lattice filled with fermionic lithium atoms at varying hole densities (see the Perspective by Schauss). The authors used pattern recognition algorithms to analyze the images, in which each lattice site was individually resolved. Comparing these patterns to the predictions of several theoretical models, they found the most consistency with the so-called geometric string model. —JS

*Science*, this issue p. 251; see also p. 218

MEDICINE

### Broadening targeted therapy in cystic fibrosis

There are more than 1700 mutations in the CFTR (cystic fibrosis transmembrane conductance regulator) gene that cause cystic fibrosis. Some mutations occur frequently, whereas others affect very few or individual patients. Modulators have been developed to target specific CFTR mutations classified according to their functional impact on the encoded protein. In a Perspective, Manfredi *et al.* discuss the emerging view that many CFTR mutations have pleiotropic effects and so more patients could benefit from modulator therapy but do not receive it. Moreover, individuals with ultrarare CFTR mutations are often not indicated to receive these targeted drugs. The authors outline the approaches needed to broaden the personalization of these modulators to treat more patients. —GKA

*Science*, this issue p. 220

SOCIAL SCIENCES

### Doesn't matter if it's fair...as long as you win

In a world of increasing economic inequality, how does the public perceive the causes and fairness of wealth disparity? To measure the effect of unequal outcomes on individual perceptions of fairness and inequality, Molina *et al.* devised a card game that separates the equality of opportunity from game outcomes, player skill, and luck. In this setting, players who were dealt a great hand were more likely to (incorrectly) attribute their winnings to their own talent rather than to external factors such as luck or structural advantages. Winning players were also more likely to view the game as fair and feel positively about the game, even when the rules were tilted in their favor. Thus, perceptions of unequal outcomes, such as wealth inequality, are biased, making efforts to equalize opportunities in social or economic systems complicated. As it turns out, "it's not just how the game is played, it's whether you win or lose." —AC

*Sci. Adv.* 10.1126/sciadv.aau1156 (2019).

CANCER

### An Achilles' heel for KRAS mutant tumors

Mutations in the *KRAS* gene drive several common and deadly cancers. Unfortunately, most mutant KRAS proteins cannot be targeted therapeutically. Because KRAS-mutant pancreatic cancers rely on the

PHOTO: CREDIT GOES HERE AS SHOWN; CREDIT GOES HERE AS SHOWN

transcription factor MYC, Blake *et al.* screened for inhibitors that decreased MYC protein abundance. An inhibitor of the cell cycle–associated kinase CDK9 decreased MYC levels independently of KRAS signaling. This finding reveals a potential therapeutic target for patients with KRAS-mutant pancreatic cancers and perhaps also those with MYC-dependent cancers. —LKF

*Sci. Signal.* **12**, eaav7259 (2019).

PHOTO: CREDIT GOES HERE AS SHOWN; CREDIT GOES HERE AS SHOWN

RESEARCH ARTICLE SUMMARY

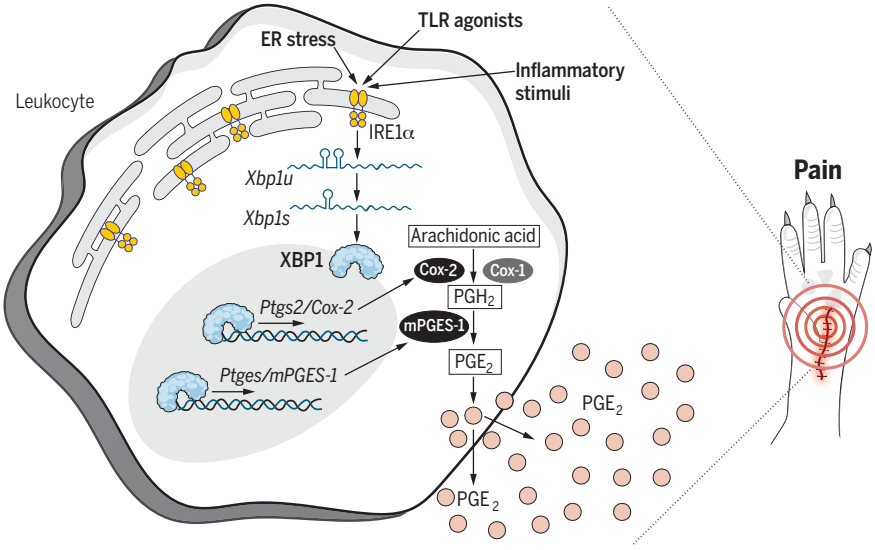
INFLAMMATORY PAIN

# IRE1α–XBP1 signaling in leukocytes controls prostaglandin biosynthesis and pain

**Sahil Chopra, Paolo Giovanelli, Perla Abigail Alvarado-Vazquez, Sara Alonso, Minkyung Song, Tito A. Sandoval, Chang-Suk Chae, Chen Tan, Miriam M. Fonseca, Silvia Gutierrez, Leandro Jimenez, Kotha Subbaramaiah, Takao Iwawaki, Philip J. Kingsley, Lawrence J. Marnett, Andrew V. Kossenkov, Mariano Sanchez Crespo, Andrew J. Dannenberg, Laurie H. Glimcher*, E. Alfonso Romero-Sandoval*, Juan R. Cubillos-Ruiz***

**INTRODUCTION:** Tissue injury triggers rapid local responses coordinated by immune cells, which dictate the maintenance and resolution of inflammation and therefore the recovery from functional impairment and pain. This inflammatory process requires high levels of protein synthesis, folding, modification, and trafficking, which are events regulated by the endoplasmic reticulum (ER). Excessive protein synthesis and handling can lead to the accumulation of misfolded proteins in this organelle, provoking a cellular state of "ER stress" and subsequent activation of the unfolded protein response (UPR). The IRE1α–XBP1 signaling pathway is an evolutionarily conserved branch of the UPR that maintains ER homeostasis while simultaneously governing various immunometabolic processes. Yet, the physiological consequences of IRE1α–XBP1 signaling in leukocytes during tissue injury and inflammation remain largely unexplored.

**RATIONALE:** IRE1α–XBP1 signaling mediates the rapid induction of pro-inflammatory cytokines in myeloid cells. This pathway has also been implicated in the regulation of lipid metabolic processes that are central for programming immune cell functions in health and disease.




**Activation of the IRE1α–XBP1 pathway in leukocytes promotes $PGE_2$ generation and pain.** ER stress, inflammatory conditions, or engagement of pattern recognition receptors, such as Toll-like receptors (TLRs), trigger IRE1α activation and generation of functional XBP1 in myeloid leukocytes. This multitasking transcription factor induces the expression of both Cox-2 and mPGES-1, which are enzymes that catalyze the synthesis of $PGE_2$ from arachidonic acid. $PGE_2$ is known to function as a potent lipid mediator that promotes pain by activating and sensitizing nociceptors. Disabling IRE1α–XBP1 signaling reduces behavioral pain responses in $PGE_2$-dependent mouse models of postsurgical and inflammatory visceral pain.

Nonetheless, whether IRE1α–XBP1 activation in leukocytes modulates the pain that can be driven by inflammatory processes has not been studied. The scarcity of pharmaceutical products that effectively manage postoperative pain has promoted the use of opioids, in turn contributing to the opioid crisis in the United States. Identifying the key molecular pathways that endow immune cells with potent pro-algesic attributes may lead to the development of more effective and safer strategies for pain treatment. We examined whether leukocyte-intrinsic IRE1α–XBP1 signaling controls transcriptional and metabolic programs that could be implicated in inflammation and pain development.

**RESULTS:** Transcriptomic analyses of mouse bone marrow–derived dendritic cells stimulated by pattern recognition receptors revealed that IRE1α was necessary for the optimal expression of gene networks involved in eicosanoid metabolism. IRE1α deficiency blunted the normal induction of prostaglandin-endoperoxide synthase 2 (*Ptgs2*/Cox-2) and prostaglandin E synthase (*Ptges*/mPGES-1) in stimulated myeloid cells. This in turn reduced the capacity of myeloid cells to produce multiple prostaglandins, including the pro-algesic lipid mediator $PGE_2$. We determined that upon activation by IRE1α, the functional form of transcription factor XBP1 bound to the human *PTGS2* and *PTGES* genes to directly induce their expression and enable robust $PGE_2$ generation. Selective loss of IRE1α or XBP1 in leukocytes decreased $PGE_2$ biosynthesis in vivo upon challenge with pro-inflammatory stimuli and reduced pain-related behaviors in $PGE_2$-dependent models of visceral and postsurgical pain. Blocking IRE1α activation by using small-molecule inhibitors evoked similar antinociceptive effects in both models of pain evaluated.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau6499

**CONCLUSIONS:** Our study demonstrates that the IRE1α–XBP1 arm of the UPR operates as a crucial mediator of eicosanoid metabolism and prostaglandin synthesis in myeloid immune cells by promoting the expression of both Cox-2 and mPGES-1. We determined that abrogating this pathway genetically or pharmacologically diminishes pain-related behaviors in mice. Modulating IRE1α–XBP1 signaling may be helpful to induce better analgesia with the goal of improved pain management and reduced opioid use. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: laurie_glimcher@dfci.harvard.edu (L.H.G.); eromeros@wakehealth.edu (E.A.R.-S.); jur2016@med.cornell.edu (J.R.C.-R.)
Cite this article as S. Chopra *et al.*, *Science* **365**, eaau6499 (2019). DOI: 10.1126/science.aau6499

RESEARCH ARTICLE

INFLAMMATORY PAIN

# IRE1α–XBP1 signaling in leukocytes controls prostaglandin biosynthesis and pain

**Sahil Chopra[1,2,3]*, Paolo Giovanelli[1,2,3], Perla Abigail Alvarado-Vazquez[4], Sara Alonso[5], Minkyung Song[2,3], Tito A. Sandoval[2,3], Chang-Suk Chae[2,3], Chen Tan[2,3], Miriam M. Fonseca[4], Silvia Gutierrez[4], Leandro Jimenez[6,7], Kotha Subbaramaiah[8], Takao Iwawaki[9], Philip J. Kingsley[10], Lawrence J. Marnett[10,11], Andrew V. Kossenkov[12], Mariano Sanchez Crespo[5], Andrew J. Dannenberg[8], Laurie H. Glimcher[13,14]†, E. Alfonso Romero-Sandoval[4]†, Juan R. Cubillos-Ruiz[1,2,3]†**

**Inositol-requiring enzyme 1[α] (IRE1[α])–X-box binding protein spliced (XBP1) signaling maintains endoplasmic reticulum (ER) homeostasis while controlling immunometabolic processes. Yet, the physiological consequences of IRE1α–XBP1 activation in leukocytes remain unexplored. We found that induction of prostaglandin-endoperoxide synthase 2 (*Ptgs2*/Cox-2) and prostaglandin E synthase (*Ptges*/mPGES-1) was compromised in IRE1α-deficient myeloid cells undergoing ER stress or stimulated through pattern recognition receptors. Inducible biosynthesis of prostaglandins, including the pro-algesic mediator prostaglandin E2 ($PGE_2$), was decreased in myeloid cells that lack IRE1α or XBP1 but not other ER stress sensors. Functional XBP1 transactivated the human *PTGS2* and *PTGES* genes to enable optimal $PGE_2$ production. Mice that lack IRE1α–XBP1 in leukocytes, or that were treated with IRE1α inhibitors, demonstrated reduced pain behaviors in $PGE_2$-dependent models of pain. Thus, IRE1α–XBP1 is a mediator of prostaglandin biosynthesis and a potential target to control pain.**

The endoplasmic reticulum (ER) ensures the proper folding and posttranslational modification of secreted and transmembrane proteins. Diverse physiological and pathological conditions can provoke the accumulation of misfolded proteins within this organelle. These, in turn, can induce ER stress and activate the unfolded protein response (UPR). The inositol-requiring enzyme 1[α] (IRE1[α])–X-box binding protein spliced (XBP1) pathway is the most evolutionarily conserved arm of the UPR (*1*). When ER homeostasis is altered, the dual-enzyme IRE1α undergoes oligomerization and autophosphorylation, activating its endoribonuclease domain to excise a 26-nucleotide fragment from unspliced *Xbp1* mRNA (*2*). This unconventional splicing event gives rise to the functional form of transcription factor XBP1, which promotes the expression of multiple genes that enhance the protein-folding capacity of the ER (*2*, *3*).

Emerging evidence indicates that IRE1α–XBP1 signaling can also control UPR-independent cellular pathways, influencing processes such as hepatic lipogenesis (*4*), hypoxia responses (*5*), angiogenesis (*6*), atherosclerosis (*7*, *8*), arthritis (*9*), and antitumor immunity (*10*–*12*). Myeloid cells stimulated through plasma membrane–bound Toll-like receptors (TLRs) rapidly and selectively activate IRE1α–XBP1. This event is required for their optimal production of some pro-inflammatory cytokines (*13*). Nevertheless, the precise transcriptional and metabolic programs coordinated by IRE1α–XBP1 signaling in leukocytes under inflammatory conditions, and their physiological consequences, remain largely unexplored.

## IRE1α controls transcriptional programs in myeloid cells stimulated through pattern recognition receptors

To understand how IRE1α–XBP1 activation influences global gene expression in myeloid cells, we performed unbiased transcriptomic analyses of wild-type (WT) and IRE1α-deficient bone marrow–derived dendritic cells (BMDCs) stimulated with bacterial lipopolysaccharide (LPS) (TLR4 agonist) or fungal zymosan (Dectin-1 and TLR2 agonist). Consistent with previous reports (*13*), WT BMDCs exposed to these microbial products exhibited IRE1α-dependent *Xbp1* splicing (fig. S1, A and B) but did not show robust induction of canonical XBP1 target genes in the ER stress response (fig. S1C) or activation of other UPR branches (fig. S1D). We did not observe signs of regulated IRE1α-dependent decay (RIDD) (*14*–*16*) upon LPS or zymosan stimulation because the expression levels of several genes reported to be potentially regulated by this process were not increased in BMDCs that lack IRE1α (fig. S1, E and F). IRE1α deficiency did not compromise normal BMDC generation or survival in response to granulocyte macrophage colony-stimulating factor (GM-CSF) (fig. S2, A and B). However, we identified 1792 and 2863 genes whose expression was significantly altered in IRE1α-deficient BMDCs stimulated with either zymosan or LPS, respectively, compared with their WT counterparts (Fig. 1A). There was a significant overlap of 1167 differentially regulated genes between the two stimuli (Fig. 1A), indicating a common effect of IRE1α deficiency independent of the agonist used. Ingenuity pathway analysis (IPA) for these commonly regulated genes revealed the enrichment of nine biological categories (Fig. 1B). As expected, IRE1α deficiency influenced transcriptional processes involved in post-translational protein modification as well as cellular maintenance and survival (Fig. 1B) (*17*). Surprisingly, biosynthesis and metabolism of eicosanoids emerged as a major cellular function potentially regulated by IRE1α in BMDCs stimulated with LPS or zymosan (Fig. 1B). We then identified 27 regulators that not only changed expression at the mRNA level but also had a significant number of known targets enriched in the gene list (Fig. 1C, top 10 regulators are shown). Confirming previous reports (*13*), expression of *Il6* and its associated target genes was significantly decreased in TLR-stimulated BMDCs that lack IRE1α, compared with their WT counterparts (Fig. 1C). Additionally, and corresponding with IPA denoting altered eicosanoid metabolism, prostaglandin-endoperoxide synthase 2 (*Ptgs2*/Cox-2) and prostaglandin E synthase (*Ptges*/mPGES-1) emerged as potential regulators that were markedly decreased in IRE1α-deficient BMDCs exposed to LPS or zymosan (Fig. 1C). We confirmed the down-regulation of these two enzymes at the mRNA and protein levels in stimulated BMDCs that lack IRE1α using quantitative reverse transcription polymerase chain reaction (RT-PCR) and immunoblot assays, respectively (Fig. 1, D and E). IRE1α deficiency did not affect the constitutive expression of *Ptgs1*/Cox-1 or *Ptges2*/mPGES-2 (fig. S2C), suggesting that this

[1]Weill Cornell Graduate School of Medical Sciences, Cornell University. New York, NY 10065, USA. [2]Department of Obstetrics and Gynecology, Weill Cornell Medicine. New York, NY 10065, USA. [3]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, USA. [4]Department of Anesthesiology, Pain Mechanisms Laboratory, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA. [5]Instituto de Biología y Genética Molecular, CSIC-Universidad de Valladolid, Valladolid, Spain. [6]Instituto Ludwig de Pesquisa Sobre o Câncer, São Paulo, Brazil. [7]Hospital Sírio-Libanês, São Paulo, Brazil. [8]Department of Medicine, Weill Cornell Medicine, New York, NY 10065, USA. [9]Division of Cell Medicine, Medical Research Institute, Kazanawa Medical University, Ishikawa, Japan. [10]Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA. [11]A. B. Hancock Jr. Memorial Laboratory for Cancer Research, Departments of Biochemistry, Chemistry and Pharmacology, Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, TN 37232, USA. [12]Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA 19104, USA. [13]Department of Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02115, USA. [14]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA.
*Present Address: Vertex Ventures HC, 345 California Avenue, Palo Alto, CA 94306, USA.
†Corresponding author. Email: laurie_glimcher@dfci.harvard.edu (L.H.G.); eromeros@wakehealth.edu (E.A.R.-S.); jur2016@med.cornell.edu (J.R.C.-R.)

ER stress sensor primarily mediates the rapid induction of *Ptgs2*/Cox-2 and *Ptges*/mPGES-1 in response to inflammatory stimuli. Consequently, we hypothesized that IRE1α may be required for normal eicosanoid production by myeloid cells.

### IRE1α–XBP1 signaling is necessary for optimal prostaglandin biosynthesis

Prostaglandins are a major class of eicosanoids whose inducible biosynthesis depends on the rapid metabolism of arachidonic acid by Cox-2 (Fig. 2A) (*18*). These bioactive lipids participate in the regulation of diverse physiological processes such as allergy, fever, vascular permeability, and pain (*19*). Lipidomic analyses revealed that IRE1α deficiency did not influence basal prostaglandin levels in untreated BMDCs (Fig. 2B). However, and consistent with the observed impairment in Cox-2 induction, we identified a profound decrease in the intracellular levels of several prostaglandins, including prostaglandin $E_1$ ($PGE_1$), $PGF_{1\alpha}$, $PGD_2$, $PGE_2$, $PGF_{2\alpha}$, 15-keto $PGF_{2\alpha}$, D12-$PGJ_2$, 13,14dh-15k $PGE_2$, and $PGD_3$ in LPS-stimulated BMDCs that lack IRE1α when compared with their WT counterparts (Fig. 2B and table S1).

Cox-2 converts arachidonic acid to prostaglandin endoperoxide $H_2$ ($PGH_2$), which is subsequently metabolized by mPGES-1 to generate the potent lipid mediator $PGE_2$ (Fig. 3A) (*18*).

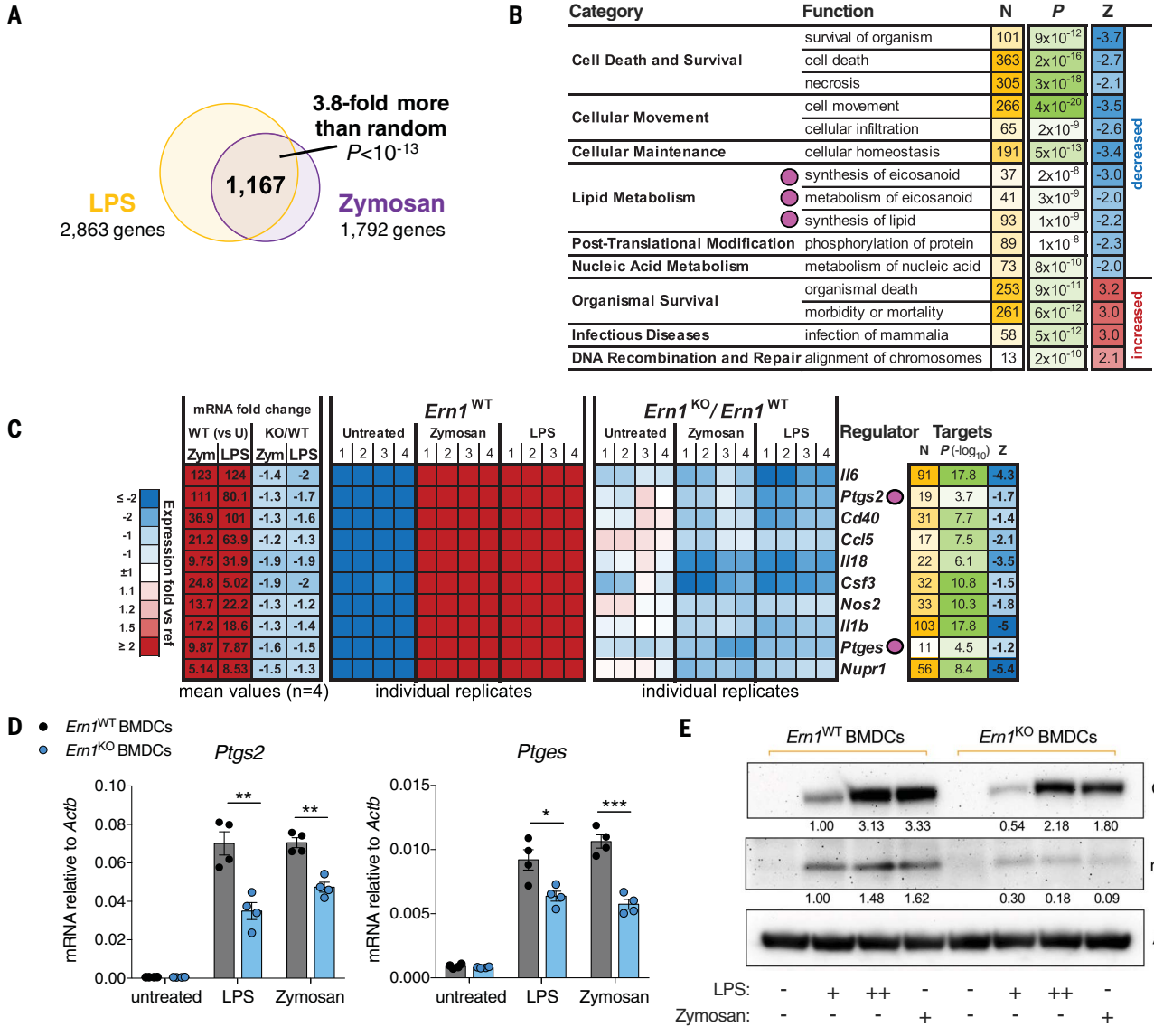



**Fig. 1. IRE1α regulates the expression of *Ptgs2* and *Ptges*.** *Ern1*$^{WT}$ or *Ern1*$^{KO}$ BMDCs were left untreated or stimulated with LPS (50 ng/ml) or zymosan (25 μg/ml) for 6 hours, and global transcriptional profiles were determined with RNA-seq. (**A**) Number of differentially regulated genes identified in IRE1α-deficient BMDCs treated with LPS or zymosan. (**B**) IPA of RNA-seq data highlighting eicosanoid metabolism as one of the central biological functions altered in LPS- or zymosan-stimulated BMDCs lacking IRE1α. (**C**) Top 10 key regulators identified with RNA-seq analysis. Individual replicate heatmaps show log ratio of expression values versus the reference, which is the mean across samples for the "*Ern1* $^{WT}$" panel, and *Ern1*$^{WT}$ for the "*Ern1*$^{KO}$/*Ern1*$^{WT}$" panel. U, untreated; Zym, zymosan. (**D**) Quantitative RT-PCR for *Ptgs2* and *Ptges* upon LPS or zymosan stimulation. (**E**) Representative immunoblot analyses for Cox-2 and mPGES-1 expression in *Ern1*$^{WT}$ and *Ern1*$^{KO}$ BMDCs stimulated with LPS (+, 10 ng/ml, or ++, 100 ng/ml) or zymosan (25 μg/ml). The density of each band was normalized to its own actin value, and numbers shown represent relative expression compared with *Ern1*$^{WT}$ BMDCs treated with 10 ng/ml LPS because there was undetectable expression of these proteins in unstimulated BMDCs. In (D) and (E), data are shown as mean ± SEM and are representative of at least three independent experiments. Two-tailed Student's *t* test was used for statistical analysis. *$P$ < 0.05, **$P$ < 0.005, ***$P$ < 0.0005.

Corresponding with the decreased induction of both Cox-2 and mPGES-1 in IRE1α-deficient BMDCs stimulated with LPS (Fig. 1), there was a marked reduction in $PGE_2$ production by these cells compared with their WT counterparts (Fig. 3, B and C). Additional IRE1α-deficient myeloid cell subsets, including primary neutrophils and macrophages, also demonstrated defective $PGE_2$ synthesis upon LPS stimulation (fig. S3, A and B). To further confirm these findings in vivo, we administered LPS intraperitoneally to transgenic mice selectively lacking IRE1α in leukocytes ($Ern1^{f/f}$ $Vav1^{cre}$) (*20*) and then quantified $PGE_2$ production in situ. LPS exposure triggered *Xbp1* splicing and the concomitant IRE1α-dependent induction of both *Ptgs2* and *Ptges* in peritoneal leukocytes (fig. S4, A to C). Accordingly, mice lacking IRE1α in leukocytes demonstrated reduced production of peritoneal $PGE_2$ upon LPS administration compared with that of their WT counterparts (fig. S4D). Confirming our transcriptional profiling by using an independent agonist (Fig. 1), $PGE_2$ synthesis was also diminished in zymosan-exposed BMDCs that lack IRE1α (Fig. 3D). Similar results were observed in vivo after the intraperitoneal administration of zymosan into mice lacking IRE1α in leukocytes (fig. S4, E to I). In this setting, lipidomic analyses further confirmed that the production of Cox-2–dependent prostaglandins ($PGE_2$, $PGD_2$, and $PGF_2\alpha$) and $TBX_2$ was reduced in cell-free peritoneal lavage from $Ern1^{f/f}$ $Vav1^{cre}$ compared with $Ern1^{f/f}$ mice (fig. S4, E to H). By contrast, lipoxygenase-dependent 15-HETE was unaltered (fig. S4I). XBP1 deletion phenocopied the same defects observed in IRE1α-deficient myeloid cells (Fig. 3E and fig. S3C), whereas the ablation of other ER stress sensors such as PERK (protein kinase RNA-like endoplasmic reticulum kinase) (encoded by *Eif2ak3*) and ATF6α (activating transcription factor 6α) did not compromise $PGE_2$ generation in response to this treatment (Fig. 3, F and G). Thus, the IRE1α–XBP1 arm of the ER stress response is selectively required for optimal $PGE_2$ production by LPS- or zymosan-stimulated myeloid cells.

IRE1α-dependent induction of $PGE_2$ was also observed in BMDCs treated with agonists for other plasma membrane–bound TLRs, whereas stimulation from endosomal TLR3, TLR8, or TLR9 had no effect (fig. S5A). These results are consistent with previous reports that demonstrated predominant IRE1α–XBP1 activation by agonists engaging plasma membrane-bound but not endosomal TLRs (*13*). $PGE_2$ induction was also reduced in IRE1α-deficient BMDCs activated with phorbol myristate acetate (PMA) (fig. S5A), thus ruling out the possibility that IRE1α ablation was compromising proximal TLR signaling. Moreover, we also found diminished $PGE_2$ production, accompanied by reduced expression of both Cox-2 and mPGES-1, in IRE1α-deficient BMDCs treated with the pharmacological ER stressor thapsigargin (fig. S5, B and C). Thus, optimal $PGE_2$ synthesis by mouse myeloid cells undergoing ER stress or stimulated through plasma membrane–bound TLRs requires IRE1α–XBP1 activation, which promotes the expression of Cox-2 and mPGES-1.

To define whether IRE1α–XBP1 signaling also controlled inducible $PGE_2$ production in human myeloid cells, we generated monocyte-derived DCs from the peripheral blood of healthy volunteers and then abrogated this pathway using gene-editing techniques. Transient transfection of primary human DCs with single-guide RNA (sgRNA)–Cas9 complexes targeting *XBP1* effectively edited this gene and prevented the generation of its spliced (active) form upon zymosan treatment (Fig. 3H). The induction of *PTGS2* and *PTGES*, as well as $PGE_2$ production, were significantly diminished in zymosan-exposed human DCs that lack XBP1, compared with their WT counterparts transfected with scrambled sgRNA-Cas9 complexes (Fig. 3, I to K). Similar effects were observed when human DCs were transfected with complexes that target *ERN1* (fig. S5, D and E). Thus, IRE1α–XBP1 signaling acts as a conserved mediator of inducible $PGE_2$ production in human DCs.

### XBP1s transactivates the human *PTGS2* and *PTGES* promoters

We sought to determine the molecular mechanism by which IRE1α-activated XBP1 (XBP1s) mediates inducible $PGE_2$ production in human myeloid cells. We analyzed the promoter regions of *PTGS2* and *PTGES* for potential XBP1s binding sites, as previously described (*21*, *22*), and found putative X-box–binding and Unfolded Protein Responses Element A (UPRE-A) sequences on the *PTGS2* promoter (Fig. 4A). Additionally, we identified an X-box–binding region and two ETS domain–binding sites in the *PTGES* promoter (Fig. 4B). We hypothesized that XBP1s could operate as a driver of *PTGS2* and *PTGES* transcription.

Chromatin immunoprecipitation (ChIP)–PCR was used to evaluate direct XBP1s binding to the



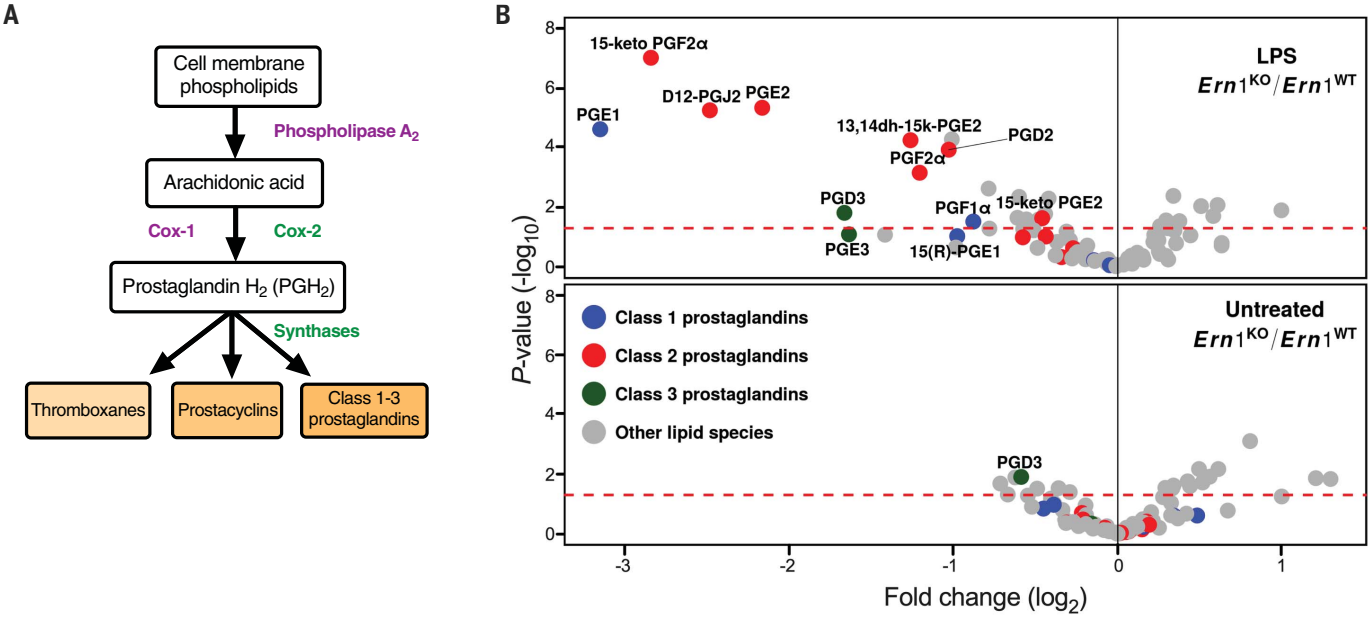



**Fig. 2. IRE1α promotes prostaglandin biosynthesis.** (**A**) Pathway depicting the main events implicated in the metabolism of arachidonic acid and prostaglandin production. Cox-1 is constitutively expressed, whereas Cox-2 is induced by pro-inflammatory stimuli. (**B**) $Ern1^{WT}$ ($n$ = 4 independent samples) or $Ern1^{KO}$ BMDCs ($n$ = 3 independent samples) were left untreated (bottom) or stimulated with LPS (50 ng/ml) for 6 hours (top), and lipidomic analyses were performed. Data are represented as volcano plots with red lines indicating a significance level of $P$ = 0.05. Two-tailed Student's $t$ test with FDR correction was used for statistical analysis.

promoter regions identified. We stimulated human monocyte-derived DCs with zymosan alone or in combination with 2-deoxy-D-glucose (2-DG), which inhibits *N*-linked protein glycosylation, producing ER stress and robust IRE1α–XBP1 activation (*23*). Zymosan exposure increased XBP1s binding to the predicted *PTGS2* and *PTGES* promoter regions, and concomitant treatment with the ER stressor 2-DG markedly enhanced these effects (Fig. 4, C and D). Disabling the IRE1α ribonuclease (RNase) domain by using the selective pharmacological inhibitor MKC8866 (*24*–*27*) abrogated XBP1s binding to these promoters in zymosan-stimulated human DC undergoing ER stress (Fig. 4, C and D). XBP1s also bound the *GFPT1* promoter, as previously reported (*23*), whereas promoter regions of *pri-miR-21* devoid of XBP1s-binding sites were not enriched in these assays (Fig. 4, E and F). Furthermore, luciferase reporter assays using human embryonic kidney (HEK) 293 cells demonstrated that XBP1s was

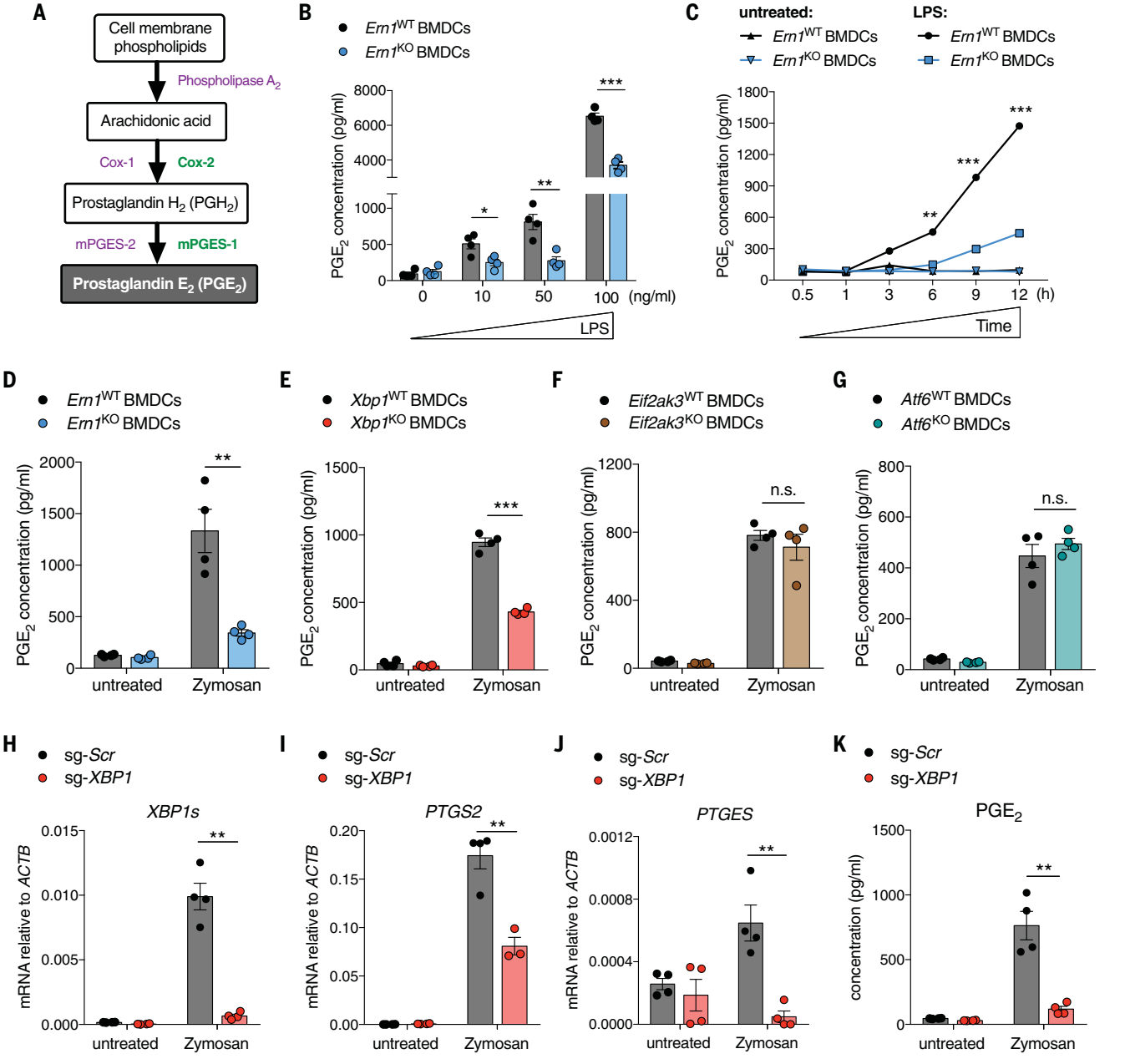



**Fig. 3. IRE1α–XBP1 is necessary for optimal $PGE_2$ production by myeloid cells.** (**A**) Pathway showing the main steps implicated in $PGE_2$ generation. Both Cox-1 and mPGES-2 are constitutively expressed, whereas Cox-2 and mPGES-1 are induced by pro-inflammatory stimuli. (**B** and **C**) ELISA-based confirmation for reduced $PGE_2$ in supernatants from *Ern1*$^{KO}$ BMDCs upon stimulation with the indicated concentrations of LPS (B), or at different time points after stimulation with LPS at 50 ng/ml (C). (**D** to **G**) Mouse BMDCs of the indicated genotypes were stimulated with zymosan (25 μg/ml) for 6 hours, and $PGE_2$ was quantified in culture supernatants by means of ELISA. Dots represent BMDCs generated from independent mice. (**H** to **K**) CRISPR/Cas9–based gene editing was used to ablate *XBP1* in human monocyte–derived DCs, and cells were then stimulated for 6 hours with zymosan (25 μg/ml). [(H) to (J)] Quantitative RT-PCR was used to assess the indicated transcript levels, and (K) $PGE_2$ levels were determined in the corresponding supernatants by means of ELISA. sg, single-guide; Scr, scrambled RNA. Dots represent human DC samples generated from independent donors. Data are shown as mean ± SEM and are representative of at least two independent experiments. Two-tailed Student's *t* test was used for (B) and (D) to (K). Two-way ANOVA (Tukey's test) was used for (C); $*P < 0.05$, $**P < 0.005$, $***P < 0.0005$.

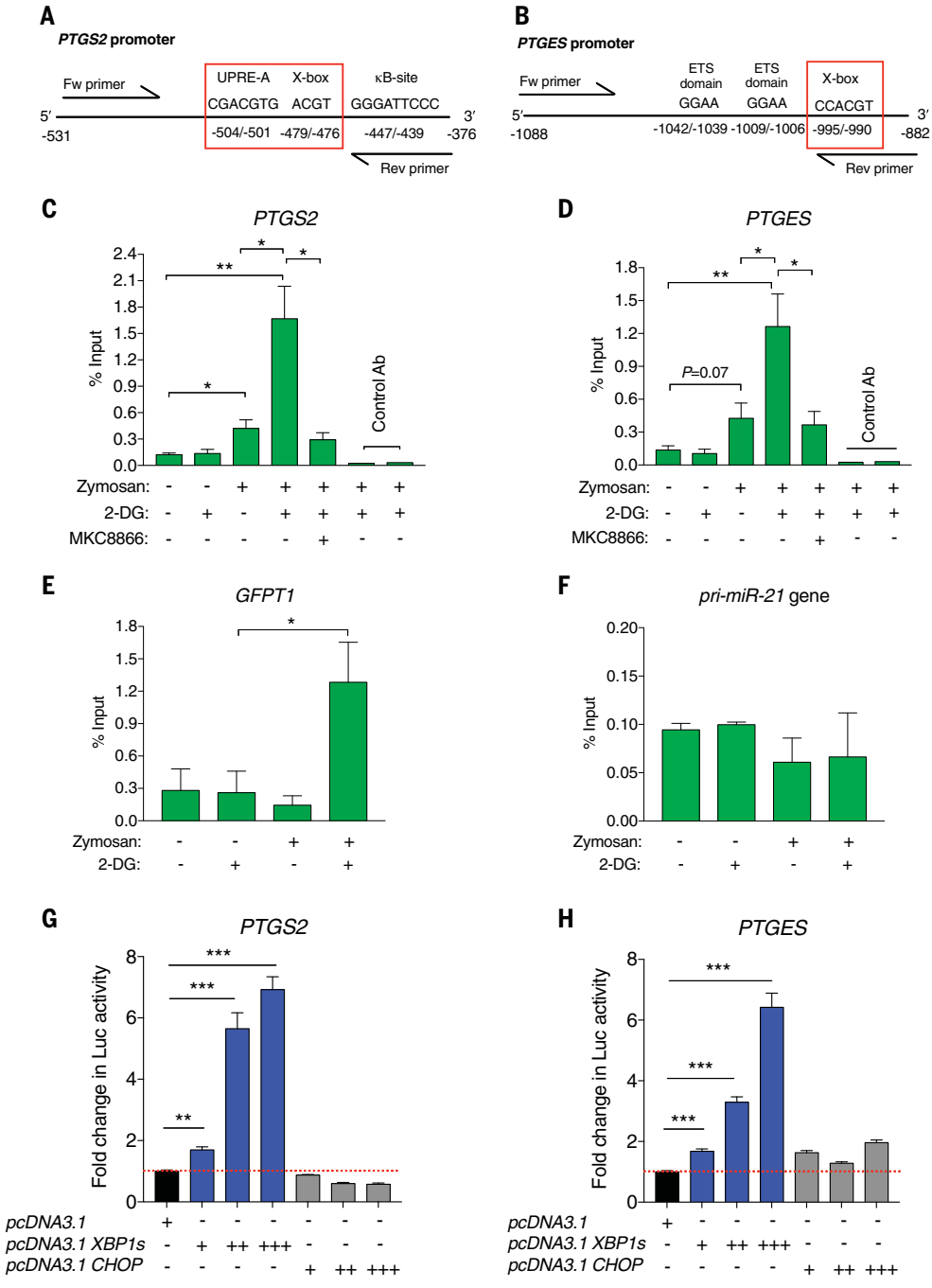



**Fig. 4. XBP1s transactivates the *PTGS2* and *PTGES* promoters.** (**A** and **B**) Predicted XBP1s-binding sites in promoter regions of human (A) *PTGS2* and (B) *PTGES*. (**C** to **F**) Human primary monocyte–derived DCs were left untreated or stimulated with zymosan in the presence or absence of the ER stressor 2-DG. ChIP assays were performed by using anti-XBP1s or isotype control antibodies (Ab). The IRE1α inhibitor MKC8866 was used as indicated, and quantitative PCR was used to determine XBP1s occupancy at these promoter regions under the conditions tested. ChIP-PCR assays were performed by using three to six independent human donors. (**G** and **H**) HEK293 cells were cotransfected with XBP1s- or CHOP-expressing plasmid vectors, and luciferase reporter constructs harboring the (G) *PTGS2* or (H) *PTGES* promoters, along with a *Renilla* luciferase construct for internal control. Firefly luciferase (Luc) activity was normalized to *Renilla* activity in each case. Data are representative of at least two independent experiments with similar results, using four technical replicates. Data are shown as mean ± SEM and expressed as fold increase in Firefly luciferase activity compared with empty vector (pcDNA3.1). Two-tailed Student's *t* test was used for statistical analysis. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$.

sufficient to transactivate the human *PTGS2* and *PTGES* promoters in a dose-dependent manner (Fig. 4, G and H). By contrast, the PERK-controlled ER stress transcription factor C/EBP homologous protein (CHOP) showed no effect in this reporter system (Fig. 4, G and H). Thus, IRE1α-activated XBP1s promotes inducible $PGE_2$ biosynthesis by directly driving the transcriptional induction of both *PTGS2* and *PTGES*.

## IRE1α–XBP1 signaling in leukocytes promotes pain behaviors

$PGE_2$ generated by induction of Cox-2 and mPGES-1 engages EP1-4 receptors on peripheral sensory neurons and the central nervous system to promote pain responses (*28*–*30*). We postulated that mice lacking IRE1α in leukocytes would demonstrate reduced pain behaviors because of their impaired capacity to induce $PGE_2$ production in response to inflammatory stimuli (fig. S4). Two classical $PGE_2$-dependent models of pain were used to test this hypothesis: an acetic acid–based model for inflammatory visceral pain (*31*–*34*) and a paw incision model of postsurgical pain (*35*). We injected 0.9% v/v of acetic acid intraperitoneally into either $Ern1^{f/f}$ or $Ern1^{f/f}$ $Vav1^{cre}$ mice, and writhing behaviors were monitored over time by a blinded observer. The number of writhing events recorded within the first 30 min (Fig. 5A), as well as $PGE_2$ levels in cell-free peritoneal lavage samples (fig. S6A), were significantly reduced in $Ern1^{f/f}$ $Vav1^{cre}$ mice compared with their IRE1α-sufficient counterparts. We also performed automated unbiased and blinded tests to evaluate how the inflammatory visceral pain caused by acetic acid administration affected the normal ambulatory capacity of the host. IRE1α-sufficient ($Ern1^{f/f}$) mice demonstrated lower displacement ability than that of their $Ern1^{f/f}$ $Vav1^{cre}$ littermates, as evidenced by a significant decrease in their total ambulatory counts and times after acetic acid injection

(Fig. 5, B and C). Reduced writhing behaviors were also observed in mice selectively lacking XBP1 in leukocytes (*Xbp1*$^{f/f}$ *Vav1*$^{cre}$) (Fig. 5D), thus confirming that canonical IRE1α–XBP1 signaling mediates this response. Peritoneal leukocytes demonstrated constitutive IRE1α-dependent *Xbp1* splicing that was maintained upon acetic acid administration (fig. S6B), and loss of IRE1α in these cells decreased their normal expression of *Ptges* by ~50% (fig. S6C). No alterations in interleukin-1β (IL-1β), IL-6, or tumor necrosis factor–α (TNFα) expression at the mRNA or protein levels were found in this milieu at the same time point analyzed (fig. S6, D to G). Thus, leukocyte-intrinsic IRE1α enabled the rapid production of $PGE_2$ without modulating these pro-inflammatory cytokines upon intraperitoneal acetic acid administration. Similar IRE1α-driven writhing responses were observed in male and female mice, indicating that this phenotype was not markedly influenced by sex (fig. S7). To determine whether disabling IRE1α–XBP1 signaling pharmacologically could reduce inflammatory visceral pain, we used two inhibitors

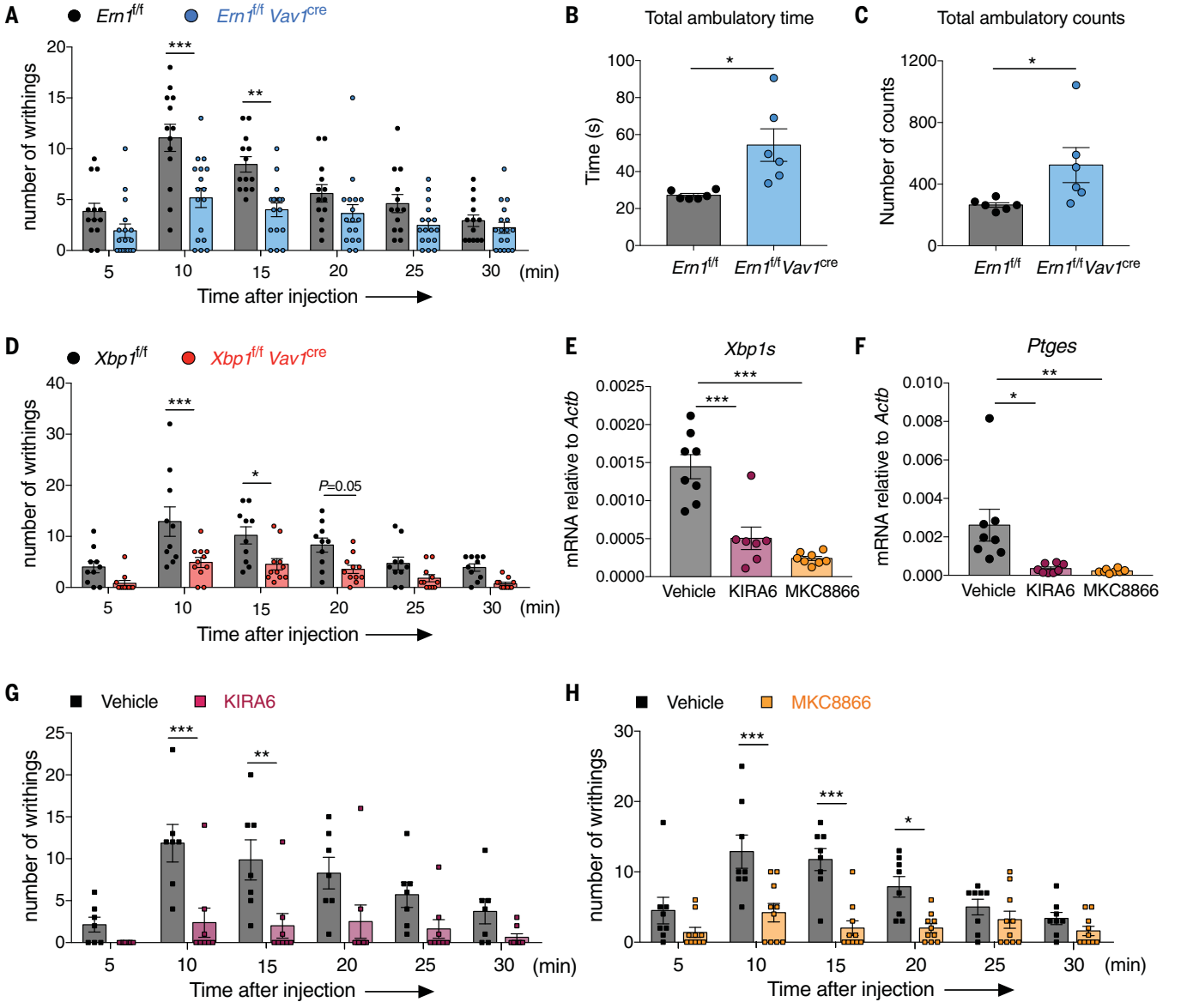



**Fig. 5. IRE1α–XBP1 signaling in leukocytes promotes inflammatory visceral pain.** (**A** to **C**) 0.9% v/v acetic acid (5 ml/kg) was administered intraperitoneally to *Ern1*$^{f/f}$ ($n$ = 13) or *Ern1*$^{f/f}$ *Vav1*$^{cre}$ ($n$ = 17) mice. Each group contained both male and female mice. (A) Writhing behaviors after acetic acid injection were recorded every 5 min for 30 min. Total ambulatory time (B) and total ambulatory counts (C) for mice of the indicated genotypes were monitored for 15 min after acetic acid injection. (**D**) 0.9% v/v acetic acid (5 ml/kg) was injected intraperitoneally into *Xbp1*$^{f/f}$ ($n$ = 10) or *Xbp1*$^{f/f}$ *Vav1*$^{cre}$ ($n$ = 11) mice (mix of males and females) and writhing was recorded every 5 min for 30 min. (**E** to **H**) Pharmacological inhibition of IRE1α reduces visceral pain. WT C57BL/6J mice were given KIRA6 (25 mg/kg) or MKC8866 (20 mg/kg) intraperitoneally 6 hours and 30 min before challenge with 0.9% v/v acetic acid (5 ml/kg). In (E) and (F), the indicated mRNA transcript levels were determined by means of quantitative RT-PCR in leukocytes recovered from peritoneal lavage samples after treatment with IRE1α inhibitors. Writhing behaviors after acetic acid administration were recorded in mice receiving (G) KIRA6 or (H) MKC8866 and compared with their corresponding vehicle controls. All data are presented as mean ± SEM. Two-way ANOVA (Sidak's test) was used for (A), (D), (G), and (H). Two-tailed Student's $t$ test was used for (B) and (C). One-way ANOVA (Dunnett's test) was used for (E) and (F). *$P$ <0.05, **$P$ < 0.005, ***$P$ < 0.0005. Data corresponding to behavioral responses were generated in two independent experiments.

of IRE1α: the kinase domain–specific inhibitor KIRA6 (*36*, *37*) and the RNase domain–specific inhibitor MKC8866 (*24*–*27*). These compounds were independently administered intraperitoneally 6 hours and 30 min before acetic acid injection. Treatment with either KIRA6 or MKC8866 reduced *Xbp1s* and *Ptges* expression in peritoneal leukocytes (Fig. 5, E and F) and significantly diminished the number of writhing behaviors after acetic acid injection (Fig. 5, G and H). The administration of a similar dose of celecoxib, a nonsteroidal anti-inflammatory drug that inhibits Cox-2 and limits prostanoid production, also decreased writhing behaviors (fig. S8), confirming the role of $PGE_2$ in this behavioral response. Thus, IRE1α–XBP1 activation promotes visceral pain in the acetic acid–based model.

We next evaluated whether IRE1α deficiency in leukocytes could also influence postoperative pain, which is a $PGE_2$-mediated process commonly treated with Cox-2 inhibitors (*38*–*40*). A surgical incision was made in the right hind paw of either *Ern1*$^{f/f}$ or *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice, and nonreflexive pain-related behaviors were monitored over time and analyzed in comparison with baseline measurements before surgery. We observed IRE1α-dependent *Xbp1* splicing in CD45$^+$ leukocytes sorted from the injury site 24 hours after surgery (fig. S9A). Although the proportions of neutrophils, macrophages, and DCs infiltrating the lesions at this time point were not altered (fig. S9B), we identified a significant reduction in the number of Cox-2–expressing leukocytes

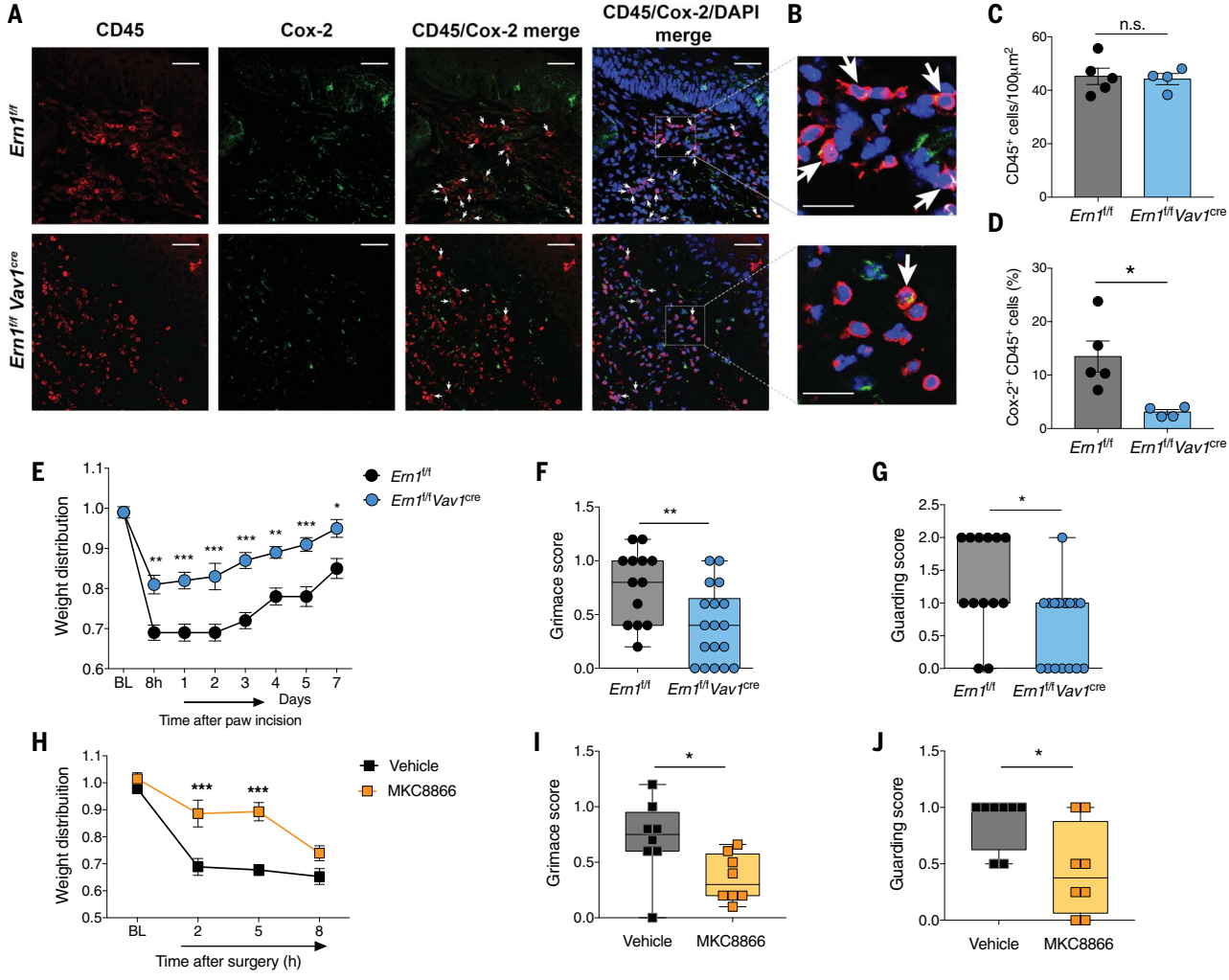



**Fig. 6. IRE1α activation mediates spontaneous pain behaviors postsurgery.** (**A** to **D**) Cox-2 expression in leukocytes infiltrating the paw after surgery. (A) Representative confocal microscopy images of ipsilateral paw tissue from *Ern1*$^{f/f}$ or *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice stained with fluorescently labeled antibodies specific for CD45 (red) or Cox-2 (green), and DAPI (blue). White arrows denote cells coexpressing CD45 and Cox-2. Scale bars, 100 μm. (B) Magnified images of the indicated insets are shown. Scale bars, 50 μm. (C) Quantification of total CD45$^+$ cells, and of Cox-2-expressing CD45$^+$ leukocytes (D) in the paw 48 hours after surgery. Data are shown as mean ± SEM. *$P$ < 0.05. (**E** to **G**) A surgical incision was made in the right hind paw of *Ern1*$^{f/f}$ ($n$ = 11 to 24) or *Ern1*$^{f/f}$ *Vav1*$^{cre}$ ($n$ = 11 to 29) mice. Each group comprised both male and female mice. Animals of the indicated genotypes were monitored for (E) spontaneous hind paw weight bearing distribution, (F) grimace score, and (G) guarding score. Grimace and guarding scores were determined 8 hours postsurgery. (**H** to **J**) Pharmacological inhibition of IRE1α reduces postoperative pain behaviors. C57BL/6J male mice ($n$ = 8/group) were administered vehicle or MKC8866 (20 mg/kg) intraperitoneally 6 hours and 30 min before a surgical incision was made in the right hind paw. Animals were monitored then for (H) spontaneous weight-bearing distribution, (I) grimace score, and (G) guarding score. Grimace and guarding scores were determined 8 hours after surgery. For (F), (G), (I) and (J), data are presented as median ± 25 to 75% confidence interval (boxes) with smallest and largest values (whiskers). Data in all other panels are shown as mean ± SEM. Two-tailed Student's $t$ test was used for (C) and (D). Two-way ANOVA (Sidak's test) was used for (E) and (H). The Mann-Whitney $U$ test was used for (F), (G), (I), and (J). *$P$ < 0.05, **$P$ < 0.005, ***$P$ < 0.0005.

infiltrating the injured tissues (surgical site) in *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice, compared with their littermate controls (Fig. 6, A to D). Weight-bearing distribution tests indicated that *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice showed a greater capacity to use their injured paws as compared with their IRE1α-sufficient counterparts. They displayed a more balanced hind-paw distribution at early stages after incision and recovered significantly faster (Fig. 6E). These effects were not associated with differential body weight in the two genotypes (fig. S10A). Additional parameters indicative of spontaneous nonevoked pain behaviors, such as facial expression by grimace score (Fig. 6F) and guarding behavior by using the guarding score (Fig. 6G), were also decreased in mice lacking IRE1α in leukocytes. Mechanical threshold, determined by using von Frey–induced withdrawal reflex analysis, was comparable in *Ern1*$^{f/f}$ versus *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice after surgery (fig. S10B), which is consistent with the negligible role of peripheral Cox-2 in evoked punctate mechanical hypersensitivity previously reported in rodent models of acute pain (*41*–*44*). The numbers of flinches (fig. S10C) and paw perimeter (fig. S10D) also remained unchanged in *Ern1*$^{f/f}$ versus *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice after paw incision. The role of leukocyte-intrinsic IRE1α in these postsurgical responses was comparable in male (fig. S11, A to G) and female (fig. S12, A to G) mice. However, we found that male *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice showed reduced impairment and faster recovery of rearing activity in comparison with their *Ern1*$^{f/f}$ counterparts (fig. S11H), whereas female mice did not (fig. S12H).

To determine whether pharmacological targeting of IRE1α could also modulate postsurgical pain, we administered MKC8866 before paw incision surgery and monitored pain responses thereafter. IRE1α inhibition improved nociceptive functional behaviors, as demonstrated by a more balanced weight distribution when compared with that of vehicle-treated mice (Fig. 6H). Grimace and guarding scores after surgery were also significantly reduced in mice receiving MKC8866 (Fig. 6, I and J). In contrast to our observations when using *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice, we found reduced flinching activity after paw incision in MKC8866-administered groups (fig. S13A), suggesting a pro-algesic role for IRE1α in additional nonleukocyte cells in this setting. Rearing activity was unchanged upon IRE1α targeting (fig. S13B), indicating that complete inhibition of IRE1α may be required for altering this specific behavior in male mice after paw incision. Consistent with our results using conditional IRE1α-deficient mice, mechanical hypersensitivity remained unaltered upon administration of MKC8866 (fig. S13C). Similar behavioral responses were observed in mice treated with KIRA6 (fig. S14). As a positive control, we administered a comparable dose of celecoxib following the same scheme and route described above. Similar to IRE1α inhibition, we observed a more balanced weight-bearing distribution as well as diminished grimace and guarding scores after paw incision in mice that received celecoxib, compared with vehicle-treated mice (fig. S15, A to C). Flinches, rearing activity, and mechanical threshold after paw incision remained unaffected upon celecoxib treatment (fig. S15, D to F). This is consistent with previous reports demonstrating that Cox-2 inhibition at the comparable dose chosen does not influence mechanical hypersensitivity in models of acute pain (*41*, *42*). Thus, mice lacking IRE1α–XBP1 in leukocytes exhibit reduced behavioral pain responses in two distinct $PGE_2$-dependent models of pain. Furthermore, targeting IRE1α pharmacologically can modulate these pain behaviors in vivo.

## Conclusions

Here, we present molecular and functional evidence that reveals an unexpected function for the ER stress sensor IRE1α as a central mediator of prostaglandin biosynthesis and behavioral pain responses in mice. Our findings suggest a previously unappreciated mechanism in which IRE1α activates transcription factor XBP1 to sustain expression of two rate-limiting enzymes that are necessary for optimal prostaglandin production—namely, Cox-2 and mPGES-1. More effective pain management strategies are needed in the clinic, especially in light of the devastating opioid crisis that the United States currently faces (*45*). The pharmacological modulation of IRE1α–XBP1 signaling may represent an alternative approach for pain control with the potential of producing better analgesia, diminished opioid requirements, and reduced opioid side effects. Future studies will be needed to determine whether IRE1α–XBP1 signaling also regulates additional physiological and pathological processes driven by prostaglandins such as pregnancy, fever, allergy, and immunosuppression in cancer.

## Materials and methods

### *RNA isolation, quantitative RT-PCR, and Xbp1* splicing assays

Total RNA was isolated using RNeasy Mini kit or QIAzol lysis reagent (Qiagen) according to the manufacturer's instructions. RNA (0.1 to 1 μg) was reverse-transcribed to generate cDNA using the qScript cDNA synthesis kit (Quantabio). Quantitative RT-PCR was performed using PerfeCTa SYBR green fastmix (Quantabio) and TaqMan Universal PCR master mix (Life Technologies) on a QuantStudio 6 Flex real-time PCR system (Applied Biosystems). Normalized gene expression was calculated by comparative threshold cycle method using *ACTB* or *Actb* as a control. *Xbp1* splicing assays were performed as previously described (*46*). PCR products were separated by electrophoresis through a 2.5% agarose gel and visualized by ethidium bromide staining. Primers used in this study are described in table S2.

### *Transgenic mice*

*Atf6*$^{f/f}$, *Eif2ak3*$^{f/f}$, *Vav1*$^{cre}$, and *CD11c*$^{cre}$ mice were obtained from The Jackson Laboratory. *Xbp1*$^{f/f}$ and *Ern1*$^{f/f}$ mice have been previously described by our groups (*4*, *47*). We generated conditional knockout mice lacking ATF6, IRE1α, or XBP1 in leukocytes by crossing *Atf6*$^{f/f}$, *Ern1*$^{f/f}$, or *Xbp1*$^{f/f}$ animals, respectively, with the *Vav1*$^{cre}$ strain that allows selective gene deletion in hematopoietic cells (*20*). Crossing *Eif2ak3*$^{f/f}$ mice with *CD11c*$^{cre}$ animals generated mice devoid of PERK in DCs. All mouse strains were on a full C57BL/6J background. Mice were housed in specific pathogen-free animal facilities at Weill Cornell Medical College, Memorial Sloan Kettering Cancer Center, and Wake Forest University. Mice were handled in compliance with Weill Cornell Institutional Animal Care and Use Committees procedures. Mice used for behavioral pain tests were housed at Wake Forest School of Medicine, in accordance with the Wake Forest University Guidelines on the ethical use of animals. The Institutional Animal Care and Use Committee of Wake Forest University approved all pain-related experiments. Animals were housed under a 12-hour light–dark cycle, with food and water ad libitum.

### *Primary cell isolation and generation*

Murine BMDCs were generated by incubation of flushed, single suspended, bone marrow cells isolated from mice of the indicated genotypes in complete RPMI media [RPMI supplemented with L-glutamine, 10% fetal bovine serum (FBS), HEPES, sodium pyruvate, non-essential amino acids, β-mercaptoethanol, and penicillin-streptomycin] containing 20 ng/ml of recombinant GM-CSF (Gemini or Peprotech). Media was replenished on day 6, and cells were harvested on day 7 and used directly for subsequent in vitro functional assays.

Human monocyte-derived DCs were generated by positively isolating $CD14^+$ cells (Miltenyi, catalog number 130-050-201) from blood/buffy coats using Ficoll-gradient centrifugation and plated in complete RPMI media containing 10% FBS and human recombinant GM-CSF (Peprotech) at 1000 IU/ml and IL-4 (Peprotech) at 500 IU/ml for 7 days. Cells were then harvested and used for subsequent in vitro assays (*48*).

Mouse primary macrophages were generated by incubating flushed, single suspended, bone marrow cells from mice of the indicated genotypes in DMEM F12 50/50 media supplemented with L-glutamine, 10% FBS and penicillin–streptomycin, and containing 20 ng/ml recombinant M-CSF (Peprotech) and 1 ng/ml recombinant IL-3 (Peprotech) for 3 days in bacteriological plates. On day 4, non-adherent cells were washed and plated in tissue culture-treated dishes at $1\times10^5$ cells/ml in media containing 20 ng/ml recombinant M-CSF. On day 6, media was replaced and cells were harvested and used for stimulation on day 7.

Primary neutrophils were isolated directly from the bone marrow of *Ern1*$^{f/f}$ or *Ern1*$^{f/f}$ *Vav1*$^{cre}$ mice using negative selection (Miltenyi, catalog #130-097-658) according to the manufacturer protocol. In all cases, isolation purity was greater than 80%. All stimulations were performed in 96-well plates at a volume of 200 μl of media and supernatants were collected after the indicated time points.

### Flow cytometry-based analysis

Murine bone marrow-derived BMDCs were washed with phosphate-buffered saline (PBS), FcγR-blocked using TruStain fcX™ (anti-mouse CD16/32, Biolegend, clone 93, 5 μg/ml) and then stained with antibodies specific for CD11c (Biolegend, clone N418, PE-Cy7, 1 μg/ml) and MHC-II (Tonbo, clone M5/114.15.2, FITC, 1.25 μg/ml), along with live/dead staining using 4′,6-diamidino-2-phenylindole dihydrochloridehydrate (DAPI) (Thermo Fischer Scientific, 0.5 μg/ml). Data were acquired on an LSR II instrument (BD Biosciences).

Single-cell suspensions from ipsilateral paws (described below) were washed, FcγR-blocked using TruStain fcX™, as described above, and stained with antibodies specific for CD45 (BD Biosciences, clone 30-F11, PE-CF594, 1 μg/ml), CD11c (Biolegend, clone N418, APC, 1 μg/ml), MHC-II (Tonbo, clone M5/114.15.2, FITC, 1.25 μg/ml), Ly-6G (Tonbo, clone 1A8, APC-Cy7, 1 μg/ml), CD11b (Tonbo, clone M1/70, PerCP-Cy5.5, 1 μg/ml), F4/80 (Biolegend, clone BM8, PE, 1 μg/ml) along with live/dead staining using DAPI (Thermo Fischer Scientific, 0.5μg/ml). Live $CD45^+$ cells were sorted using BD Aria II SORP cell sorter at the Flow Cytometry Core facility of Weill Cornell Medicine. All flow cytometry data were analyzed with FlowJo software V10 (TreeStar).

### Lipidomic analyses

$5\times10^6$ *Ern1*$^{WT}$ or *Ern1*$^{KO}$ BMDCs were stimulated with 50 ng/ml LPS in six-well plates. Cells were collected after 6 hours, and washed with ice-cold PBS. Cell pellets were then frozen at −80°C until further analysis. Cell pellets were suspended in 850 μl of ice-cold PBS and homogenized using a probe sonicator on ice (three cycles of 10 s each, power and frequency). The homogenate was diluted with 150 μl of methanol containing 10 ng each of prostaglandin $E_1$-d4, resolvin $D_1$-d5, leukotriene $B_4$-d4, 15-HETE-d8, and arachidonic acid-d8, and 100 ng each of cholesteryl heptadecanoate and triheptadecanoyl glycerol (all served as internal standards for the LC-MS analysis). The samples were applied to C18 solid phase extraction cartridge (StrataX C18, Phenomenex) and the lipids were extracted following published procedures (*49*, *50*) with following modifications: The SPE cartridges were eluted with isooctane–ethyl acetate (9:1) first for non-polar lipids (sterol esters, neutral sphingolipids, and triglycerides) before eluting the fatty acyl lipidome with methanol containing 0.1% formic acid. The lipidomic analysis was performed by the Lipidomics Core Facility at Wayne State University by LC-MS using standard protocols. The procedures followed were essentially as described earlier for eicosanomic analysis (*51*–*53*) and by other published procedures for fatty acids, sterol esters, triacyl glycerols, and sphingolipids (*54*–*56*). Volcano plots were generated in R studio using the bioconductor limma package (*57*).

### Immunoblot assays

BMDCs were washed twice in cold PBS and cell pellets were lysed using RIPA lysis buffer (150 mM sodium chloride, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, and 50 mM Tris pH 8.0) supplemented with protease and phosphatase inhibitors (Roche). Homogenates were centrifuged at 22,000×*g* rpm for 30 min at 4°C, and the supernatants were collected. Protein concentrations were determined using BCA protein assay kit (Thermo Fisher Scientific). Equivalent amounts of protein were separated via SDS–PAGE and transferred to PVDF membranes (Immobilon, Millipore) following standard protocols. Membranes were blotted with primary antibodies anti-Cox-2 (Cell Signaling, catalog #12282, 1:1,000), anti-mPGES-1 (Cayman Chemicals, catalog #160140, 1:200), and anti-β actin (Cell Signaling, catalog #4967, 1:1,000), followed by incubation with an HRP-conjugated anti-rabbit secondary antibody (Thermo Fischer Scientific, catalog # G-21234, 1:5,000). SuperSignal West Pico and Femto chemiluminescent substrates (Thermo Fisher Scientific) were used to image blots in a FlourChemE instrument (ProteinSimple).

### $PGE_2$ ELISA

$2.5\times10^5$ cells were stimulated with the indicated compounds and time points, and $PGE_2$ in the supernatants were measured using the $PGE_2$ enzyme-linked immunosorbent assay (ELISA) kit (Enzo Lifesciences, Cat# ADI-900-001). If different number of cells were plated, $PGE_2$ levels were normalized to $2.5\times10^5$ cells/well. Cell viability counts were comparable in all cases. Plates were read at 405 nm using a Varioskan Instrument (Thermo Fischer Scientific). LPS (cat# L2630) and zymosan (cat# Z4250) were purchased from Sigma-Aldrich (St. Louis, MO). Pam3CSK4 (100 ng/ml), HKLM ($5x10^7$ cells), Poly I:C HMW (2 μg/ml), Poly I:C LMW (2 μg/ml), FLA-ST (500 ng/ml), FSL-1 (50 ng/ml), ssRNA (40 μg/ml), ODN-GPG (5 μM) were all purchased from Invivogen (cat# tlrl-kit1mw). PMA (50 μg/ml) was purchased from EMD Millipore (cat# 524400).

### In vivo measurement of prostaglandins using mass spectrometry

#### Materials

HPLC-grade methanol, acetonitrile, and water used for sample purification and LC–MS/MS analysis were JT Baker-brand (ThermoFisher Scientific). The deuterated internal standards $PGE_2$-d4, $PGD_2$-d4, $PGF_{2\alpha}$-d4, 6-keto-$PGF_{1\alpha}$-d4, thromboxane $B_2$-d4 and 15-HETE-d8 were purchased from Cayman Chemicals. Authentic standards of these analytes were also purchased from Cayman. Formic and acetic acid were purchased from Sigma-Aldrich. Acquity chromatography columns were purchased from Waters Corp. (Milford, MA). Solid phase extraction (SPE) cartridges (Strata C18-E, 200 mg/6 ml) were purchased from Phenomenex (Torrance, CA). LC–MS/MS analysis was performed on a Shimadzu Nexera system in-line with a SCIEX 6500 QTrap. The QTrap was equipped with a TurboV Ionspray source and operated in negative ion mode. SCIEX Analyst software (ver 1.6.2) was used to control the instruments and acquire and process the data.

#### Sample collection and purification

Peritoneal lavages were obtained by flushing the abdominal cavity with 3 ml of PBS (pH 7.4). The wash was centrifuged at 400×*g* for 5 min and supernatants were stored at −80°C. One milliliter of cell-free supernatant was removed from −80°C storage, placed on ice, and allowed to thaw. A 10-μl aliquot of internal standard solution was then added to each sample and the sample was vortexed. The samples were then transferred to a 13 × 100-mm test tube. The original sample vials were rinsed with 0.2 ml of 1:1 water:methanol and vortexed. This rinse was also transferred to the 13 × 100-mm test tube. The transferred samples were then diluted with 4.0 ml of 2% acetic acid (aqueous). The diluted samples were applied to pre-conditioned Strata C18 SPE cartridges. The loaded cartridges were washed with 3 ml of 2% acetic acid (aq) followed by 2 ml of 2% acetic acid (aq) with 10% methanol. Finally, the analytes were eluted from cartridges into 12 × 75-mm test tubes with 2 ml of acetonitrile. The samples were dried under nitrogen, capped and stored at −20°C until LC–MS analysis.

#### Prostaglandin Analysis

Immediately prior to LC–MS analysis, the samples were reconstituted in 50 μl of MeOH and 60 μl of $H_2O$, vortexed and transferred onto a 96-well plate. The samples were analyzed on the above-referenced LC–MS system. The analytes were chromatographed on an Acquity UPLC BEH C18 reversed-phase column (5.0- × 0.21-cm; 1.7 μm) which was held at 40°C. A gradient elution profile was applied to each sample; after an initial hold for 0.4 min, %B was increased from 8% (initial conditions) to 70% over 4.0 min, and held at 70% for 1.1 min. Then the column was returned to initial conditions for 1.5 min prior to the next injection. The flow rate was 330 μl/min and component A was water with 0.05% formic acid, whereas component B was acetonitrile with 0.05% formic acid.

All analytes were detected by the SCIEX 6500 QTrap via selected reaction monitoring (SRM) in negative ion mode. The SRM transition (Da) and collision energy (volts) are given for each analyte: $PGE_2$ (351.2 to 233.1, −15); $PGE_2$-d4 (355.2 to 237.1, −15); $PGD_2$ (351.2 to 189.1, −28); $PGD_2$-d4 (355.2 to 193.1, −28); $PGF_{2\alpha}$ (353.2 to 193.1, −34); $PGF_{2\alpha}$-d4 (357.2 to 197.1, −34); 6-keto-$PGF_{1\alpha}$ (369.2 to 163.1, −34); 6-keto-$PGF_{1\alpha}$-d4 (373.2 to 167.1, −34); Thromboxane $B_2$ (369.2 to 169.1, −20); Thromboxane $B_2$-d4 (373.2 to 173.1, −20); 15-HETE (319.2 to 219.1, −25); 15-HETE-d8 (327.2 to 226.1, −25). Analytes were quantitated by stable isotope dilution against their deuterated internal standard. Data were normalized to total number of peritoneal lavage cells recovered.

### ChIP assays

Human monocyte-derived DC were incubated in complete RPMI medium (11.1 mM glucose and 4 mM L-glutamine) in the presence or absence of 10 mM 2-DG and treated with 1 mg/ml zymosan, as previously described (*23*). Cells were then

washed and fixed in 1% formaldehyde for ChIP assays. Cross-linking was terminated using 0.125 M glycine. Nuclear extracts were collected and resuspended in a lysis buffer containing a high salt concentration. Chromatin sonication was carried out using a Bioruptor device from Diagenode (Liege, Belgium). The chromatin solution was precleared by adding Protein A/G PLUS-Agarose for 30 min at 4°C under continuous rotation. After bead removal, isotype control or anti-human XBP1s antibodies (Clone Poly6195, Biolegend) was added for overnight incubation at 4°C, and then incubation with Protein A/G PLUS-Agarose was carried out for 2 hours at 4°C. Beads were pelleted by centrifugation at 16,000×*g* and sequentially washed with lysis buffer high salt, wash buffer, and elution buffer. Cross-links were reversed by heating at 67°C in a water bath, and the DNA bound to the beads isolated by extraction with phenol/chloroform/isoamylalcohol. Irrelevant antibody (Ab) and sequences of the *Pri-miR-21* promoter were used as control of binding specificity. The IRE1α-specific inhibitor utilized in these assays was MKC8866 (*24*–*27*). Results are expressed as percentage of input. Primers used for ChIP-PCR are in table S2. Primer sequences are numbered in bp from the transcription initiation site, but in the case of *Pri-miR-21* it is numbered from the mRNA sequence, which is encoded in chromosome 17, GRCh38.p7. This was selected because of its lack of putative XBP1s-binding sequences.

### RNA-sequencing and bioinformatic analyses

RNA was isolated using RNeasy MinElute kit (Qiagen) from LPS- or zymosan-stimulated murine bone marrow-derived DCs. All samples passed RNA quality control examined by Agilent Bioanalyzer 2100, and mRNA libraries were generated and sequenced at the Weill Cornell Epigenomics Core Facility. RNA-sequencing (RNA-seq) data was aligned using bowtie2 (*58*) against the mm10 mouse genome and RSEM v1.2.12 software (*59*) was used to estimate gene-level read counts using Ensemble transcriptome information. DESeq2 (*60*) was used to estimate significance of differential expression difference between any two experimental groups and gene expression changes of at least 1.2 fold were considered significant if passed false discovery rate (FDR) < 5% thresholds. Gene set enrichment analysis was performed using QIAGEN's Ingenuity® Pathway Analysis software (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) using "Canonical Pathways," "Diseases & Functions," and "Upstream Regulators" options. Enrichment results with at least ten deregulated genes were considered and pathways that passed FDR<5%, functions with *P*-value < $10^{-7}$ and regulators with *P*-value<0.001 were considered significant. Only functions and regulators with significant predicted activation states (|Z-score|>2) were reported. Functions were additionally filtered to remove entries specific to cancer cell lines and immune cell types. Significance of overlap was calculated with hypergeometric test. RNA-seq data was deposited to Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo) under accession no. GSE131404.

### Gene editing in human monocyte-derived DCs

Human CD14$^{+}$ monocytes were isolated from peripheral blood and plated at a density of 5 × $10^{6}$ cells in 3 ml of RPMI supplemented with human recombinant GM-CSF at 1000 IU/ml and IL-4 at 500 IU/ml as described above. On day 6, DCs were prepared for transfection by washing with serum-free PBS and resuspension in RPMI medium supplemented with human recombinant GM-CSF and IL-4, at the same concentrations described above. DCs were then reverse-transfected on a 96-well plate by adding 2.5 × $10^{5}$ cells in suspension onto 150 nM sgRNA-CAS9 ribonucleoprotein complexes containing lipofectamine CRISPRMAX transfection reagent (Invitrogen). All materials for sgRNA-Cas9 complex generation were purchased from Integrated DNA Technologies, and prepared as instructed (*61*). The final sgRNA-Cas9 and CRISPRMAX complex concentrations per well were 50 nM and 1% (vol/vol), respectively. Forty-eight hours post-transfection, genetic ablation of target genes was assessed via quantitative RT-PCR.

The 20-nucleotide CRISPR-RNA (crRNA) targeting human XBP1 (*Homo sapiens* chromosome 22, GRCh38.p12, NC_000022.11) is directed at the genomic sequence 5′-TGCACGTAGTCTGAGTGCTG**CGG**-3′ (the 3 additional nucleotides highlighted in bold represent the protospacer adjacent motif, or PAM). This target sequence corresponds to exon 4 of the human *XBP1* transcript and was manually chosen by identifying a 20-base pair fragment immediately upstream of the highlighted PAM (*62*). The PAM was selected such that Cas9-mediated target DNA cleavage and resulting nucleotide would perturb *XBP1u* recognition and splicing by activated IRE1α (*2*, *63*). The most likely on- and off-target effects of the manually selected CRISPR sequence were then analyzed using the Broad Institute's Genetic Perturbation Platform (*64*). To validate the genomic editing capacity of the crRNA, quantitative RT-PCR was performed on total RNA isolated from cells transfected with sgRNA-Cas9 complexes containing the *XBP1* crRNA described above. The reverse primer for *XBP1s* quantification via quantitative RT-PCR anneals to the same nucleotides as the *XBP1* crRNA target site. Therefore, the primers can only efficiently amplify intact, unperturbed *XBP1s* cDNAs. The primers for evaluating deletion efficacy are listed in Table S2.

The genomic target sequence for the crRNA directed at human *ERN1* (*Homo sapiens* chromosome 17, GRCh38.p12, NC_000017.11) is 5′-ATGTAGAGGATTCCATCTGA**CCC**-3′. This sequence was generated and chosen using the Zhang Lab's crRNA design tool (*65*). To validate the genomic editing capacity of this crRNA, quantitative RT-PCR was performed on total RNA isolated from cells transfected with sgRNA-Cas9 complexes containing *ERN1* crRNA. *XBP1s* levels were used to assess the genetic perturbation of IRE1α (*2*, *63*), using the primer pair specified in table S2. The scrambled crRNA contains a 20-nucleotide sequence that was computationally designed to be non-targeting within the human genome (*61*). The RNA sequence for this non-targeting control is 5′-CGUUAAUCGCGUAUAAUACG-3′.

### Plasmid constructs and luciferase reporter assays

Expression constructs used for luciferase-based assays are pcDNA3.1 XBP1s (NM_001079539.1), pcDNA3.1 CHOP (NM_001195053.1) while reporter constructs used are pGL3-PTGS2 promoter (−1.2 kb/+137) and pGL3-PTGES promoter (−1.3 kb/+35). All plasmids were generated at VectorBuilder.

For dual luciferase assays, 2 × $10^{4}$ HEK293FT cells were plated overnight in a 96-well plate and were transfected with the indicated plasmids using Lipofectamine 3000 (Thermo Fischer Scientific). Eighteen ng of reporter and 2 ng of *Renilla* plasmid were cotransfected with various ratios (w:w) of expression plasmids (reporter: expression plasmid = 1:1, 1:3 or 1:5) and pcDNA3.1, which was added to reach a total of 200 ng of DNA/well. After 36 to 48 hours, cells were washed with PBS and were lysed in Passive Lysis Buffer according to the manufacturer's protocol (Dual luciferase reporter assay system, Promega, catalog #E1960) (*3*). Firefly and *Renilla* luciferase activities were measured in 96-well plates using an automated luminometer (Luminoskan Ascent, Thermo Fischer). Firefly luciferase activity was normalized to its own *Renilla* activity.

### Administration of pharmacological inhibitors

C57BL/6J mice were twice injected intraperitoneally with small-molecule inhibitors (at 6 hours and then 30 min) before acetic acid challenge or paw incision. KIRA6 (Medchem Express, cat# HY-19708), MKC8866 (Medchem Express, cat# HY-104040), or celecoxib (Sigma-Aldrich, cat# PZ0008) were used as indicated. The vehicle used for KIRA6, MKC8866, and celecoxib administration was 7% Tween-80, 3% ethanol, and 90% saline, as previously described (*36*). Briefly, anhydrous compounds were dissolved at 50 mg/ml concentration in dimethyl sulfoxide (DMSO), and then further dissolved in the vehicle at the desired concentrations so that a final volume of 200 μl was administered to each mouse via intraperitoneal injection.

### Single-cell suspensions from mouse paws

Mice were perfused transcardially with 20 ml of 0.1 M phosphate buffer one day after paw incision. Both anterior and posterior parts of the injured or non-injured paw were dissected in a petri dish containing 2 ml of RPMI 164 medium (Gibco). Tissue was dissected into small pieces using surgical scissors, then transferred to a tube containing 2 ml of 0.5 mg/ml of Type II collagenase (Worthigton) in RPMI 1640 (Gibco) and incubated for 2 hours at 37°C shaking at

700 rpm. The enzymatic reaction was stopped by adding 4 ml of 2% FBS (Sigma) in 0.1 M phosphate buffer. Digested tissue was passed through a 40-μm nylon mesh (BD Biosciences) using a syringe plunger. Cell suspension was centrifuged at 450×*g* for 5 min at 4°C and resuspended in 1 ml of 2% FBS in 0.1 M phosphate buffer. Total cell number and cellular viability were determined using trypan blue staining and a hemocytometer. Cells were stored at −80°C in FBS containing 10% DMSO until subsequent flow cytometry analyses were performed.

### Paw incision surgery in mice

Plantar incision surgery was performed as previously described (*35*). Briefly, mice were anesthetized with isoflurane in oxygen (4% induction, 1.5 to 2% for maintenance) and the right hind paw was aseptically cleaned with 10% povidone–iodine solution. Then, a 5-mm incision was made in the glabrous hind-paw skin from the heel to the base of the toes using a No. 11 scalpel and sterile technique. The underlying muscle and ligaments were elevated with a curved forceps and stretched for 6 to 8 s, without incising them. The incision was closed using 5.0 nylon mattress sutures.

### Paw inflammation

Paw perimeter was measured in both left and right hind paws before the surgery and after every behavioral evaluation. The procedure was performed in a consistent manner using a 4.0 silk thread, which was placed around the center of the surgery in the right paw and at the same level in the paw contralateral to surgery. An increase in the paw perimeter was considered to reflect augmented inflammation of the affected paw.

### Behavioral tests

All behavioral measurements were performed by a blinded observer before and after surgery (postoperative days 1 to 21) or acetic acid intraperitoneal injection (0 to 30 min). Animals were acclimated to the testing devices and/or places for 3 days, and baseline measurements were taken for at least 4 consecutive days before surgery or acetic acid injection.

Writhing spontaneous pain behaviors were evaluated after intraperitoneal injection of 0.9% acetic acid (v/v, 5 ml/kg). The number of writhing responses was quantified immediately after acetic acid injection for 30 min in 5-min intervals by an observer blinded to genotype. Writhings induced by acetic acid are overt stretching behaviors indicative of abdominal pain, a phenomenon that is dependent upon mPGES-1 and $PGE_2$ (*31*, *32*). Spontaneous exploratory activity (ambulatory counts and ambulatory time) was measured for 15 min immediately after injection of acetic acid in independent groups. We used a commercially available equipment and software, which was set according to the manufacturer's instructions (Med Associates, St. Albans, VT). Briefly, mice were placed in the center of an acrylic box (42.5 × 42.5 × 37.5 cm) with banks of 16 beams transmitters spaced 2.5 cm apart in both the *x* and *y* directions and 7 cm above the floor of the box (*z* plane).

Various spontaneous (non-evoked) pain-related behaviors were also evaluated postsurgery: weight bearing, grimace score, guarding score, rearing, and flinching activity. Mice were placed in individual acrylic chambers on an elevated mesh floor for 30 to 45 min before testing. Following the acclimatization period, the number of total vertical rearings and paw flinches were quantified during a 2-min period. Vertical rearings were defined as the number of times that the animal stood supporting its weight on both hind limbs. Vertical rearings are a normal behavior in rodents. Thus, a reduction in this behavior was indicative of a protective response to pain produced by movement, which mimics pain induced by surgeries in humans. Spontaneous flinching of the affected paw was quantified every time that the animal shaked the affected paw without any stimulation. Flinches of the injured paw is a pain-related behavior that is indicative of breakthrough pain, similar to intense spontaneous spike of pain in humans with postoperative pain. When all animals had been evaluated for flinching behaviors, they were observed again in sequential order for guarding behaviors using the guarding score method as follows (*66*). Mice were observed for 1 min every 10 min for a total of three times and scores are averaged for a final value. A score was given depending on the position in which the injured paw is found during the majority of the 1-min scoring period, and the following score system was used: 0 = full weight bearing of the paw is present if the wound is blanched or distorted by the mesh; 1 = the area of the wound touched the mesh without blanching or distorting; and 2 = the paw is completely off the mesh. After completion of guarding assessment, the animals were observed again to evaluate facial expressions using the Mouse Grimace Scale (*67*). Five facial expression parameters were evaluated: orbital tightening, nose bulge, cheek bulge, ear position (ear pulled backward), and whisker position (backward, forward, clumped together). Each parameter was scored using the following scale: 0 = not present; 1 = moderately present, and 2 = severely present. This evaluation was performed two times 5 min apart, and the values were averaged to have a single score per time point after surgery. These values were used to subtract the baseline value of the grimace scale (before surgery), and therefore the presented value for each animal is the mouse grimace scale difference score.

Mechanical hypersensitivity, an evoked pain-related behavior, was evaluated after all previous assessments. Mechanical withdrawal thresholds were calculated using the up-down method and applying force with calibrated von Frey filaments (0.07 G, 0.17 G, 0.40 G, 0.60 G, 1.04 G, 1.37 G, and 2.0 G, Stoeling, Wood Dale, IL, United States) to the plantar aspect of the paw for 5 s. Withdrawal of paws or flinching in response to a given applied force was noted as a positive response.

Hind paw weight bearing distribution was determined using an incapacitance tester apparatus (Stoelting, IL, version 5.64). This is a test for non-reflexive behaviors that represents a spontaneous pain-related behavior that mimics postoperative pain behaviors in humans (protection of the surgery site from normal activities). Before surgery, animals were habituated for at least 3 days to the apparatus, in which animals stand with each hind paw resting on individual weight plates inside an acrylic chamber. The apparatus measures the body weight distributed between the two hind paws over a 3-s period, and provide the average measurement. We used the average value of each hind paw to determine the weight distribution ratio (ipsilateral/contralateral side). A ratio below one indicates a greater weight bearing on the contralateral paw and was therefore considered as a pain-related behavior.

### Immunohistochemistry

Mice were anesthetized with isoflurane (3 to 4% in oxygen) and perfused transcardially with 20 ml of filtered 0.1 M phosphate-buffered saline (PBS) followed by 20 ml of 4% formaldehyde. Tissue around the injured paw was collected by making a rectangular incision around the injury about 1.5 mm apart from the center of the surgery. Skin and muscle associated with the incision were collected and postfixed for 3 hours in 4% formaldehyde at 4°C. Tissue was stored at 4°C in 30% sucrose solution for 72 hours before sectioning. Slices of tissue were cut at 18 μm using optimal cutting temperature compound (Sakura Finetek) in a Leica cryostat and placed on coated slides (VWR International). Slides were then washed three times for 5 min with 0.1 M PBS and blocked using a solution of 3% normal donkey serum (NDS) + 0.3% Triton X-100 in 0.1 M PBS for 1 hour at room temperature. Primary antibodies used were rabbit anti-Cox-2 (Cell Signaling, catalog #12282, 0.203 μg/ml) and rat anti-CD45 (BioRad, catalog #MCA1388, 5 μg/ml). Tissues with primary antibodies were incubated overnight at 4°C. Tissues were then washed three times for 5 min with 0.1 M PBS and incubated 2 hours at room temperature with corresponding secondary antibodies: Cyanine 2-conjugated donkey anti-rabbit immunoglobulin G (IgG) (Jackson Immuno Research Labs, catalog #711-225-152, 3.75 μg/ml) and Cyanine 3-conjugated donkey anti-rat IgG (Jackson Immuno Research Labs, catalog #712-165-150, 3.75 μg/ml). Finally, slides were rinsed three times and mounted using anti-fade medium containing DAPI (Invitrogen) to allow visualization of cell nuclei.

At least three images per slide were taken at 20X from areas adjacent to the incision, using a Nikon Eclipse Ni fluorescent microscope system (Nikon, Japan) equipped with a Nikon DS-Qi1Mc digital camera (Nikon). All images were acquired using the NIS elements software Version 4.2 (Nikon, Japan). In each micrograph, $CD45^+$ or $Cox\text{-}2^+$ cells were quantified by a blinded observer in three random squares of 100 $\mu m^2$ each using Adobe Photoshop CS6 (Adobe Systems). The percentage of $Cox\text{-}2^+$ cells was then calculated in relationship to the total $CD45^+$ cells by

a blinded observer. For colocalization studies, images were acquired with an Olympus FV1200 confocal microscope using a 40X objective lens (UPLFLN40XO/1.3) and 405 nm, 488 nm, and 559 nm laser lines. Confocal images were prepared with Olympus Fluoview Version 4.2b software and Adobe Photoshop software. All images were taken from adjacent areas of the surgical wound ipsilateral to paw incision.

### Statistical analyses

All statistical analyses were performed using the GraphPad Prism 7.0 software. Comparisons between two groups were assessed using unpaired two-tailed Student's *t*-test, unless otherwise stated. All grouped data (time course) were analyzed using two-way analysis of variance (ANOVA) and Sidak's multiple comparisons test. Grimace and guarding scores postsurgery were analyzed using the Mann–Whitney *U* test. Data are presented as mean ± standard error of the mean (SEM), unless otherwise stated. *P*-values lower than 0.05 were considered to be statistically significant.

## REFERENCES AND NOTES


1. S. E. Bettigole, L. H. Glimcher, Endoplasmic reticulum stress in immunity. *Annu. Rev. Immunol.* **33**, 107–138 (2015). doi: 10.1146/annurev-immunol-032414-112116; pmid: 25493331
2. H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107**, 881–891 (2001). doi: 10.1016/S0092-8674(01)00611-0; pmid: 11779464
3. A. H. Lee, N. N. Iwakoshi, L. H. Glimcher, XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol. Cell. Biol.* **23**, 7448–7459 (2003). doi: 10.1128/MCB.23.21.7448-7459.2003; pmid: 14559994
4. A. H. Lee, E. F. Scapa, D. E. Cohen, L. H. Glimcher, Regulation of hepatic lipogenesis by the transcription factor XBP1. *Science* **320**, 1492–1496 (2008). doi: 10.1126/science.1158042; pmid: 18556558
5. X. Chen *et al*., XBP1 promotes triple-negative breast cancer by controlling the HIF1α pathway. *Nature* **508**, 103–107 (2014). doi: 10.1038/nature13119; pmid: 24670641
6. Q. Duan *et al*., Deregulation of XBP1 expression contributes to myocardial vascular endothelial growth factor-A expression and angiogenesis during cardiac hypertrophy in vivo. *Aging Cell* **15**, 625–633 (2016). doi: 10.1111/acel.12460; pmid: 27133203
7. A. X. Zhou, I. Tabas, The UPR in atherosclerosis. *Semin. Immunopathol.* **35**, 321–332 (2013). doi: 10.1007/s00281-013-0372-x; pmid: 23553213
8. O. Tufanli *et al*., Targeting IRE1 with small molecules counteracts progression of atherosclerosis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1395–E1404 (2017). doi: 10.1073/pnas.1621188114; pmid: 28137856
9. Q. Qiu *et al*., Toll-like receptor-mediated IRE1α activation as a therapeutic target for inflammatory arthritis. *EMBO J.* **32**, 2477–2490 (2013). doi: 10.1038/emboj.2013.183; pmid: 23942232
10. J. R. Cubillos-Ruiz, S. E. Bettigole, L. H. Glimcher, Tumorigenic and immunosuppressive effects of endoplasmic reticulum stress in cancer. *Cell* **168**, 692–706 (2017). doi: 10.1016/j.cell.2016.12.004; pmid: 28187289
11. J. R. Cubillos-Ruiz *et al*., ER stress sensor XBP1 controls anti-tumor immunity by disrupting dendritic cell homeostasis. *Cell* **161**, 1527–1538 (2015). doi: 10.1016/j.cell.2015.05.025; pmid: 26073941
12. M. Song *et al*., IRE1α-XBP1 controls T cell function in ovarian cancer by regulating mitochondrial activity. *Nature* **562**, 423–428 (2018). doi: 10.1038/s41586-018-0597-x; pmid: 30305738
13. F. Martinon, X. Chen, A. H. Lee, L. H. Glimcher, TLR activation of the transcription factor XBP1 regulates innate immune responses in macrophages. *Nat. Immunol.* **11**, 411–418 (2010). doi: 10.1038/ni.1857; pmid: 20351694
14. J. S. So *et al*., Silencing of lipid metabolism genes through IRE1α-mediated mRNA decay lowers plasma lipids in mice. *Cell Metab.* **16**, 487–499 (2012). doi: 10.1016/j.cmet.2012.09.004; pmid: 23040070
15. J. Hollien *et al*., Regulated Ire1-dependent decay of messenger RNAs in mammalian cells. *J. Cell Biol.* **186**, 323–331 (2009). doi: 10.1083/jcb.200903014; pmid: 19651891
16. F. Osorio *et al*., The unfolded-protein-response sensor IRE-1α regulates the function of CD8α$^+$ dendritic cells. *Nat. Immunol.* **15**, 248–257 (2014). doi: 10.1038/ni.2808; pmid: 24441789
17. C. Hetz, F. R. Papa, The unfolded protein response and cell fate control. *Mol. Cell* **69**, 169–181 (2018). doi: 10.1016/j.molcel.2017.06.017; pmid: 29107536
18. A. J. Dannenberg, K. Subbaramaiah, Targeting cyclooxygenase-2 in human neoplasia: Rationale and promise. *Cancer Cell* **4**, 431–436 (2003). doi: 10.1016/S1535-6108(03)00310-6; pmid: 14706335
19. E. Ricciotti, G. A. FitzGerald, Prostaglandins and inflammation. *Arterioscler. Thromb. Vasc. Biol.* **31**, 986–1000 (2011). doi: 10.1161/ATVBAHA.110.207449; pmid: 21508345
20. J. de Boer *et al*., Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33**, 314–325 (2003). doi: 10.1002/immu.200310005; pmid: 12548562
21. D. Acosta-Alvear *et al*., XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks. *Mol. Cell* **27**, 53–66 (2007). doi: 10.1016/j.molcel.2007.06.011; pmid: 17612490
22. I. M. Clauss, M. Chu, J. L. Zhao, L. H. Glimcher, The basic domain/leucine zipper protein hXBP-1 preferentially binds to and transactivates CRE-like sequences containing an ACGT core. *Nucleic Acids Res.* **24**, 1855–1864 (1996). doi: 10.1093/nar/24.10.1855; pmid: 8657566
23. S. Márquez *et al*., Endoplasmic reticulum stress sensor IRE1α enhances IL-23 expression by human dendritic cells. *Front. Immunol.* **8**, 639 (2017). doi: 10.3389/fimmu.2017.00639; pmid: 28674530
24. K. Volkmann *et al*., Potent and selective inhibitors of the inositol-requiring enzyme 1 endoribonuclease. *J. Biol. Chem.* **286**, 12743–12755 (2011). doi: 10.1074/jbc.M110.199737; pmid: 21303903
25. S. E. Logue *et al*., Inhibition of IRE1 RNase activity modulates the tumor cell secretome and enhances response to chemotherapy. *Nat. Commun.* **9**, 3267 (2018). doi: 10.1038/s41467-018-05763-8; pmid: 30111846
26. X. Sheng *et al*., IRE1α-XBP1s pathway promotes prostate cancer by activating c-MYC signaling. *Nat. Commun.* **10**, 323 (2019). doi: 10.1038/s41467-018-08152-3; pmid: 30679434
27. N. Zhao *et al*., Pharmacological targeting of MYC-regulated IRE1/XBP1 pathway suppresses MYC-driven breast cancer. *J. Clin. Invest.* **128**, 1283–1299 (2018). doi: 10.1172/JCI95873; pmid: 29480818
28. C. D. Funk, Prostaglandins and leukotrienes: Advances in eicosanoid biology. *Science* **294**, 1871–1875 (2001). doi: 10.1126/science.294.5548.1871; pmid: 11729303
29. J. L. Stock *et al*., The prostaglandin E2 EP1 receptor mediates pain perception and regulates blood pressure. *J. Clin. Invest.* **107**, 325–331 (2001). doi: 10.1172/JCI6749; pmid: 11160156
30. A. Kawabata, Prostaglandin E2 and pain—An update. *Biol. Pharm. Bull.* **34**, 1170–1173 (2011). doi: 10.1248/bpb.34.1170; pmid: 21804201
31. D. Kamei *et al*., Reduced pain hypersensitivity and inflammation in mice lacking microsomal prostaglandin e synthase-1. *J. Biol. Chem.* **279**, 33684–33695 (2004). doi: 10.1074/jbc.M400199200; pmid: 15140897
32. C. E. Trebino *et al*., Impaired inflammatory and pain responses in mice lacking an inducible prostaglandin E synthase. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9044–9049 (2003). doi: 10.1073/pnas.1332766100; pmid: 12835414
33. H. O. Collier, L. C. Dinneen, C. A. Johnson, C. Schneider, The abdominal constriction response and its suppression by analgesic drugs in the mouse. *Br. J. Pharmacol. Chemother.* **32**, 295–310 (1968). doi: 10.1111/j.1476-5381.1968.tb00973.x; pmid: 4230818
34. Z. H. Lu *et al*., Evaluation of 2 celecoxib derivatives: Analgesic effect and selectivity to cyclooxygenase-2/1. *Acta Pharmacol. Sin.* **26**, 1505–1511 (2005). doi: 10.1111/j.1745-7254.2005.00222.x; pmid: 16297351
35. E. M. Pogatzki, S. N. Raja, A mouse model of incisional pain. *Anesthesiology* **99**, 1023–1027 (2003). doi: 10.1097/00000542-200310000-00041; pmid: 14508341
36. R. Ghosh *et al*., Allosteric inhibition of the IRE1α RNase preserves cell viability and function during endoplasmic reticulum stress. *Cell* **158**, 534–548 (2014). doi: 10.1016/j.cell.2014.07.002; pmid: 25018104
37. A. M. Keestra-Gounder *et al*., NOD1 and NOD2 signalling links ER stress with inflammation. *Nature* **532**, 394–397 (2016). doi: 10.1038/nature17631; pmid: 27007849
38. E. M. Pogatzki-Zahn, D. Segelcke, S. A. Schug, Postoperative pain—From mechanisms to treatment. *Pain Rep.* **2**, e588 (2017). doi: 10.1097/PR9.0000000000000588; pmid: 29392204
39. G. T. Whiteside *et al*., Pharmacological characterisation of a rat model of incisional pain. *Br. J. Pharmacol.* **141**, 85–91 (2004). doi: 10.1038/sj.bjp.0705568; pmid: 14597606
40. A. Buvanendran *et al*., Upregulation of prostaglandin E2 and interleukins in the central nervous system and peripheral tissue during and after surgery in humans. *Anesthesiology* **104**, 403–410 (2006). doi: 10.1097/00000542-200603000-00005; pmid: 16508385
41. Y. Q. Zhao *et al*., The Analgesic Effects of Celecoxib on the formalin-induced short- and long-term inflammatory pain. *Pain Physician* **20**, E575–E584 (2017). pmid: 28535566
42. J. J. Chen *et al*., The pre-emptive analgesic effect of a cyclooxygenase-2 inhibitor in a rat model of acute postoperative pain. *Anaesthesia* **67**, 1225–1231 (2012). doi: 10.1111/j.1365-2044.2012.07246.x; pmid: 22823006
43. D. Vardeh *et al*., COX2 in CNS neural cells mediates mechanical inflammatory pain hypersensitivity in mice. *J. Clin. Invest.* **119**, 287–294 (2009). pmid: 19127021
44. A. F. Gaspar, W. A. Prado, Comparison of pre- versus post-incision administration of intraplantar indomethacin and MK886 in a rat model of postoperative pain. *Braz. J. Med. Biol. Res.* **40**, 1141–1147 (2007). doi: 10.1590/S0100-879X2006005000136; pmid: 17665052
45. P. S. Ostling *et al*., America's opioid epidemic: A comprehensive review and look into the rising crisis. *Curr. Pain Headache Rep.* **22**, 32 (2018). doi: 10.1007/s11916-018-0685-5; pmid: 29619569
46. A. H. Lee, N. N. Iwakoshi, K. C. Anderson, L. H. Glimcher, Proteasome inhibitors disrupt the unfolded protein response in myeloma cells. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9946–9951 (2003). doi: 10.1073/pnas.1334037100; pmid: 12902539
47. T. Iwawaki, R. Akai, S. Yamanaka, K. Kohno, Function of IRE1 alpha in the placenta is essential for placental development and embryonic viability. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16657–16662 (2009). doi: 10.1073/pnas.0903775106; pmid: 19805353
48. S. Nair, G. E. Archer, T. F. Tedder, Isolation and generation of human dendritic cells. *Curr. Protoc. Immunol.* **chap. 7**, unit7 32 (2012).
49. K. R. Maddipati, S. L. Zhou, Stability and analysis of eicosanoids and docosanoids in tissue culture media. *Prostaglandins Other Lipid Mediat.* **94**, 59–72 (2011). doi: 10.1016/j.prostaglandins.2011.01.003; pmid: 21236355
50. J. F. Markworth *et al*., Human inflammatory and resolving lipid mediator responses to resistance exercise and ibuprofen treatment. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **305**, R1281–R1296 (2013). doi: 10.1152/ajpregu.00128.2013; pmid: 24089379
51. K. R. Maddipati *et al*., Eicosanomic profiling reveals dominance of the epoxygenase pathway in human amniotic fluid at term in spontaneous labor. *FASEB J.* **28**, 4835–4846 (2014). doi: 10.1096/fj.14-254383; pmid: 25059230
52. K. R. Maddipati *et al*., Lipidomic analysis of patients with microbial invasion of the amniotic cavity reveals up-regulation of leukotriene B4. *FASEB J.* **30**, 3296–3307 (2016). doi: 10.1096/fj.201600583R; pmid: 27312808
53. K. R. Maddipati *et al*., Clinical chorioamnionitis at term: The amniotic fluid fatty acyl lipidome. *J. Lipid Res.* **57**, 1906–1916 (2016). doi: 10.1194/jlr.P069096; pmid: 27538821
54. R. L. Shaner *et al*., Quantitative analysis of sphingolipids for lipidomics using triple quadrupole and quadrupole linear ion trap mass spectrometers. *J. Lipid Res.* **50**, 1692–1707 (2009). doi: 10.1194/jlr.D800051-JLR200; pmid: 19036716
55. C. Hellmuth, M. Weber, B. Koletzko, W. Peissner, Nonesterified fatty acid determination for functional lipidomics: Comprehensive ultrahigh performance liquid chromatography-tandem mass spectrometry quantitation, qualification, and parameter prediction. *Anal. Chem.* **84**, 1483–1490 (2012). doi: 10.1021/ac202602u; pmid: 22224852
56. P. M. Hutchins, R. M. Barkley, R. C. Murphy, Separation of cellular nonpolar neutral lipids by normal-phase chromatography and analysis by electrospray ionization mass spectrometry.

*J. Lipid Res.* **49**, 804–813 (2008). doi: 10.1194/jlr.M700521-JLR200; pmid: 18223242
57. https://bioconductor.org/packages/release/bioc/html/limma.html.
58. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
59. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). doi: 10.1186/1471-2105-12-323; pmid: 21816040
60. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
61. https://sfvideo.blob.core.windows.net/sitefinity/docs/default-source/user-guide-manual/alt-r-crispr-cas9-user-guide-ribonucleoprotein-transfections-recommended.pdf?sfvrsn=1c43407_24.
62. F. A. Ran *et al*., Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013). doi: 10.1038/nprot.2013.143; pmid: 24157548
63. M. Calfon *et al*., IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415**, 92–96 (2002). doi: 10.1038/415092a; pmid: 11780124
64. https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design.
65. http://crispr.mit.edu.
66. T. J. Brennan, E. P. Vandermeulen, G. F. Gebhart, Characterization of a rat model of incisional pain. *Pain* **64**, 493–501 (1996). doi: 10.1016/0304-3959(95)01441-1; pmid: 8783314
67. D. J. Langford *et al*., Coding of facial expressions of pain in the laboratory mouse. *Nat. Methods* **7**, 447–449 (2010). doi: 10.1038/nmeth.1455; pmid: 20453868

## ACKNOWLEDGMENTS

We thank J. McCormick for expert assistance with flow cytometry and all members of the Weill Cornell Epigenomics Core Facility for help with RNA-seq. We thank K. Maddipati and the lipidomics facility at Wayne State University for their support with lipidomics experiments and analyses. We also thank D. Simmons, S. Bettigole, and all members of the Cubillos-Ruiz, Glimcher, and Romero-Sandoval laboratories for helpful suggestions and critical reading of this manuscript. We are grateful to D. K. Morales for creating the model shown in the summary page. **Funding:** Our research was supported by the Early-Career Investigator Award W81XWH-16-1-0438 of the Department of Defense (J.R.C.-R.), The Pershing Square Sohn Cancer Research Alliance (J.R.C.-R.), Weill Cornell Medicine Funds (J.R.C.-R. and L.H.G.), Department of Anesthesiology-Wake Forest School of Medicine Funds (E.A.R.-S.), NIH grant R01CA112663 (L.H.G.), Plan Nacional de Salud y Farmacia Grant SAF2017-83079-R (M.S.C.), and NIH grant 1S10OD017997-01A1 (L.J.M.). **Author contributions:** S.C. conceived, designed and conducted in vitro and in vivo experiments, analyzed data, and wrote the manuscript. P.G. designed and performed CRISPR-related experiments along with other in vitro and in vivo analyses. P.A.A.-V., M.M.F., and E.A.R.-S. performed pain-related in vivo and in vitro experiments, conducted behavioral tests in mice, and analyzed data. M.S. performed in vitro and in vivo experiments. L.J. performed analysis of lipidomic data. S.A. and M.S.C. designed and performed experiments related to ChIP-PCR. S.G. and M.M.F. performed pain-related and immunofluorescence experiments. C-S.C., T.A.S., and C.T. performed diverse in vitro experiments. T.I. provided *Ern1*-floxed transgenic mice. P.J.K. and L.J.M. performed and analyzed mass spectrometry assays for in vivo samples. A.V.K. carried out advanced computational analyses of RNA-seq data. A.J.D. and K.S. contributed with ideas and experimental systems, analyzed data, and reviewed the manuscript. L.H.G. provided *Xbp1*-floxed mice and critical resources, analyzed and interpreted data, and reviewed the manuscript. E.A.R.-S. and J.R.C.-R. conceived and designed the research, performed experiments, analyzed and interpreted data, wrote the manuscript, and administered the project. **Competing interests:** L.H.G. and J.R.C.-R are cofounders of and scientific advisors for Quentis Therapeutics. L.H.G. is a former director of Bristol-Myers Squibb Pharmaceuticals, currently serves on the Board of Directors of and holds equity in GlaxoSmithKline Pharmaceuticals and Waters Company, and is on the scientific advisory boards of Abpro, Kaleido, and Repare Therapeutics. **Data and materials availability:** All data are available in the main text or the supplementary materials. RNA-seq data was deposited to GEO (www.ncbi.nlm.nih.gov/geo) under accession no GSE131404. *Ern1*-floxed or *Xbp1*-floxed mice may be obtained under a materials transfer agreement from T.I. or L.H.G., respectively.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/eaau6499/suppl/DC1
Figs. S1 to S15
Tables S1 and S2

3 July 2018; resubmitted 27 March 2019
Accepted 10 June 2019
10.1126/science.aau6499

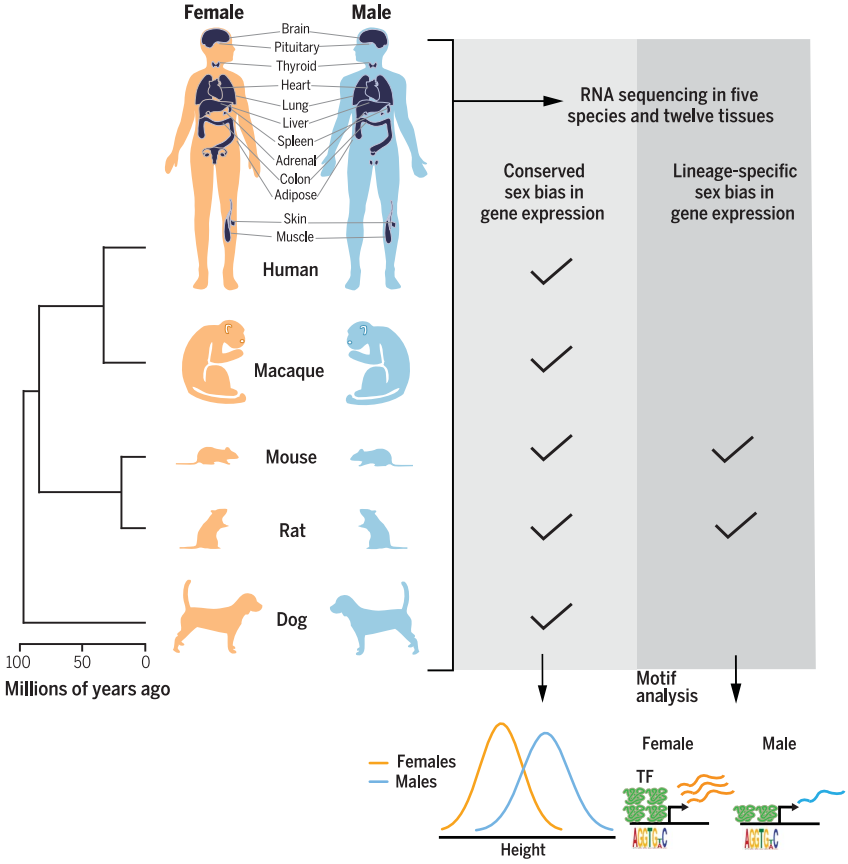
RESEARCH ARTICLE SUMMARY

COMPARATIVE GENETICS

# Conservation, acquisition, and functional impact of sex-biased gene expression in mammals

**Sahin Naqvi, Alexander K. Godfrey, Jennifer F. Hughes, Mary L. Goodheart, Richard N. Mitchell, David C. Page***

**INTRODUCTION:** Sex differences are widespread in humans and other mammals. For example, the distribution of height or body size is shifted upwards in males relative to females, and sex differences are found in the immune and cardiovascular systems as well as in metabolism. However, little is known about how gene expression differs between the sexes in a broad range of mammalian tissues and species. A catalog of such sex-biased gene expression could help us understand phenotypic sex differences. Assessing the extent to which sex-biased gene expression is conserved across the body could also have important implications for the use of nonhuman mammals as models of sex-biased human biology.




**RNA sequencing of male and female samples in 12 tissues and five species reveals the functional impact and mechanistic underpinnings of sex-biased gene expression.** A survey of sex differences in gene expression using RNA sequencing data (left) leads to the discovery of both conserved (species-shared) and lineage- or species-specific sex biases in expression across the genome. Genes with conserved sex bias contribute to the sex difference in mean height in humans and other mammals, whereas lineage-specific changes can be partially explained by gains and losses of motifs for sex-biased TFs.

**RATIONALE:** To identify both conserved and lineage- or species-specific sex differences in gene expression, we sequenced RNA from male and female samples in 12 tissues in each of four nonhuman mammals (cynomolgus macaque, mouse, rat, and dog) and analyzed these data jointly with publicly available data from postmortem male and female human tissues. To assess the impact of sex-biased gene expression on the sex difference in mean human height, we applied methods that integrate the effects of genetic variation on both gene expression and phenotype (height in this case). We sought to understand which transcription factors (TFs) contribute to evolutionary changes in sex bias by analyzing motifs gained or lost concurrently with lineage- or species-specific changes in sex bias.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw7317

**RESULTS:** Linear modeling revealed ~3000 genes with conserved (species-shared) sex bias in gene expression, most of which was tissue specific. The cumulative effects of conserved sex bias explain ~12% of the sex difference in mean human height, and cases such as that of *LCORL*, a TF with conservation of both female-biased expression and genetic association with height, suggest a contribution to sex differences in body size beyond humans. However, most sex-biased gene expression (~77%) was specific to single species or subsets of species, implying that it arose more recently during evolution. We identified 83 instances where TFs showed sex-biased expression in the same tissue, in which their motifs were associated with gain or loss of sex bias at other genes, accounting for a significant portion (~27%) of lineage-specific changes in sex bias.

**CONCLUSION:** By conducting a 12-tissue, five-species survey of sex differences in gene expression, we found that although conserved sex bias in gene expression exists throughout the body, most sex bias has been acquired more recently during mammalian evolution. Height is likely subject to opposing selective pressures in males and females; our study thus documents how such selective forces can result in sex-biased expression which, when layered upon genetic pathways acting identically in males and females, can lead to trait distributions shifted between the sexes. Our findings also suggest that, in many cases, molecular sex differences observed in humans may not be mirrored in nonhuman mammals. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dcpage@wi.mit.edu
Cite this article as S. Naqvi *et al.*, *Science* **365**, eaaw7317 (2019). DOI: 10.1126/science.aaw7317

RESEARCH ARTICLE

COMPARATIVE GENETICS

# Conservation, acquisition, and functional impact of sex-biased gene expression in mammals

Sahin Naqvi[1,2], Alexander K. Godfrey[1,2], Jennifer F. Hughes[1], Mary L. Goodheart[1,3], Richard N. Mitchell[4], David C. Page[1,2,3]*

**Sex differences abound in human health and disease, as they do in other mammals used as models. The extent to which sex differences are conserved at the molecular level across species and tissues is unknown. We surveyed sex differences in gene expression in human, macaque, mouse, rat, and dog, across 12 tissues. In each tissue, we identified hundreds of genes with conserved sex-biased expression—findings that, combined with genomic analyses of human height, explain ~12% of the difference in height between females and males. We surmise that conserved sex biases in expression of genes otherwise operating equivalently in females and males contribute to sex differences in traits. However, most sex-biased expression arose during the mammalian radiation, which suggests that careful attention to interspecies divergence is needed when modeling human sex differences.**

Males and females exhibit differences across a wide range of biological processes. Studies in humans have documented sex differences in anthropometric traits (*1*), energy metabolism (*2*), brain morphology (*3*), and immune (*4*) and cardiac (*5*) function. Sex differences are also evident in the incidence, prevalence, and mortality across diseases, including autoimmune disorders (*6*), cardiovascular diseases (*7*), and autism (*8*). Sex differences are common in other mammals besides humans, many of which are models of sex-biased human traits and diseases (*9*). For example, males are larger than females in most mammalian species (*10*), whereas sex differences in brain structures (*11*) and immune (*12*) and cardiac (*13*) function have been observed in rodents. These phenotypic sex differences are likely associated with, and may be caused by, sex differences in gene activity or function.

The sex chromosomes are one source of sex differences in gene activity. The Y chromosome harbors male-specific genes (*14*), some broadly expressed (*15*). Incomplete inactivation of the second X chromosome in females results in female-biased expression of some X-linked genes (*16*). However, given the scale and complexity of gene networks, and the greater number of autosomal genes, it is unlikely that sexually dimorphic expression of sex-linked genes accounts for all phenotypic sex differences in mammals. Understanding the molecular origins of these sex differences therefore requires a genome-wide, multitissue, and comparative approach to sex biases in gene expression.

Our understanding of sex bias in mammalian gene expression is lacking in three regards. First, the degree to which sex-biased expression is conserved across the mammalian lineage and the extent of conservation in different tissues and organ systems are unknown. Multitissue studies of sex bias in gene expression focused on humans (*17*, *18*) or mice (*19*). Multispecies studies in *Drosophila* (*20*–*24*) examined RNA from whole carcasses or gonads, whereas studies in mammals that examined nonreproductive tissues focused on single tissues (*25*, *26*). Second, little is known about how sex differences in gene expression across the body cumulatively result in phenotypic sex differences. Sex-biased expression of the autosomal genes *VGLL3* (*27*) and *IL-33* (*28*), as well as the X-linked gene *TLR7* (*29*), appears to contribute to sexually dimorphic immune phenotypes. However, most complex traits are polygenic and underpinned by variation in hundreds or even thousands of genes (*30*). Third, apart from single-gene studies in *Drosophila* (*31*), lineage-specific regulatory changes that drive the evolution of sex-biased expression remain unexplored. Progress has been made in understanding mechanisms of X-linked dosage compensation (*32*, *33*), the lack of which can lead to sex-biased expression on the X chromosome, but additional mechanisms likely contribute to genome-wide sex-biased gene expression. Thus, previous studies sought to understand the extent of sex-biased expression across either tissues (*17*–*19*) or species (*25*, *26*), or they explored its phenotypic impact (*27*–*29*) or underlying evolutionary mechanisms (*31*) for individual genes. Assessing sex-biased expression across tissues and species, together with its cumulative contribution to phenotypic sex differences, would advance our understanding of molecular differences between males and females.

## Results

### A five-species, 12-tissue survey of sex differences in gene expression

To assess sex differences in nonhuman mammals, we collected RNA sequencing data from three males and three females from cynomolgus macaque (*Macaca fascicularis*, cyno), mouse (*Mus musculus*), rat (*Rattus norvegicus*), and dog (*Canis familiaris*). Together with humans, these five species, whose last common ancestor lived 80 to 100 million years ago, span the evolution of the Boreoeutheria, including all placental mammals except Afrotheria and Xenarthra (which include the elephant and anteater, respectively). We sampled 12 tissues from each individual: adipose, adrenal gland, brain, colon, heart, liver, lung, muscle, pituitary, skin, spleen, and thyroid. These tissues represent many organ systems and all three germ layers (Fig. 1A). We designed tissue collection and processing procedures to minimize biological and technical variation (*34*) (table S1). We used our RNA sequencing (RNA-seq) data to systematically improve the transcriptome annotations of each nonhuman mammalian species, which we then assessed using the percentage of reads from independent studies that mapped to our annotations versus existing annotations (e.g., a 16% increase in read mapping rate in dog) (fig. S1).

To assess sex differences in humans, we analyzed RNA-seq data from the Genotype-Tissue Expression Consortium (GTEx, v6p release) (*35*). To reduce the possibility of sex biases in cell-type composition, pathology, or other factors driving our results, we performed stringent quality control for samples from each of the 12 target tissues using individual- and sample-level metadata from GTEx and our own evaluation of histological images (*34*) (table S2). We adjusted gene expression values using top principal components to remove variation due to hidden technical or biological confounders. In three tissues (adipose, brain, and skin) for which expression data from purified cell populations is available, there is a correlation between sample-level cell-type proportions estimated by CIBERSORT (*36*) and top principal component loadings (fig. S2). Although this approach controls for variation in cell-type composition in the human samples, we acknowledge that some sex biases, especially those specific to nonhuman mammals, could reflect sex differences in cell-type composition.

We removed outlier samples (*34*) to obtain 740 human and 277 nonhuman RNA-seq samples (see table S3 for human sample sizes by sex and tissue). We clustered all nonhuman samples and a randomly chosen subset of human samples, using the expression levels of 12,939 one-to-one orthologous protein-coding genes.

[1]Whitehead Institute, Cambridge, MA 02142, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Howard Hughes Medical Institute, Whitehead Institute, Cambridge, MA 02142, USA. [4]Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.
*Corresponding author. Email: dcpage@wi.mit.edu

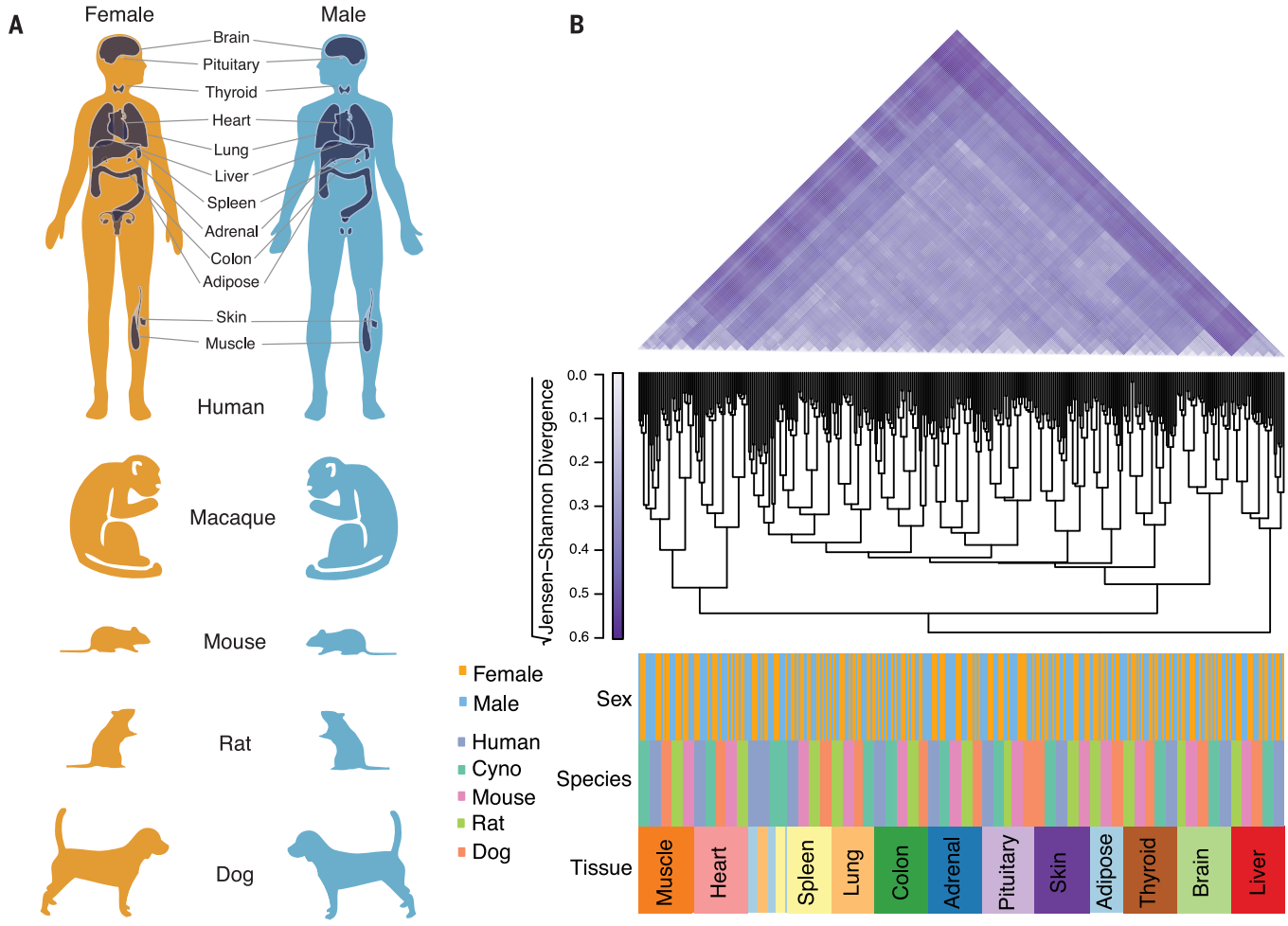



**Fig. 1. Five-species, 12-tissue survey of sex differences in gene expression.** (**A**) Schematic of study design, with tissues chosen for analysis in all five species highlighted in humans. (**B**) Hierarchical clustering of 349 RNA-seq samples. (Top) Pairwise estimates of Jensen-Shannon divergence (JSD) between pairs of samples. Six random human samples per tissue, in addition to all nonhuman samples, were included for display purposes. (Middle) Tree dendrogram obtained by hierarchical clustering (average linkage) based on pairwise JSD values. (Bottom) Sample labels by tissue, species, and sex.

With the exception of human adipose tissue and lung, which cluster closely together, samples cluster first by tissue and then by species (Fig. 1B). This tissue-dominated clustering agrees with prior studies (*37*, *38*) and indicates consistent sampling of tissues across species and also that the nonhuman data generated in this study are comparable to the human data from GTEx (*35*). There are no cases where samples cluster by sex before tissue or species, indicating that species effects dominate over sex effects. Nevertheless, sex contributes significantly to gene expression variation as pairwise within-sex distances in each tissue-species combination are significantly lower than pairwise between-sex distances (fig. S3).

Both our reanalysis of GTEx data and our analysis of our own data replicated published estimates of sex bias in six human and mouse tissues (*27*, *39*–*45*) (Pearson's correlation coefficient $r$ = 0.29 to 0.92) (figs. S4 and S5). These results indicate that expression values are comparable across species and yield reproducible estimates of sex bias.

### *Conserved sex-biased gene expression exists across the body*

Within each tissue, we used a linear mixed model to identify genes that showed a consistent sex bias [false discovery rate (FDR) 4.5%, as estimated by permutation of male and female sample labels] across species while controlling for differences in expression variability and sample size between species. Further, we required that genes show a fold change ≥ 1.05 in the same direction in at least four of the five species studied. We assume that such genes likely had a conserved sex bias in the common ancestor of Boreoeutheria (example in Fig. 2A). Of 113,853 expressed gene-tissue pairs, 3885 pairs (corresponding to 3161 genes) show a conserved sex bias. We used a rank-based statistic to confirm that gene-tissue pairs with conserved sex bias also have low $P$ values for sex bias in each of the individual species (fig. S6). Conserved sex bias is generally of modest magnitude (~90% of sex-biased gene-tissue pairs had a less than twofold change between the sexes) (fig. S7) but reproducible in independent datasets (Pearson's $r$ = 0.18 to 0.78) (fig. S8). The number of genes with conserved sex bias per tissue varies from 128 in colon to 805 in pituitary (Fig. 2B and table S4) and is not correlated with tissue sample size or rates of between-species gene expression divergence (Pearson's $r$ = 0.093 and 0.0083 and $P$ = 0.77 and 0.97, respectively) (fig. S9). A naïve approach, requiring $P < 0.05$ in at least four of five species for each tissue, found a smaller number of gene-tissue pairs with conserved sex bias but revealed between-tissue patterns that were correlated with results from the linear mixed model (fig. S10). Of genes with conserved sex bias in any of the 12 tissues examined, 562 genes (18%) are sex-biased in more than one tissue (Fig. 2C). In cases of multitissue sex bias, the bias is significantly more likely to be in the same direction in multiple tissues ($P$ = 0.00035, two-sided Fisher's exact test) (Fig. 2D). Thus, conserved sex bias in gene expression is mostly tissue-specific, but a significant minority of genes shows concordant sex bias across multiple tissues, implying that some regulatory factors result in similar profiles of sex-biased expression in multiple tissues or cell types.

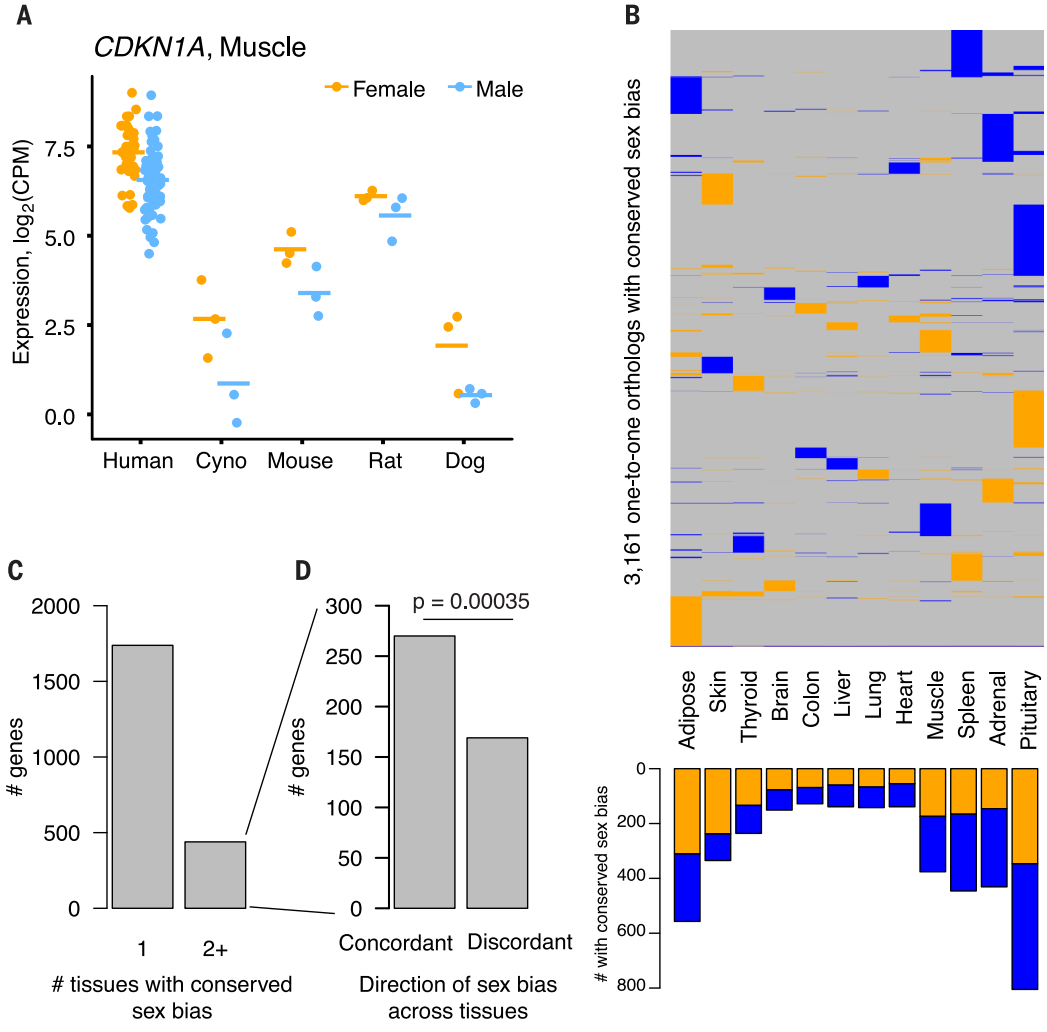



**Fig. 2. Conserved sex bias in gene expression across the body**. (**A**) Example of gene with conserved female-biased expression. CPM, counts per million. (**B**) Heatmap of conserved male (blue) and female (orange) sex bias across genes (rows) and tissues (columns). (**C**) The *y* axis represents number of genes with conserved sex bias in one (left) or multiple (right) tissues. (**D**) Of genes with conserved sex bias in multiple tissues, the number concordant (same direction) or discordant (opposite direction) in multiple tissues is plotted. Significance as assessed by two-sided Fisher's exact test comparing to equal proportions.

We considered the extent to which genes with conserved sex-biased expression were enriched for sex linkage. All assayed Y-linked genes are male-biased (fig. S11A), as expected, whereas X-linked genes are significantly enriched for conserved female bias (2.1- to 10.2-fold increase relative to autosomes two-sided Fisher's exact test) (fig. S11B). The enrichment for X-linked genes is driven by genes that escape X inactivation in females. In turn, the enrichment for X-escape genes is largely driven by the subset of X-escape genes that have a nonrecombining Y-linked homolog in mammals (two-sided Fisher's exact test) (fig. S11B). Despite these enrichments, most (85 to 95%, depending on the tissue) genes with conserved sex bias are autosomal (fig. S12). We compared the magnitude of sex bias between autosomal and X-linked genes using independent, publicly available datasets (*27*, *39*, *40*, *43*, *44*) (seven mouse, three human) to avoid ascertainment bias. X-linked genes show significantly higher magnitudes of sex bias in four of the ten datasets (adjusted $P < 0.05$, two-sided Wilcoxon rank-sum test) (fig. S13). Thus, the sex chromosomes, primarily as a result of harboring genes with both X- and Y-linked homologs, contribute a small but significant fraction of conserved sex bias in gene expression.

### *Most sex bias in gene expression has arisen since the last common ancestor of boreoeutherian mammals*

We investigated sex-biased gene expression specific to subsets of the five species, mindful that differences in statistical power between species could result in false positive calls of lineage-specific sex bias. For example, a gene with true primate-specific male bias might falsely appear to have a human-specific male bias if its expression is significantly biased in humans but does not reach statistical significance in cyno. At the same time, false positive calls of sex bias in single species will by necessity appear to be species specific. We used mashr (*46*) to model the covariation in sex bias across tissues and species and to more confidently determine the lineage of sex bias. We repeated the mashr procedure using permuted male/female sample labels to empirically estimate the FDR for any given set of sex-biased genes (*34*). This increased the number of rodent-specific gains of sex bias in most tissues (fig. S14). After using mashr to estimate sex bias in each tissue-species combination, we assigned each sex-biased gene-tissue pair (other than those with conserved sex bias) to one of 12 lineage-specific categories by parsimony: primate-specific gains or losses, rodent-specific gains or losses, gains specific to one of the five species, multiple gains or losses, and more complex patterns of sex bias inconsistent with single gains or losses (examples in Fig. 3A and table S4). In each category, we used the permutation-estimated FDR to estimate the

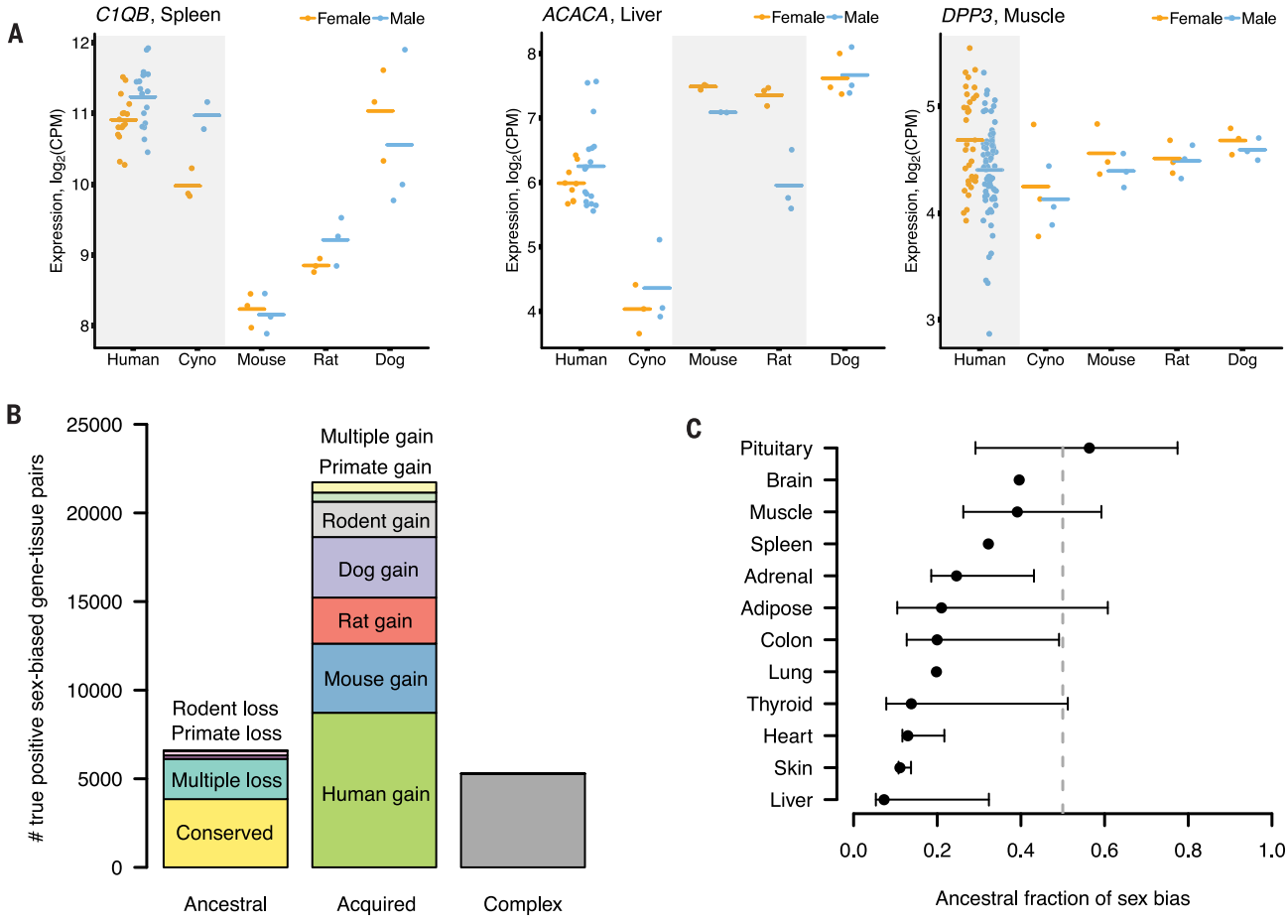



**Fig. 3. Most sex bias in gene expression has arisen since the last common ancestor of Boreoeutheria.** (**A**) Examples of genes with lineage-specific sex bias. (**B**) Number of true-positive sex-biased gene-tissue pairs (*y* axis) in each evolutionary class was calculated as the difference between the total number discovered across all tissues using true or permuted sex labels. Evolutionary classes defined in main text are designated as ancestral, acquired, or complex relative to last common ancestor of Boreoeutheria (the five species considered here). (**C**) Comparisons of ancestral to acquired sex biases as in (B), but performed in each tissue separately. Upper and lower confidence intervals represent fraction of sex bias estimated to be ancestral when counting all complex events as ancestral or acquired, respectively.

number of true positive sex-biased gene-tissue pairs.

We assessed how much of the sex-biased gene expression observed in the five species was present in the common ancestor of Boreoeutheria (i.e., ancestral). Instances of ancestrally sex-biased expression included gene-tissue pairs that we previously identified as having a "conserved" sex bias as well as gene-tissue pairs that lost sex bias in the primate or rodent lineages or in multiple lineages. Instances of acquired sex bias included gene-tissue pairs with primate-, rodent-, or species-specific sex bias, as well as multiple gains of sex bias. By this logic, 6539 (23%) of sex-biased gene-tissue pairs were likely sex-biased in the common ancestor, and 22,194 (77%) likely acquired sex bias after divergence from a common ancestor. An additional 8495 gene-tissue pairs exhibited more complex patterns and could not be confidently assigned as ancestrally sex-biased or acquired (Fig. 3B). If all such "complex" events were ancestral, the ancestral fraction of sex bias would be 40%, whereas if they were acquired, the fraction would be 18%. Performing these calculations in each tissue separately, we found that ancestral sex bias constituted the minority of total sex bias in all tissues except the pituitary (Fig. 3C). We also quantified the fraction of ancestral bias using a range of fold-change cutoffs up to 1.5 and found that, for all cases, ancestral sex bias was in the minority (fig. S15). Repeating this analysis with conserved sex bias called by mashr rather than the linear mixed model yielded similar results, with both methods detecting similar numbers of gene-tissue pairs with conserved sex bias (fig. S15). We conclude that most sex bias in gene expression in nonreproductive tissues arose during, rather than before, the boreoeutherian radiation.

### *Sex-biased gene expression is associated with reduced selective constraint*

We assessed the degree of selective constraint operating on sex-biased gene expression. Reasoning that genes functioning across many tissues and cell types face increased selective constraint on gene expression levels, we compared the breadth of expression of genes with and without sex bias, in each tissue. Sex-biased genes showed significantly lower expression breadth than genes with no bias, with the exception of lung, where sex-biased genes were more broadly expressed (adjusted $P < 0.05$, two-sided Wilcoxon rank-sum test) (Fig. 4A). These differences in expression breadth could either be downstream consequences of, or have predated, the observed sex bias. We thus analyzed expression breadth in chicken, an evolutionary outgroup to mammals, reasoning that patterns found in both human and chicken were likely present in the common mammalian ancestor before the acquisition of sex bias. Again, sex-biased genes in mammals showed almost uniformly lower expression breadth in chicken than unbiased genes (adjusted $P < 0.05$, two-sided Wilcoxon rank-sum test) (fig. S16).

To assess conservation of expression levels in a tissue-specific manner, we used estimates of

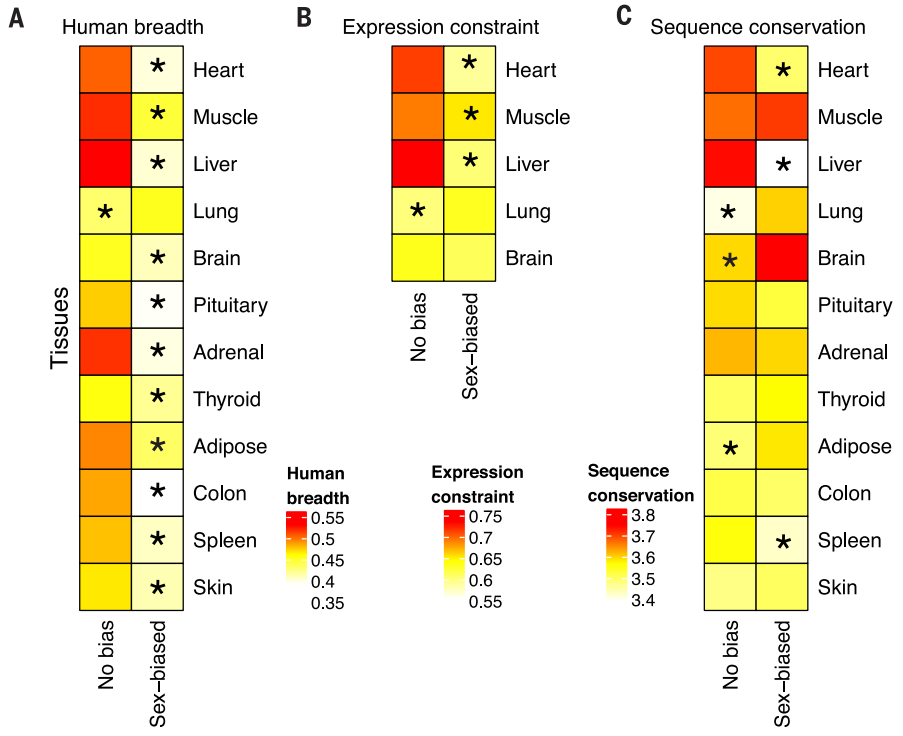



**Fig. 4. Sex-biased gene expression is associated with reduced selective constraint.** In each tissue, genes were binned as showing no sex bias or showing sex bias of any evolutionary type. Human breadth (**A**) was calculated on the basis of median expression values in the 12 selected GTEx tissues (*34*), expression constraint (**B**) represents the genome-wide percentile, and sequence conservation (**C**) is calculated as the mean coding phyloP score (*34*). In each heatmap, the group median of the indicated gene-level trait is plotted; asterisks indicate a Benjamini-Hochberg–adjusted $P < 0.05$ from a two-sided Wilcoxon rank-sum test, placed on the group ("No bias" or "Sex-biased") with the lower value of the gene-level trait.

mammalian gene expression-level constraint learned from 16 species (*47*) and seven tissues, five of which were also assessed in our study. As with expression breadth, sex-biased genes showed lower constraint than unbiased genes in heart, muscle, and liver but higher constraint in lung (adjusted $P < 0.05$, two-sided Wilcoxon rank-sum test) (Fig. 4B).

We observed that genes sex-biased in heart, spleen, and liver showed lower sequence conservation than unbiased genes, whereas genes sex-biased in adipose, brain, and lung showed higher sequence conservation than unbiased genes (adjusted $P < 0.05$, two-sided Wilcoxon rank-sum test) (Fig. 4C). Thus, some sex-biased genes are relatively strongly constrained at the sequence level, perhaps because they perform important or pleiotropic functions. However, considering both expression levels and sequence conservation, our findings indicate that sex-biased gene expression is primarily associated with reduced selective constraint, from before the divergence of the boreoeutherian lineages.

### *Conserved sex bias in autosomal gene expression contributes to sex differences in mammalian height and body size*

Males are larger than females in most mammalian taxa (*10*). Human males are, on average, 10 to 15 cm (7 to 13%) taller than females, but the distributions of height in males and females overlap substantially (Fig. 5A). The genetic architecture of human height is polygenic and largely shared between the sexes. A recent meta-analysis reported 712 genome-wide significant loci (*48*), and only a handful of sex-specific associations with height have been discovered (*49*, *50*); recent human studies reported a between-sex genetic correlation of 0.96 (*50*, *51*). Studies in humans and other mammals have concluded that height is likely subject to opposing selective pressures between the sexes, where increased height enhances reproductive success in males and decreased height favors reproductive success in females (*52*, *53*).

Could sex-biased gene expression contribute to the sex difference in height and body size observed in humans and other mammals? To link variation in gene expression to variation in height, we turned to transcriptome-wide association studies (TWAS) that integrate an expression quantitative trait loci (eQTL) study from a given tissue or cell type and genome-wide association studies (GWAS) of a given trait (*54*).

If sex-biased gene expression contributes to sex differences in height, genes with male-biased expression levels should mostly identify height-increasing effects, as measured by TWAS, whereas female-biased genes should identify height-decreasing effects. We considered genes with genome-wide significant associations for height, as annotated in the NHGRI-EBI GWAS catalog (*55*). We used TWAS to combine height GWAS statistics from a meta-analysis (*48*) of data from the UK Biobank (*56*) and GIANT consortium (*57*) (which we verified were correlated; Pearson's $r = 0.83$, $P < 2.2 \times 10^{-16}$) (fig. S17) with reference eQTL panels from 43 different tissues generated using data from GTEx (*35*). TWAS $z$-scores largely agree in sign across the 43 tissues (figs. S18 and S19; see table S5 for all TWAS $z$-scores). We therefore combined $z$-scores for each gene across tissues by meta-analysis. Sixty-two genome-wide significant height genes have both computed TWAS $z$-scores and conserved sex bias in at least one tissue. Genes with conserved male-biased expression have more-positive TWAS $z$-scores than genes with conserved female-biased expression (mean $z$-score difference = 18, $P = 0.023$, group permutation test), but this difference was not seen when analyzing genes with human-specific or primate-specific sex bias (Fig. 5B). Expanding our analyses to include TWAS results for all genes allowed for greater stringency by only considering TWAS $z$-scores calculated for the same tissue in which sex-biased expression was observed. Five hundred sixty gene-tissue pairs have both computed TWAS $z$-scores and conserved sex bias; these are distributed across all 12 tissues, with the largest numbers in muscle, adipose, and pituitary (fig. S20), and they are enriched for metabolic functions (adjusted $P$ value < 0.05, two-sided Fisher's exact test) (table S6). Gene-tissue pairs with conserved male bias have more-positive TWAS $z$-scores than those with conserved female bias (mean $z$-score difference = 0.7, $P = 0.039$, group permutation test), but this difference was not seen when considering gene-tissue pairs with human- or primate-specific sex bias (Fig. 5C). Together, these results indicate that genes with conserved male-biased expression show height-increasing effects, whereas genes with conserved female-biased expression show height-decreasing effects.

We sought to quantify the fraction of sex difference in height explained by conserved sex bias in gene expression, focusing on cases where the sex bias was in the same tissue as the TWAS $z$-score (i.e., Fig. 5C). We estimated the contribution of conserved sex bias to the height sex difference with two approaches. One approach used a physical scale with the effect sizes of eQTLs in GWAS, and the other examined a relative fold-change on the basis of TWAS $z$-scores (*34*) (fig. S21). The two approaches yielded similar estimates of the contribution of conserved sex-biased gene expression: ~1.6 cm, or 12%, of the observed sex difference in mean height.

Genes with conserved male and female bias show the largest difference in height TWAS $z$-scores, suggesting a contribution to sex differences in size in other mammals. Indeed, all five species assessed in this study exhibit sex differences in size (fig. S22). Consider the transcription factor *LCORL*, which shows conserved female

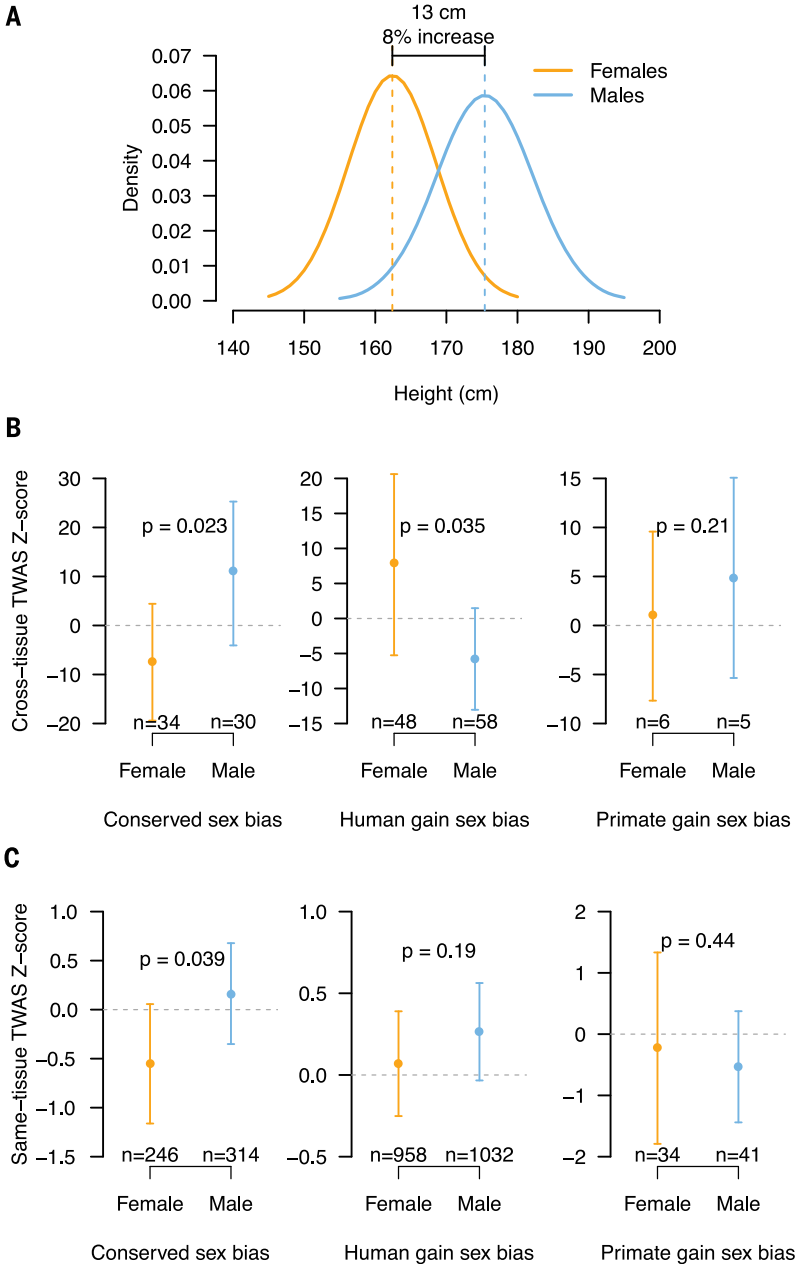



**Fig. 5. Conserved sex bias in autosomal gene expression contributes to sex differences in human height.** (**A**) Overlapping but shifted distributions of male and female heights. Theoretical normal distributions using published means and standard deviations of male and female heights in individuals of European ancestry from the United Kingdom (*53*). (**B**) TWAS *z*-scores for genome-wide significant height genes with either female (orange) or male (blue) bias in one of 12 tissues, either conserved across mammals (left), specific to humans (middle), or specific to primates (right). For each gene, TWAS *z*-scores were meta-analyzed across 48 GTEx tissues. (**C**) TWAS *z*-scores for gene-tissue pairs with either female (orange) or male (blue) bias, either conserved across mammals (left), specific to humans (middle), or specific to primates (right), in all cases in same tissue as computed TWAS *z*-score. Points represent group means; whiskers represent 95% confidence intervals. *P* value for mean difference calculated by 1000 permutations of male and female point labels.

bias in the pituitary (fig. S23A) and is height-decreasing in humans (cross-tissue TWAS *z*-score = −28.7). Although *LCORL* lacks a predictive TWAS model in the pituitary, an allele at the *LCORL* locus associated with increased expression in the pituitary is associated with decreased height (fig. S23B). Notably, genetic variation at the *LCORL* locus has been associated with height or body size in dogs (*58*), cattle (*59*), and horses (*60*). Reanalysis of publicly available RNA-seq data (*61*) shows that *LCORL* is one of the most strongly female-biased autosomal genes in the cattle pituitary (1.6-fold higher in females) (fig. S23C and table S7). An allele associated with increased body size in horses is associated with decreased *LCORL* expression in hair root (*62*), indicating that the negative association between *LCORL* expression and height likely extends beyond humans. These observations suggest that female-biased expression of *LCORL* contributes to sex differences in size in multiple species. Beyond *LCORL*, studies have observed significant overlap in genome-wide-significant height loci between humans (*57*), dogs (*58*), and cattle (*59*) (table S8), suggesting a broader contribution of conserved sex-biased gene expression to sex differences in body size in a range of mammals.

### *Conserved and acquired sex biases in gene expression are similarly enriched for specific biological pathways and show similar magnitudes of bias*

Our results indicate that although sex-biased gene expression overall shows signs of lowered selective constraint, conserved sex bias contributes to sex differences in height or body size. This raises the possibility that more-recently acquired sex bias in expression has little or no functional impact. To assess this possibility, we compared genes with either conserved or acquired sex bias with respect to (i) overrepresentation in specific biological pathways via Gene Ontology (GO) category enrichment and (ii) the magnitude of their sex bias.

We observed similar degrees of enrichments for biological pathways among conserved and acquired sex biases. Genes with conserved male bias in pituitary are enriched for cyclic adenosine monophosphate signaling, which functions in response to stress (*63*). Genes with conserved female bias in colon and thyroid are enriched for adaptive immune pathways, whereas genes with conserved female bias in adipose tissue are enriched for mitochondrial translation and ribosomal RNA processing. At the same time, genes with acquired male bias in the liver, adipose tissue, and heart are enriched for functions related to fatty acid metabolism, regulation of hormone secretion, and nucleotide metabolism, respectively, and genes with acquired female bias in the liver are enriched for extracellular matrix organization (adjusted $P < 0.05$, two-sided Fisher's exact test) (table S6).

We compared the magnitude of conserved and acquired sex bias using 10 independent human and mouse datasets to minimize differences due to ascertainment; we found no significant differences (adjusted $P > 0.05$, two-sided Wilcoxon rank-sum test) (fig. S24). Considering that conserved sex bias, although generally small in magnitude, can nevertheless contribute to sex differences in height, these results suggest that acquired sex bias could also be functionally consequential. Additional studies are needed to demonstrate the functional impact of acquired sex-biased gene expression.

### Evolutionary turnover of motifs for sex-biased transcription factors reflects lineage-specific changes in sex bias

One mechanism by which sex-biased expression could evolve is via sex-biased transcription factors (TFs). For example, male-biased TF expression in muscle would result in higher TF activity in male muscle. Genes that acquired motifs for this TF in, for example, the primate lineage would then show a primate-specific sex bias in muscle. To test this idea, we searched for motifs enriched in the promoters of sex-biased genes with lineage-specific changes in a given tissue, relative to their unbiased orthologs. We repeated this analysis with random, equally sized sets of genes showing no lineage-specific sex bias in order to calculate an empirical *P* value for motif enrichment (*34*). Because of the nonparametric nature of this *P* value calculation and our desire to analyze enriched motifs inclusively as a set, we considered motifs at a 10% FDR. We found 83 instances in which such motifs matched predicted binding sites of TFs with sex bias in the same tissue (Fig. 6A and table S9). This was significantly more ($P$ = 0.014, tissue permutation test) than the ~67 instances of matches expected when randomly assigning the tissue of sex bias for each TF (Fig. 6B), which, combined with the 10% FDR for motif discovery, yields ~6.7 matches expected by chance. By quantifying the enrichment of each motif in its corresponding set of sex-biased orthologs (*34*), we estimated that these 83 instances account for the lineage-specific sex bias of 6073 gene-tissue pairs, or 27% of all lineage-specific sex bias. Furthermore, 13 TFs showed matches to enriched motifs in more than one tissue, significantly more than the approximately one such TF expected by chance ($P$ = 0.032, tissue permutation test), as determined by the motif discovery FDR and permuting the tissue of TF sex bias as described above (fig. S25). This suggests that gains and losses of motifs for sex-biased TFs could, in some cases, coordinate the evolution of sex-biased gene expression across multiple tissues or cell types.

To confirm that genes with lineage-specific gains or losses of motifs for sex-biased TFs are TF-bound in living cells, we leveraged publicly available data from chromatin immunoprecipitation sequencing (ChIP-seq) in human and mouse (table S9). Although these assays were almost invariably performed in a different cell type than the tissue of TF sex bias and motif gain, we reasoned that sex-biased genes with gained motifs in a given tissue should nevertheless show enrichment for TF ChIP-seq signal. Eleven of 15 cases with available data showed significant enrichment of ChIP-seq peaks in the promoters of genes with a gain or loss of sex bias and the relevant motif, relative to a background set of

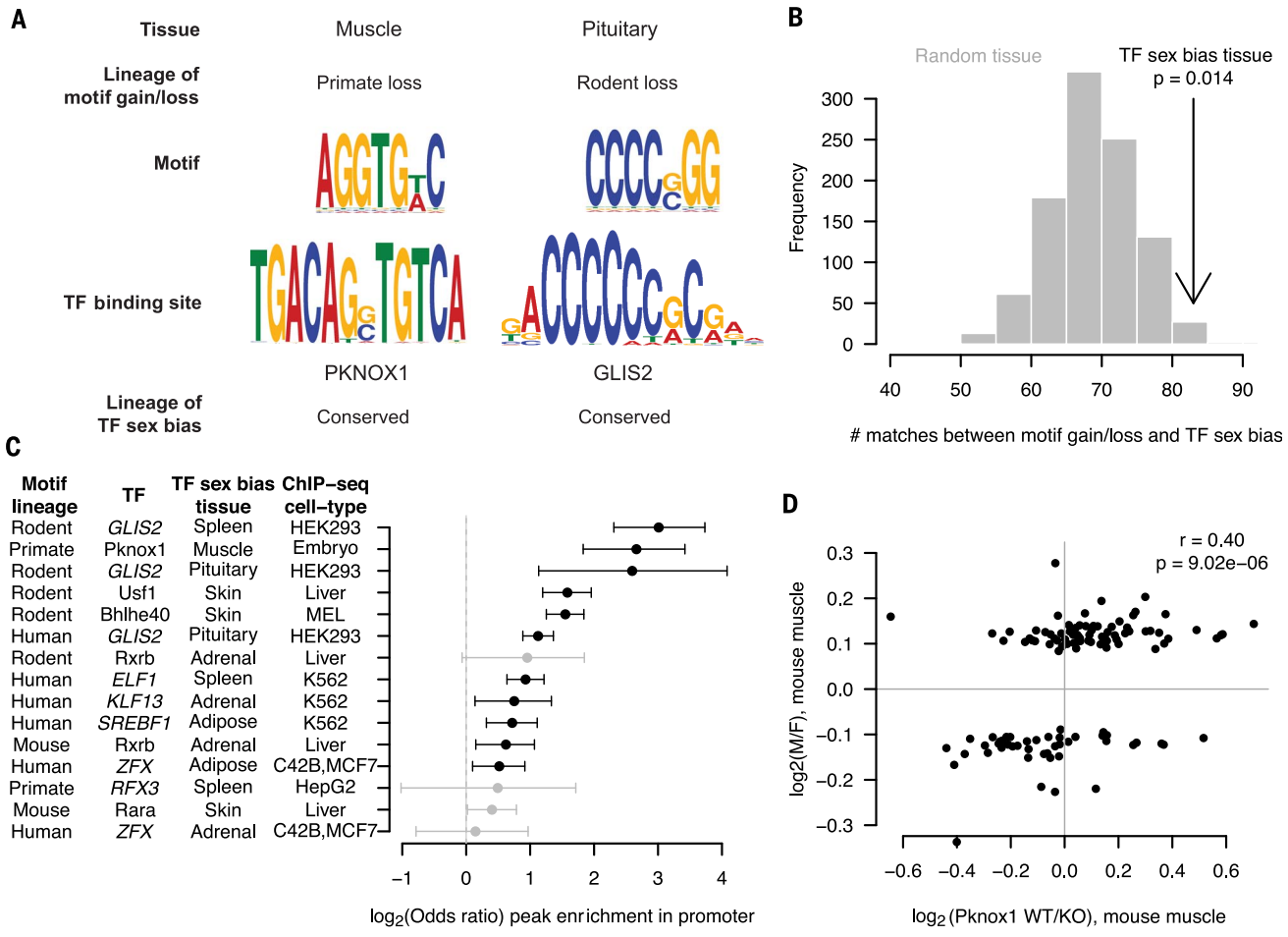



**Fig. 6. Evolutionary turnover of motifs for sex-biased transcription factors is associated with gains and losses of sex bias.** (**A**) Representative gained or lost motifs in promoters of genes with lineage-specific gains or losses of sex bias (top) aligned with motifs for sex-biased TFs in same tissue (bottom). The lineage of sex bias gain or loss is indicated above each motif; the sex-biased TF and lineage of its sex bias are indicated below. (**B**) Total number of matches between gained or lost motifs and sex-biased TFs when considering tissue of TF sex bias (black) or randomly chosen tissues (gray). (**C**) Enrichment of ChIP-seq peaks in promoters of genes with lineage-specific sex biases containing gained or lost motifs for the TF. The sex-biased TF, along with tissue of sex bias and motif gain or loss and cell type in which ChIP-seq was performed, are indicated to left. The $\log_2$ odds ratio for genes with lineage-specific sex bias and containing the motif as compared with a background set of genes with no motif is shown on the *x* axis, with 95% confidence intervals by Fisher's exact test. (**D**) Effect of *Pknox1* knockout (*x* axis) (*64*) versus sex bias (*y* axis), both in mouse muscle, for genes that show loss of sex bias in primate lineage and contain a motif for PKNOX1 in mouse.

genes lacking the motif (two-sided Fisher's exact test), which is consistent with this prediction (Fig. 6C). Thus, the evolutionary gains and losses of motifs we observed likely correspond to gains and losses of binding by cognate TFs.

If gains and losses of motifs for sex-biased TFs contribute to lineage-specific changes in sex bias in their target genes, there should be directional agreement between the activating or repressive effect of the TF, the sex bias of the TF, and the sex bias of the target gene. For example, target genes activated (or repressed) by a male-biased TF should be male (or female) biased, and the opposite should be true for female-biased TFs. Rigorously testing this prediction requires experimental manipulation of the TF in the tissue where lineage-specific changes in sex bias are observed. Such data are available for PKNOX1, a homeobox TF with conserved male-biased expression in muscle (*64*) (fig. S26). Genes with loss of muscle-specific sex bias in the primate lineage show depletion (at a stringent 5% FDR) of PKNOX1-matching motifs relative to mouse, rat, and dog (Fig. 5A, examples of PKNOX1 targets in fig. S26). Genes with a PKNOX1-matching motif show significant positive correlation between the effect of *Pknox1* knockout (*64*) and the effect of sex on muscle gene expression (Pearson's $r = 0.40$, $P = 9.02 \times 10^{-6}$) (Fig. 6D). Thus, both ChIP-seq and TF knockout data confirm that gains and losses of regulation by sex-biased TFs have contributed to the evolution of sex bias.

## Discussion

Comparative studies of sex-biased gene expression have implications for the use of nonhuman mammals as models of sex-biased human traits or diseases. Conserved sex bias in gene expression across the body indicates that certain molecular sex differences in humans are amenable to study in a wide range of mammalian model organisms. However, in many cases, nonhuman models may not adequately recreate the human sex differences in question. This is supported by two lines of evidence: (i) in each tissue, samples cluster by species rather than sex, and (ii) most sex bias in gene expression has arisen recently and is thus not shared between most mammals. For example, genetic variants that decrease expression of the TF *KLF14* in human adipose tissue tend to increase insulin resistance and risk for type 2 diabetes only in females, but elimination of *Klf14* expression in mouse adipose tissue leads to analogous phenotypes in both sexes (*65*). Nonhuman mammals may still be useful as models of physiological or systems-level sex differences, but caution should be exercised when extrapolating specific molecular findings to humans.

We find that conserved sex bias in autosomal gene expression explains ~12% of the sex difference in mean human height, whereas all common single-nucleotide polymorphisms are thought to explain 60% of the heritability of height (*66*). Although these two numbers are not directly comparable (the former relates to between-group differences, the latter to between-individual variation), these height heritability estimates suggest that additional genes and instances of sex bias relevant to sex differences in height remain to be discovered. Deletions of the *SHOX* gene, located in the pseudoautosomal region of the human X and Y chromosomes, contribute to short stature in Turner syndrome (*67*), whereas increases in sex chromosome number (and thus *SHOX* dosage) increase height (*68*). Although the height GWAS used here excluded the pseudoautosomal regions, precluding analysis of *SHOX*, targeted studies indicate that *SHOX* dosage is positively correlated with height (*67*, *68*). In light of reports that expression of *SHOX* is male-biased in multiple tissues (*16*), it may be that *SHOX* contributes a fraction of the sex difference in height [discussed further in (*69*)].

Studies of selection on height have illustrated how males and females can have different optimal values for a quantitative trait, with increased height favored in males and decreased height favored in females (*53*). Our finding that conserved sex bias in gene expression contributes to sex differences in height suggests one way in which such optimal values can be reached—through the acquisition and maintenance of sex-biased gene expression (*70*). Thus, although some conserved sex bias in gene expression may have arisen through selectively neutral processes, opposing selective forces between the sexes appear to have been at work here. Height is also subject to balancing selection, in which extreme variation in either direction negatively impacts reproductive fitness (*53*). A recent study in *Drosophila* found a strong signature of balancing selection at loci with opposite fitness effects in females and males, establishing that sexual antagonism and balancing selection can coincide (*71*). Future studies may identify mechanisms that reduce fitness at the extremes of the height distributions in both sexes. Whereas our study focused exclusively on height, genetic pleiotropy may broaden the reach of our findings. Sex-biased gene expression resulting from opposing selective pressures on male and female height could result in sex differences in phenotypes yet to be identified.

Our results also illustrate one way in which sex-biased gene expression can lead to phenotypic sex differences: autosomal genes, operating identically in males and females to influence a trait, can be expressed more abundantly in one sex. Although most genetic variation influencing height acts identically between the sexes, pronounced sex-specific genetic effects have been demonstrated for waist-to-hip ratio, body mass index (*72*–*74*), thyroid hormone levels (*75*), and obsessive-compulsive disorder (*76*). Fully accounting for such sex differences in genetic architecture in association mapping (*77*), and integrating this information with sex biases in gene expression, may reveal additional mechanisms underlying phenotypic sex differences.

Our finding of sex-biased TFs underlying lineage-specific changes in sex bias provides molecular insight into mechanisms underlying the evolution of sex-biased gene expression in nonreproductive mammalian tissues. We focused on regulatory changes in promoter regions because of the lack of tissue-specific enhancer annotations in cyno, rat, or dog. However, single-gene studies in *Drosophila* indicate that sex-biased gene expression can evolve through more complex changes in cis-regulatory elements at larger genomic distances from their target gene (*31*). Studying gains and losses of TF binding motifs in promoters, although an important first step, is a simplifying approach. It is thus necessary to catalog both tissue and species specificity of mammalian enhancers to enable detailed analyses of the cis-regulatory changes driving gains or losses of sex-biased gene expression during mammalian evolution. Most of the sex-biased TFs we identified as contributing to lineage-specific evolution of sex bias are autosomal, indicating that their sex biases could arise as a result of trans-regulatory effects of sex chromosomes or sex hormones. Distinguishing between these two possibilities is an important future direction for research.

## Materials and methods summary

Human (GTEx) samples were filtered on the basis of cause of death, medical history, and notes from GTEx pathologists (*35*), and additional detailed evaluations were conducted on samples with available histology. Samples from cynomolgus macaque, mouse, rat, and dog were collected within 1 hour of euthanizing healthy animals, and only tissues from nonestrous females were used. RNA extraction, library preparation, and RNA-seq of nonhuman mammals were performed in batches randomized with respect to tissue, species, and sex. Analyzing the combined human and nonhuman dataset, we used a linear mixed model (*78*) to identify genes showing consistent sex bias across species in each tissue, and we used mashr to identify lineage-specific changes in sex bias. These analyses were repeated with permuted male/female sample labels to empirically estimate FDRs. Magnitudes of sex bias were assessed in independent datasets by reanalyzing raw data, where available. For lineage-specific sex biases in each tissue, motif analysis (*79*, *80*) was used to identify TF binding sites enriched in the set of sex-biased orthologs relative to the unbiased orthologs. Height-increasing or -decreasing effects of gene-tissue pairs were determined by combining publicly available TWAS predictive models (*54*) based on eQTL information from GTEx with height GWAS summary statistics from the UK Biobank (*56*), GIANT consortium (*57*), and a meta-analysis of the two studies (*48*).

### REFERENCES AND NOTES


1. J. C. K. Wells, Sexual dimorphism of body composition. *Best Pract. Res. Clin. Endocrinol. Metab.* **21**, 415–430 (2007). doi: 10.1016/j.beem.2007.04.007; pmid: 17875489
2. H. J. Green, I. G. Fraser, D. A. Ranney, Male and female differences in enzyme activities of energy metabolism in vastus lateralis muscle. *J. Neurol. Sci.* **65**, 323–331 (1984). doi: 10.1016/0022-510X(84)90095-9; pmid: 6238135
3. A. N. V. Ruigrok *et al.*, A meta-analysis of sex differences in human brain structure. *Neurosci. Biobehav. Rev.* **39**, 34–50 (2014). doi: 10.1016/j.neubiorev.2013.12.004; pmid: 24374381

4. S. L. Klein, K. L. Flanagan, Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016). doi: 10.1038/nri.2016.90; pmid: 27546235
5. C. S. Hayward, W. V. Kalnins, R. P. Kelly, Gender-related differences in left ventricular chamber function. *Cardiovasc. Res.* **49**, 340–350 (2001). doi: 10.1016/S0008-6363(00)00280-7; pmid: 11164844
6. S. T. Ngo, F. J. Steyn, P. A. McCombe, Gender differences in autoimmune disease. *Front. Neuroendocrinol.* **35**, 347–369 (2014). doi: 10.1016/j.yfrne.2014.04.004; pmid: 24793874
7. V. Regitz-Zagrosek *et al*., Gender in cardiovascular diseases: Impact on clinical manifestations, management, and outcomes. *Eur. Heart J.* **37**, 24–34 (2016). doi: 10.1093/eurheartj/ehv598; pmid: 26530104
8. D. M. Werling, D. H. Geschwind, Sex differences in autism spectrum disorders. *Curr. Opin. Neurol.* **26**, 146–153 (2013). doi: 10.1097/WCO.0b013e32835ee548; pmid: 23406909
9. H. Olson *et al*., Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.* **32**, 56–67 (2000). doi: 10.1006/rtph.2000.1399; pmid: 11029269
10. P. Lindenfors, J. L. Gittleman, K. Jones, in *Evolutionary Studies of Sexual Size Dimorphism* (Oxford Univ. Press, 2007), pp. 16–26.
11. R. A. Gorski, J. H. Gordon, J. E. Shryne, A. M. Southam, Evidence for a morphological sex difference within the medial preoptic area of the rat brain. *Brain Res.* **148**, 333–346 (1978). doi: 10.1016/0006-8993(78)90723-0; pmid: 656937
12. R. S. Scotland, M. J. Stables, S. Madalli, P. Watson, D. W. Gilroy, Sex differences in resident immune cell phenotype underlie more efficient acute inflammatory responses in female mice. *Blood* **118**, 5918–5927 (2011). doi: 10.1182/blood-2011-03-340281; pmid: 21911834
13. K. M. Shioura, D. L. Geenen, P. H. Goldspink, Sex-related changes in cardiac function following myocardial infarction in mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **295**, R528–R534 (2008). doi: 10.1152/ajpregu.90342.2008; pmid: 18550865
14. H. Skaletsky *et al*., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003). doi: 10.1038/nature01722; pmid: 12815422
15. D. W. Bellott *et al*., Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014). doi: 10.1038/nature13206; pmid: 24759411
16. T. Tukiainen *et al*., Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017). doi: 10.1038/nature24265; pmid: 29022598
17. M. Melé *et al*., The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: 10.1126/science.aaa0355; pmid: 25954002
18. M. Gershoni, S. Pietrokovski, The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol.* **15**, 7 (2017). doi: 10.1186/s12915-017-0352-z; pmid: 28173793
19. X. Yang *et al*., Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006). doi: 10.1101/gr.5217506; pmid: 16825664
20. J. M. Ranz, C. I. Castillo-Davis, C. D. Meiklejohn, D. L. Hartl, Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745 (2003). doi: 10.1126/science.1085881; pmid: 12805547
21. Y. Zhang, D. Sturgill, M. Parisi, S. Kumar, B. Oliver, Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* **450**, 233–237 (2007). doi: 10.1038/nature06323; pmid: 17994089
22. S. Grath, J. Parsch, Rate of amino acid substitution is influenced by the degree and conservation of male-biased transcription over 50 myr of *Drosophila* evolution. *Genome Biol. Evol.* **4**, 346–359 (2012). doi: 10.1093/gbe/evs012; pmid: 22321769
23. R. Assis, Q. Zhou, D. Bachtrog, Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol. Evol.* **4**, 1189–1200 (2012). doi: 10.1093/gbe/evs093; pmid: 23097318
24. J. C. Perry, P. W. Harrison, J. E. Mank, The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Mol. Biol. Evol.* **31**, 1206–1219 (2014). doi: 10.1093/molbev/msu072; pmid: 24526011
25. B. Reinius *et al*., An evolutionarily conserved sexual signature in the primate brain. *PLOS Genet.* **4**, e1000100 (2008). doi: 10.1371/journal.pgen.1000100; pmid: 18566661
26. R. Blekhman, J. C. Marioni, P. Zumbo, M. Stephens, Y. Gilad, Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010). doi: 10.1101/gr.099226.109; pmid: 20009012
27. Y. Liang *et al*., A gene network regulated by the transcription factor VGLL3 as a promoter of sex-biased autoimmune diseases. *Nat. Immunol.* **18**, 152–160 (2017). doi: 10.1038/ni.3643; pmid: 27992404
28. A. E. Russi, M. E. Ebel, Y. Yang, M. A. Brown, Male-specific IL-33 expression regulates sex-dimorphic EAE susceptibility. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1520–E1529 (2018). doi: 10.1073/pnas.1710401115; pmid: 29378942
29. M. Souyris *et al*., *TLR7* escapes X chromosome inactivation in immune cells. *Sci. Immunol.* **3**, eaap8855 (2018). doi: 10.1126/sciimmunol.aap8855; pmid: 29374079
30. E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017). doi: 10.1016/j.cell.2017.05.038; pmid: 28622505
31. T. M. Williams *et al*., The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**, 610–623 (2008). doi: 10.1016/j.cell.2008.06.052; pmid: 18724934
32. C. M. Disteche, Dosage compensation of the sex chromosomes. *Annu. Rev. Genet.* **46**, 537–560 (2012). doi: 10.1146/annurev-genet-110711-155454; pmid: 22974302
33. C. M. Disteche, Dosage compensation of the sex chromosomes and autosomes. *Semin. Cell Dev. Biol.* **56**, 9–18 (2016). doi: 10.1016/j.semcdb.2016.04.013; pmid: 27112542
34. Materials and methods are available as supplementary materials.
35. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
36. A. M. Newman *et al*., Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015). doi: 10.1038/nmeth.3337; pmid: 25822800
37. J. Merkin, C. Russell, P. Chen, C. B. Burge, Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012). doi: 10.1126/science.1228186; pmid: 23258891
38. D. Brawand *et al*., The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011). doi: 10.1038/nature10532; pmid: 22012392
39. D. M. Werling, N. N. Parikshak, D. H. Geschwind, Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat. Commun.* **7**, 10717 (2016). doi: 10.1038/ncomms10717; pmid: 26892004
40. M. E. Lindholm *et al*., The human skeletal muscle transcriptome: Sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. *FASEB J.* **28**, 4571–4581 (2014). doi: 10.1096/fj.14-255000; pmid: 25016029
41. M. S. Newman, T. Nguyen, M. J. Watson, R. W. Hull, H.-G. Yu, Transcriptome profiling reveals novel BMI- and sex-specific gene expression signatures for human cardiac hypertrophy. *Physiol. Genomics* **49**, 355–367 (2017). doi: 10.1152/physiolgenomics.00122.2016; pmid: 28500252
42. N. Viguerie *et al*., Determinants of human adipose tissue gene expression: Impact of diet, sex, metabolic status, and cis genetic regulation. *PLOS Genet.* **8**, e1002959 (2012). doi: 10.1371/journal.pgen.1002959; pmid: 23028366
43. R. Marin *et al*., Convergent origination of a *Drosophila*-like dosage compensation mechanism in a reptile lineage. *Genome Res.* **27**, 1974–1987 (2017). doi: 10.1101/gr.223727.117; pmid: 29133310
44. B. Li *et al*., A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. *Sci. Rep.* **7**, 4200 (2017). doi: 10.1038/s41598-017-04520-z; pmid: 28646208
45. M. D. Franco *et al*., Transcriptome of normal lung distinguishes mouse lines with different susceptibility to inflammation and to lung tumorigenesis. *Cancer Lett.* **294**, 187–194 (2010). doi: 10.1016/j.canlet.2010.01.038; pmid: 20189714
46. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019). doi: 10.1038/s41588-018-0268-8; pmid: 30478440
47. J. Chen *et al*., A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res.* **29**, 53–63 (2019). doi: 10.1101/gr.237636.118; pmid: 30552105
48. L. Yengo *et al*., Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018). doi: 10.1093/hmg/ddy271; pmid: 30124842
49. T. Tukiainen *et al*., Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage compensation. *PLOS Genet.* **10**, e1004127 (2014). doi: 10.1371/journal.pgen.1004127; pmid: 24516404
50. J. Sidorenko, I. Kassam, K. Kemper, J. Zeng, L. Lloyd-Jones, G. W. Montgomery, G. Gibson, A. Metspalu, T. Esko, J. Yang, A. F. McRae, P. M. Visscher, The effect of X-linked dosage compensation on complex trait variation. bioRxiv 433870 [Preprint]. 3 October 2018. https://doi.org/10.1101/433870
51. K. Rawlik, O. Canela-Xandri, A. Tenesa, Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol.* **17**, 166 (2016). doi: 10.1186/s13059-016-1025-x; pmid: 27473438
52. G. Stulp, L. Barrett, Evolutionary perspectives on human height variation. *Biol. Rev. Camb. Philos. Soc.* **91**, 206–234 (2016). doi: 10.1111/brv.12165; pmid: 25530478
53. J. S. Sanjak, J. Sidorenko, M. R. Robinson, K. R. Thornton, P. M. Visscher, Evidence of directional and stabilizing selection in contemporary humans. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 151–156 (2018). doi: 10.1073/pnas.1707227114; pmid: 29255044
54. A. Gusev *et al*., Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016). doi: 10.1038/ng.3506; pmid: 26854917
55. J. MacArthur *et al*., The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45** (D1), D896–D901 (2017). doi: 10.1093/nar/gkw1133; pmid: 27899670
56. P. R. Loh, G. Kichaev, S. Gazal, A. P. Schoech, A. L. Price, Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018). doi: 10.1038/s41588-018-0144-6; pmid: 29892013
57. A. R. Wood *et al*., Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014). doi: 10.1038/ng.3097; pmid: 25282103
58. J. J. Hayward *et al*., Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* **7**, 10460 (2016). doi: 10.1038/ncomms10460; pmid: 26795439
59. A. C. Bouwman *et al*., Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* **50**, 362–367 (2018). doi: 10.1038/s41588-018-0056-5; pmid: 29459679
60. H. Signer-Hasler *et al*., A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLOS ONE* **7**, e37282 (2012). doi: 10.1371/journal.pone.0037282; pmid: 22615965
61. M. Seo *et al*., Comprehensive identification of sexually dimorphic genes in diverse cattle tissues using RNA-seq. *BMC Genomics* **17**, 81 (2016). doi: 10.1186/s12864-016-2400-4; pmid: 26818975
62. J. Metzger, R. Schrimpf, U. Philipp, O. Distl, Expression levels of LCORL are associated with body size in horses. *PLOS ONE* **8**, e56497 (2013). doi: 10.1371/journal.pone.0056497; pmid: 23418579
63. N. Stroth, Y. Holighaus, D. Ait-Ali, L. E. Eiden, PACAP: A master regulator of neuroendocrine stress circuits and the cellular stress response. *Ann. N. Y. Acad. Sci.* **1220**, 49–59 (2011). doi: 10.1111/j.1749-6632.2011.05904.x; pmid: 21388403
64. T. Kanzleiter *et al*., Pknox1/Prep1 regulates mitochondrial oxidative phosphorylation components in skeletal muscle. *Mol. Cell. Biol.* **34**, 290–298 (2014). doi: 10.1128/MCB.01232-13; pmid: 24216763
65. K. S. Small *et al*., Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018). doi: 10.1038/s41588-018-0088-x; pmid: 29632379
66. J. Yang *et al*., Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010). doi: 10.1038/ng.608; pmid: 20562875
67. E. Rao *et al*., Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat. Genet.* **16**, 54–63 (1997). doi: 10.1038/ng0597-54; pmid: 9140395
68. T. Ogata, N. Matsuo, Sex chromosome aberrations and stature: Deduction of the principal factors involved in the determination of adult height. *Hum. Genet.* **91**, 551–562 (1993). doi: 10.1007/BF00205079; pmid: 8340109
69. A. K. San Roman, D. C. Page, A strategic research alliance: Turner syndrome and sex differences. *Am. J. Med. Genet. C. Semin. Med. Genet.* **181**, 59–67 (2019). doi: 10.1002/ajmg.c.31677; pmid: 30790449

70. J. Parsch, H. Ellegren, The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* **14**, 83–87 (2013). doi: 10.1038/nrg3376; pmid: 23329110
71. F. Ruzicka *et al*., Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. *PLOS Biol.* **17**, e3000244 (2019). doi: 10.1371/journal.pbio.3000244; pmid: 31022179
72. J. C. Randall *et al*., Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLOS Genet.* **9**, e1003500 (2013). doi: 10.1371/journal.pgen.1003500; pmid: 23754948
73. T. W. Winkler *et al*., The influence of age and sex on genetic associations with adult body size and shape: A large-scale genome-wide interaction study. *PLOS Genet.* **11**, e1005378 (2015). doi: 10.1371/journal.pgen.1005378; pmid: 26426971
74. D. Shungin *et al*., New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015). doi: 10.1038/nature14132; pmid: 25673412
75. E. Porcu *et al*., A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLOS Genet.* **9**, e1003266 (2013). doi: 10.1371/journal.pgen.1003266; pmid: 23408906
76. E. A. Khramtsova *et al*., Sex differences in the genetic architecture of obsessive-compulsive disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 10.1002/ajmg.b.32687 (2018). doi: 10.1002/ajmg.b.32687; pmid: 30456828
77. E. Y. Kang *et al*., An Association Mapping Framework To Account for Potential Sex Difference in Genetic Architectures. *Genetics* **209**, 685–698 (2018). pmid: 29752291
78. M. E. Ritchie *et al*., *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015). doi: 10.1093/nar/gkv007; pmid: 25605792
79. T. L. Bailey, DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (2011). doi: 10.1093/bioinformatics/btr261; pmid: 21543442
80. R. C. McLeay, T. L. Bailey, Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010). doi: 10.1186/1471-2105-11-165; pmid: 20356413
81. S. Naqvi, A. Godfrey, J. Hughes, M. Goodheart, R. Mitchell, D. Page, Conservation, acquisition, and functional impact of sex-biased gene expression in mammalian tissues [Data set], Version 1, Zenodo (2019); http://doi.org/10.5281/zenodo.2658829.

## ACKNOWLEDGMENTS

We thank D. W. Bellott, L. Chmatal, and R. Ransohoff for critically reading the manuscript. **Funding:** S.N. and A.K.G. were supported by a research grant from Biogen, Inc. This work was supported by Biogen, Whitehead Institute, National Institutes of Health (grants R01HG007852 and U01HG007857), Howard Hughes Medical Institute, generous gifts from Brit and Alexander d'Arbeloff and Arthur W. and Carol Tobin Brill. **Author contributions:** S.N., A.K.G., J.F.H., and D.C.P. designed the study. J.F.H. procured cyno tissue samples. M.L.G. procured mouse and rat tissue samples, with assistance from S.N. S.N. processed tissue samples and performed computational analyses, with assistance from A.K.G. R.N.M. performed histological evaluations on human tissue sections. D.C.P. supervised work. S.N. and D.C.P. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw RNA-seq data (.fastq files) have been deposited in the Gene Expression Omnibus (GSE125483). Processed data (expression count and TPM matrices and sample metadata) and code required to reproduce the analyses are available at http://pagelab.wi.mit.edu/page/papers/Naqvi_et_al_2019 and at Zenodo (*81*).

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/eaaw7317/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S9
References (*82*–*106*)

19 January 2019; accepted 12 June 2019
10.1126/science.aaw7317

RESEARCH ARTICLE SUMMARY

ALZHEIMER'S DISEASE

# Amyloid β oligomers constrict human capillaries in Alzheimer's disease via signaling to pericytes

**Ross Nortley, Nils Korte*, Pablo Izquierdo*, Chanawee Hirunpattarasilp*, Anusha Mishra*, Zane Jaunmuktane*, Vasiliki Kyrargyri*, Thomas Pfeiffer, Lila Khennouf, Christian Madry, Hui Gong, Angela Richard-Loendt, Wenhui Huang, Takashi Saito, Takaomi C. Saido, Sebastian Brandner, Huma Sethi, David Attwell†**

**INTRODUCTION:** In Alzheimer's disease (AD), the production of amyloid β (Aβ) oligomers and downstream tau dysfunction are thought to cause neuronal damage, in particular a loss of synapses and synaptic plasticity, which results in cognitive impairment. However, epidemiological data show that vascular factors are important contributors to AD risk, and biomarker research has shown that the first change in AD is a decrease of cerebral blood flow. Because most of the vascular resistance within the brain is located in capillaries, this could reflect a dysfunction of contractile pericytes on capillary walls. Indeed, pericytes are known to regulate cerebral blood flow physiologically and to severely restrict blood flow after stroke.

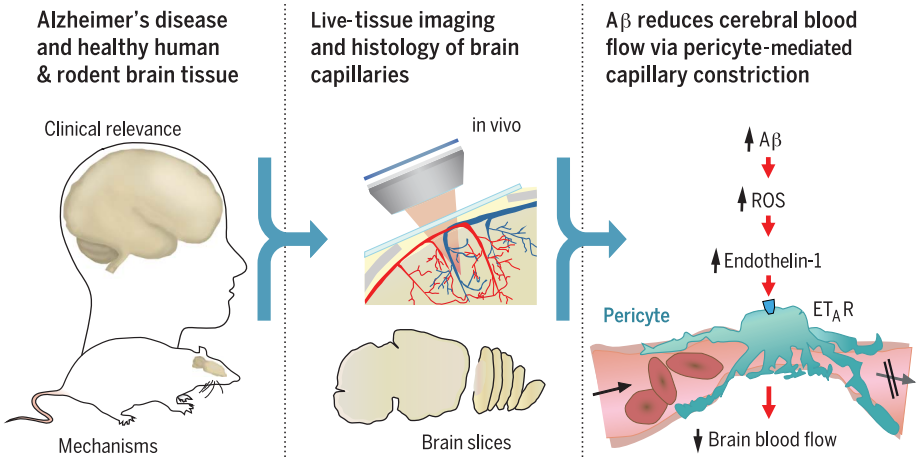
**RATIONALE:** We examined the role of pericytes in Alzheimer's disease by examining cerebral capillaries in humans and mice developing AD, and by applying Aβ to capillaries. We used freshly fixed brain biopsies from cognitively impaired living humans who were depositing Aβ plaques, and also carried out in vivo imaging in a knock-in mouse model of AD. We measured capillary diameters at positions near pericytes in order to assess whether the capillaries became constricted in AD, because this would lead to a decrease of cerebral blood flow and hence a decrease of the glucose and oxygen supply to the brain tissue. In addition, to investigate one mediator already thought to be important in AD, we applied Aβ to human brain slices made from normal tissue that was removed from patients undergoing neurosurgical glioma resection, as well as to rodent brain slices. Aβ was applied in the oligomeric form, which is thought to contribute to cognitive decline. This allowed us to examine whether Aβ itself might alter cerebral blood flow, and to use pharmacology to investigate the mechanism of any such effect.

**RESULTS:** Both in humans developing AD and in the mouse model of AD, capillaries were constricted specifically at pericyte locations, but arterioles and venules were unchanged in diameter. Thus, the reduction of cerebral blood flow known to occur in AD is produced by capillaries rather than by arterioles. The capillary constriction increased rapidly with the severity of Aβ deposition, and we calculated that in the human cortex this constriction would have the effect of reducing cerebral blood flow by approximately half; this is comparable to the decrease of blood flow measured experimentally in affected parts of the AD brain. In the AD mouse cerebellum, which lacks Aβ deposition at the age examined, there was no capillary constriction, supporting the idea of a causal link between Aβ level and constriction of capillaries. Aβ itself was found to constrict both human and rodent capillaries through a mechanism involving the generation of reactive oxygen species (ROS), mainly by NOX4 (reduced nicotinamide adenine dinucleotide phosphate oxidase 4). The ROS then triggered the release of endothelin-1, which acted on $ET_A$ receptors to evoke pericyte contraction, thus causing capillary constriction. The Aβ-evoked constriction could be halted by blocking NOX4 and $ET_A$ receptors, and was reversed by applying the vasodilator C-type natriuretic peptide.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aav9518

**CONCLUSION:** These data reconcile genetic evidence for a role of Aβ in triggering neuronal damage and cognitive decline in AD with the fact that a decrease of cerebral blood flow is the first clinically detectable change in AD. They imply that attention should be given to vascular mechanisms in AD as well as to signaling pathways that act directly on neurons or glia, and suggest novel therapeutic approaches for treating early AD by targeting drugs to brain pericytes. Our findings also raise the question of what fraction of the damage to synapses and neurons in AD reflects direct actions of Aβ and downstream tau, and what fraction is a consequence of the decrease of energy supply that Aβ produces by constricting capillaries. ■




**Live human and rodent brain capillaries become constricted in Alzheimer's disease.** Tissue from humans and rodents (left) that were healthy or developing Alzheimer's disease (AD) was imaged in vivo and as brain slices (center), revealing that pericytes constrict brain capillaries early in AD via a mechanism involving ROS generation and release of endothelin-1, which activates $ET_A$ receptors (right).

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: d.attwell@ucl.ac.uk
Cite this article as R. Nortley *et al.*, *Science* **365**, eaav9518 (2019). DOI: 10.1126/science.aav9518

RESEARCH ARTICLE

ALZHEIMER'S DISEASE

# Amyloid β oligomers constrict human capillaries in Alzheimer's disease via signaling to pericytes

**Ross Nortley[1], Nils Korte[1]*, Pablo Izquierdo[1]*, Chanawee Hirunpattarasilp[1]*, Anusha Mishra[2]*, Zane Jaunmuktane[3,4]*, Vasiliki Kyrargyri[1]*†, Thomas Pfeiffer[1], Lila Khennouf[1], Christian Madry[1]‡, Hui Gong[1], Angela Richard-Loendt[3], Wenhui Huang[5], Takashi Saito[6], Takaomi C. Saido[6], Sebastian Brandner[3,7], Huma Sethi[8], David Attwell[1]§**

**Cerebral blood flow is reduced early in the onset of Alzheimer's disease (AD). Because most of the vascular resistance within the brain is in capillaries, this could reflect dysfunction of contractile pericytes on capillary walls. We used live and rapidly fixed biopsied human tissue to establish disease relevance, and rodent experiments to define mechanism. We found that in humans with cognitive decline, amyloid β (Aβ) constricts brain capillaries at pericyte locations. This was caused by Aβ generating reactive oxygen species, which evoked the release of endothelin-1 (ET) that activated pericyte $ET_A$ receptors. Capillary, but not arteriole, constriction also occurred in vivo in a mouse model of AD. Thus, inhibiting the capillary constriction caused by Aβ could potentially reduce energy lack and neurodegeneration in AD.**

Vascular compromise occurs early in Alzheimer's disease (AD) (*1*, *2*). Cerebral blood flow in the gray matter can be reduced by more than 40% (*3*); indeed, reduced cerebral blood flow is the earliest biomarker of the development of the disease (*1*). Around the time that the amyloid hypothesis for AD was proposed (*4*–*6*), it was reported that capillaries in the brains of AD patients showed an abnormal, focally constricted morphology (*7*). Although much succeeding work focused on amyloid β (Aβ)– and tau-evoked damage to neurons, increasing evidence suggests a role for vascular disturbance in the onset of AD (*8*, *9*). Exogenous Aβ can reduce cerebral blood flow (*10*–*12*), and reduced blood flow increases Aβ production (*13*, *14*).

[1]Department of Neuroscience, Physiology and Pharmacology, University College London, London WC1E 6BT, UK. [2]Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR 97239, USA. [3]Division of Neuropathology, National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK. [4]Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, Queen Square, London WC1N 3BG, UK. [5]Molecular Physiology, CIPMM, University of Saarland, D-66421 Homburg, Germany. [6]Laboratory for Proteolytic Neuroscience, RIKEN Centre for Brain Science, Wako, Saitama 351-0198, Japan. [7]Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, Queen Square, London WC1N 3BG, UK. [8]Division of Neurosurgery, UCL Queen Square Institute of Neurology, Queen Square, London WC1N 3BG, UK.
*These authors contributed equally to this work. †Present address: Laboratory of Molecular Genetics, Department of Immunology, Hellenic Pasteur Institute, 115 21 Athens, Greece. ‡Present address: Institute of Neurophysiology, Charité-Universitätsmedizin, 10117 Berlin, Germany.
§Corresponding author. Email: d.attwell@ucl.ac.uk

Investigations of the vascular effects of exogenous Aβ have focused on arteries and arterioles (*12*, *15*), but the majority of the vascular resistance within the brain is located in capillaries (*16*). Capillary dysfunction correlates with cognitive decline in AD (*17*). This suggests that capillaries could be the most important locus where Aβ produced within the brain can act to decrease cerebral blood flow. A subset of pericytes on capillary walls is contractile [these are the only contractile cells on capillaries (*18*)] and can alter cerebral blood flow by adjusting their contractile tone (*18*–*20*). In a rodent model of AD, there are disturbances of unknown origin in the control of capillary blood flow (*21*). We therefore investigated how pericytes were affected by exogenous and endogenously generated Aβ, and in particular by $Aβ_{1–42}$ oligomers, the molecular species believed to be responsible for Aβ's toxic effects in AD (*22*, *23*). To maximize the relevance to human disease, we used living human brain slices derived from neurosurgically resected brain tissue to study acute responses to Aβ, and rapidly fixed human brain biopsy tissue (from living patients with or without Aβ deposition) to assess pericyte responses to long-term accumulation of endogenous Aβ in AD. The effects seen in human tissue were also seen in vivo in a transgenic mouse model of AD and were analyzed mechanistically in brain slices.

## Amyloid β constricts human capillaries at pericytes

Living human brain cortex slices were obtained from tissue removed during neurosurgical operations to access tumors (see materials and methods) and either fixed for immunohistochemistry or imaged live to study pericyte properties. Labeling the basement membrane with fluorescently tagged isolectin $B_4$ ($IB_4$), or immunolabeling for the pericyte marker PDGFRβ (platelet-derived growth factor receptor β), revealed pericyte morphology. Pericytes were observed with a classical "bump-on-a-log" morphology on the straight parts of capillaries, or at their branch points, with processes extending along and around the capillaries (Fig. 1, A and B). With experience, morphology alone was sufficient to identify pericytes reliably in brain slices (fig. S1). The mean distance between human pericytes was 65.3 ± 0.4 μm (for 94 pericytes imaged in tissue from two patients), 30% larger than in rodents (*19*). As for arteriole smooth muscle cells (Fig. 1C), the processes of 36% of pericytes could be labeled for α smooth-muscle actin (Fig. 1D) [the real percentage may be higher with different fixation techniques (*24*)], providing a mechanistic basis for the Aβ-evoked contraction (see below). In human brain slices, as previously reported for rodent capillaries (*18*, *19*), superfused noradrenaline constricted and glutamate dilated the capillaries at pericyte locations (Fig. 1, E and F). This is consistent with the circumferential processes of pericytes (which are oriented to be able to reduce capillary diameter) being preferentially found near pericyte somata (fig. S2), so that capillary constriction by pericytes occurs predominantly near these somata (fig. S3). Thus, the surgery-derived human tissue had functioning contractile pericytes (Fig. 1, E and F).

Aβ was oligomerized (see materials and methods), and silver staining of SDS–polyacrylamide gel electrophoresis (PAGE) gels was used to assess the degree of aggregation of the Aβ isoforms. The predominant species produced (other than monomers) for $Aβ_{1–42}$ and $Aβ_{1–40}$ had a molecular weight 2 to 4 times that of monomers, whereas scrambled $Aβ_{1–42}$ formed mainly monomers (Fig. 1G). Applying soluble $Aβ_{1–42}$ (oligomeric + monomeric, 72 nM calculated from the monomeric molecular weight) to human brain slices evoked a slowly developing constriction of all four capillaries tested, which reduced their diameter by ~25% after 40 min (Fig. 1H, significantly reduced, $P = 0.01$).

Because the limited availability of live human tissue precluded detailed analysis of the mechanism underlying the Aβ-evoked constriction, we carried out experiments on rat cortical slices. As for human capillaries, $Aβ_{1–42}$ evoked a constriction of rat capillaries near pericyte locations that was visible using either bright-field illumination or two-photon fluorescence imaging of $IB_4$ (Fig. 2, A, B, C, and G). Of 20 capillaries tested, 16 (80%) showed a >5% constriction in response to $Aβ_{1–42}$. The time course of the mean $Aβ_{1–42}$-evoked constriction (including all vessels) was similar (Fig. 2C) to that in human cortex (Fig. 1H), reaching ~15% after 1 hour ($P = 0.006$). $Aβ_{1–40}$ also evoked a similar constriction (Fig. 2, C and G; $P = 0.048$) in five of six capillaries tested (83%). Capillaries monitored for an hour without applying Aβ, or those to which a version of $Aβ_{1–42}$

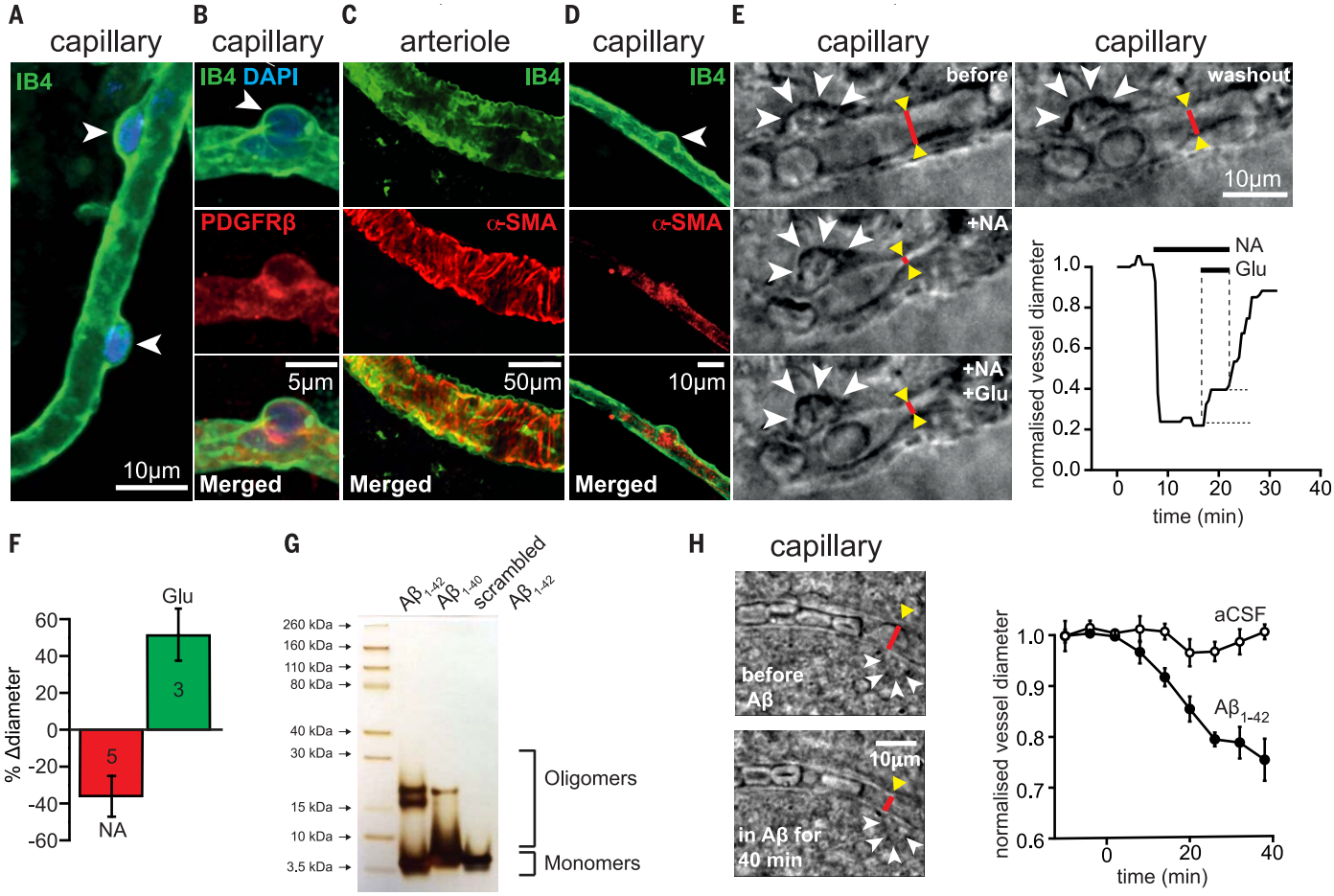



**Fig. 1. Oligomeric Aβ acts on pericytes to constrict capillaries in human brain slices.** (**A**) $IB_4$-labeled capillary in a human cortical slice, with two pericyte somata (white arrowheads) outlined by their basement membrane. Nuclei are stained with DAPI (blue). (**B**) Pericyte labeled with antibody to PDGFRβ. (**C** and **D**) Arteriole (C) and pericyte (D) labeled with $IB_4$ and antibody to α smooth muscle actin (α-SMA, localized in processes originating from the pericyte soma). (**E**) Images of a capillary (red lines between yellow arrowheads indicate diameter) and pericyte soma (white arrowheads) in a live human brain slice before drug application, in the presence of 2 µM superfused noradrenaline (+NA), with 2 µM NA and 500 µM glutamate superfused (+NA +Glu), and after stopping drug superfusion (washout). Graph shows time course of capillary diameter at red line throughout the experiment. (**F**) Mean (± SEM) glutamate-evoked dilation and noradrenaline-evoked constriction in experiments as in (E) (numbers of pericytes on bars; change in diameter was quantified relative to that before application of each drug; relative to the pre-noradrenaline diameter, the glutamate-evoked dilation was 26.8 ± 7.7%). (**G**) Silver staining of an SDS-PAGE gel for Aβ solutions prepared as in materials and methods. (**H**) Images of a human capillary before and after superfusion of 72 nM $Aβ_{1–42}$, showing a region (red line) being constricted by a pericyte (arrowheads). Graph shows mean (±SEM) diameter change at four pericyte locations from four slices treated with Aβ and three pericyte locations from three slices superfused with aCSF lacking Aβ (significantly reduced at 40 min in Aβ, $P = 0.01$).

with a scrambled sequence was applied (prepared as for the Aβ oligomers), showed no significant diameter change (Fig. 2, C and G). Scrambled $Aβ_{1–42}$ mainly formed monomers (Fig. 1G; see also materials and methods), unlike $Aβ_{1–42}$ and $Aβ_{1–40}$, which may indicate that oligomer formation is obligatory for an effect on pericytes. The pericyte-mediated constriction evoked by $Aβ_{1–42}$ showed a Michaelis-Menten dependence on Aβ concentration, with an apparent $EC_{50}$ (the concentration for a half-maximal response, equal to the Michaelis constant $K_m$ obtained by fitting a Michaelis-Menten relation to the data) of 4.7 nM (Fig. 2D).

## Reactive oxygen species and endothelin-1 generate the capillary constriction

We blocked $Aβ_{1–42}$-evoked capillary constriction in rat cortical slices by means of the endothelin-1 (ET) type A receptor blocker BQ-123 (1 µM, $P = 0.008$; Fig. 2, E and G); by application of superoxide dismutase 1 (SOD1, 150 units/ml; $P = 3.7 \times 10^{-6}$; Fig. 2, E and G), which scavenges reactive superoxide generated when Aβ activates reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (and prevents hydroxyl radical formation by the Fenton reaction); or by inhibiting NADPH oxidase with diphenyleneiodonium (DPI, 10 µM, $P = 0.032$; Fig. 2, F and G). In contrast, applying these agents alone did not affect capillary diameter (changes after 1 hour: BQ-123, −0.7 ± 5.2%, $n = 13$, $P = 0.9$; SOD1, 3.4 ± 5.8%, $n = 9$, $P = 0.57$; DPI, −7.4 ± 3.2%, $n = 5$, $P = 0.082$). The $Aβ_{1–40}$-evoked capillary constriction was also abolished by BQ-123 (a 6.2 ± 1.6% dilation was seen after 1 hour of $Aβ_{1–40}$ applied in BQ-123, $n = 5$), contradicting the suggestion (*25*) that $Aβ_{1–40}$ does not evoke ET release. These results suggest the involvement of NADPH oxidase–mediated reactive oxygen species (ROS) generation and ET release in the Aβ-evoked capillary constriction. Reactive nitrogen species derived from superoxide were not involved, because inhibiting nitric oxide synthase (NOS) with 100 µM $N_ω$-nitro-L-arginine (L-NNA) had no effect ($P = 0.83$) on the Aβ-evoked constriction (Fig. 2, F and G); L-NNA alone had no effect (after 1 hour, the diameter change was −0.5 ± 7.9%, $n = 6$, $P = 0.99$). The fact that Aβ evoked constrictions even in the presence of L-NNA also rules out the possibility that Aβ-evoked ROS production caused constriction (*26*) by ROS binding to, and removing, vasodilatory NO. The NOX4 (NADPH oxidase 4) blocker GKT137831 (0.45 µM) abolished the Aβ-evoked capillary constriction (Fig. 2, F and G;

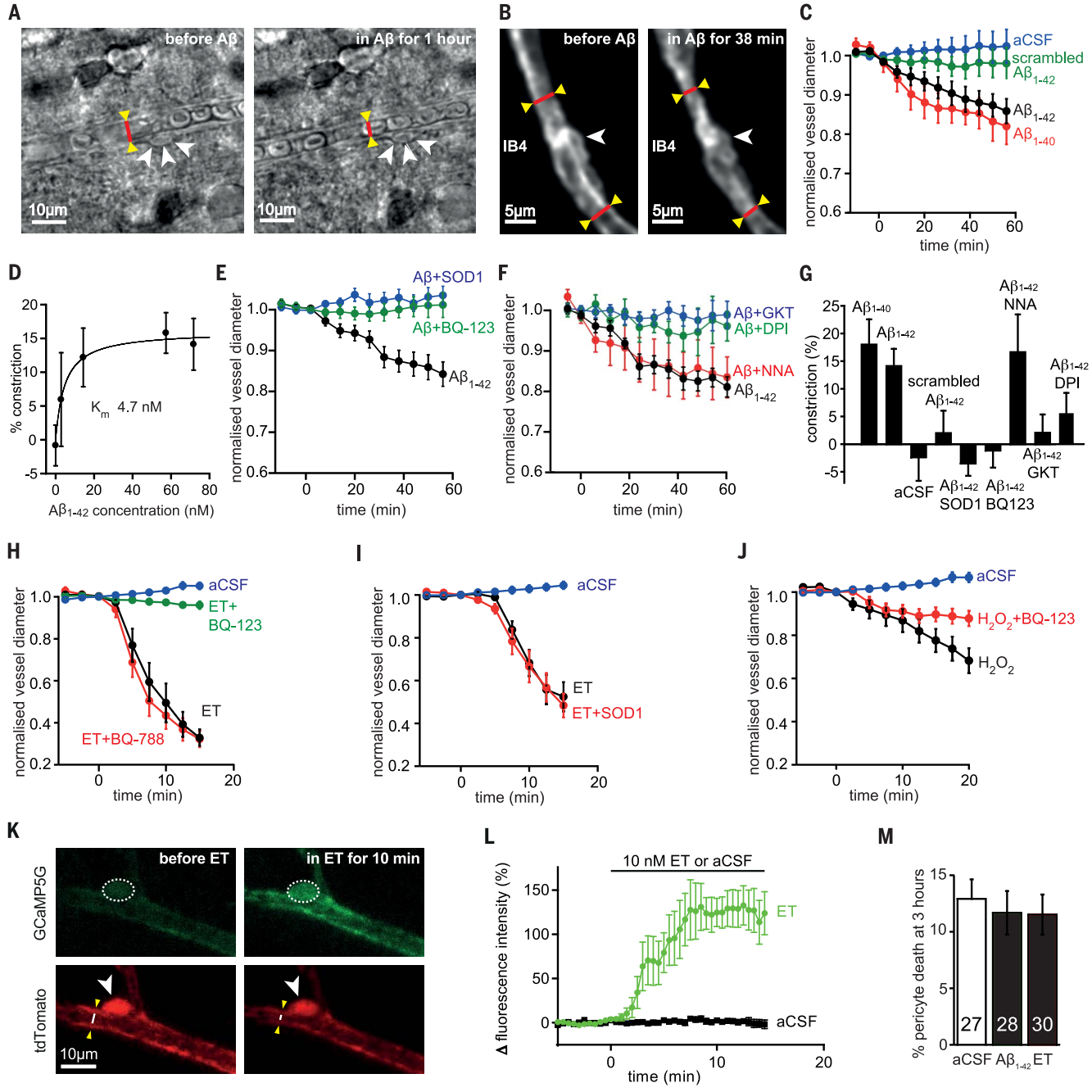



**Fig. 2. Aβ acts via ROS and $ET_A$ receptors.** (**A** and **B**) Bright-field images (A) and two-photon–evoked $IB_4$ fluorescence (B) of capillaries in rat cortical slices in aCSF and after applying 72 nM $Aβ_{1–42}$, showing constriction (yellow arrowheads and red lines) near pericytes (white arrowheads, compare with figs. S2 and S3). (**C**) Mean (± SEM) time course of capillary diameter during superfusion with aCSF ($n$ = 51 vessels), 109 nM scrambled $Aβ_{1–42}$ ($n$ = 32), 72 nM $Aβ_{1–42}$ ($n$ = 20), or 100 nM $Aβ_{1–40}$ ($n$ = 6). (**D**) Constriction evoked after 1 hour by different concentrations of $Aβ_{1–42}$ (0 nM, $n$ = 51; 2.9 nM, $n$ = 11; 14 nM, $n$ = 10; 57 nM, $n$ = 19; 72 nM, $n$ = 20). Curve is a Michaelis-Menten relation with a $K_m$ of 4.7 nM and a maximum of 16.1%. (**E** to **J**) Time course of diameter when applying the following agents (experiments in each panel were interleaved; blockers were present for 5 to 15 min before Aβ). (E) 57 nM $Aβ_{1–42}$ alone ($n$ = 19) or in the presence of SOD1 (150 units/ml, $n$ = 19) or the $ET_A$ blocker BQ-123 (1 μM, $n$ = 14). (F) 72 nM $Aβ_{1–42}$ alone ($n$ = 7) or in the presence of the NOS blocker L-NNA (100 μM, $n$ = 6), the NADPH oxidase blocker DPI (10 μM, $n$ = 5), or the NOX4 blocker GKT137831 (0.45 μM, $n$ = 7). (G) Constriction produced at 60 min for (C) to (F). (H) Effect of aCSF ($n$ = 10), ET alone (10 nM, $n$ = 10), or ET in the presence of the $ET_A$ blocker BQ-123 (1 μM, $n$ = 10) or the $ET_B$ blocker BQ-788 (1 μM, $n$ = 12). (I) aCSF or ET (5 nM) in the absence ($n$ = 12) or presence of SOD1 (150 units/ml, $n$ = 8). (J) aCSF or the ROS generator $H_2O_2$ (1 mM, $n$ = 9, which evokes constriction: $P = 1.1 \times 10^{-5}$ at 20 min) or $H_2O_2$ with the $ET_A$ blocker BQ-123 (1 μM, $n$ = 11, constriction is reduced, $P$ = 0.009). (**K**) Two-photon image of mouse cortical pericyte expressing GCaMP5G (green), before and while applying ET (10 nM), which raises $[Ca^{2+}]_i$ (increase in green intensity) in pericyte soma (arrowhead; dashed line shows ROI analyzed) and processes, and constricts the capillary (see white line on image of the tdTomato reporter of GCaMP5G expression, red). (**L**) Mean $[Ca^{2+}]_i$ time course in eight pericyte somata in response to ET (significantly elevated, $P$ = 0.0014) and in seven somata in aCSF (no significant change, $P$ = 0.74). (**M**) Incubating rat brain slices (numbers on bars) with $Aβ_{1–42}$ oligomers (1.4 μM) or ET (100 nM) for 3 hours does not increase pericyte death.

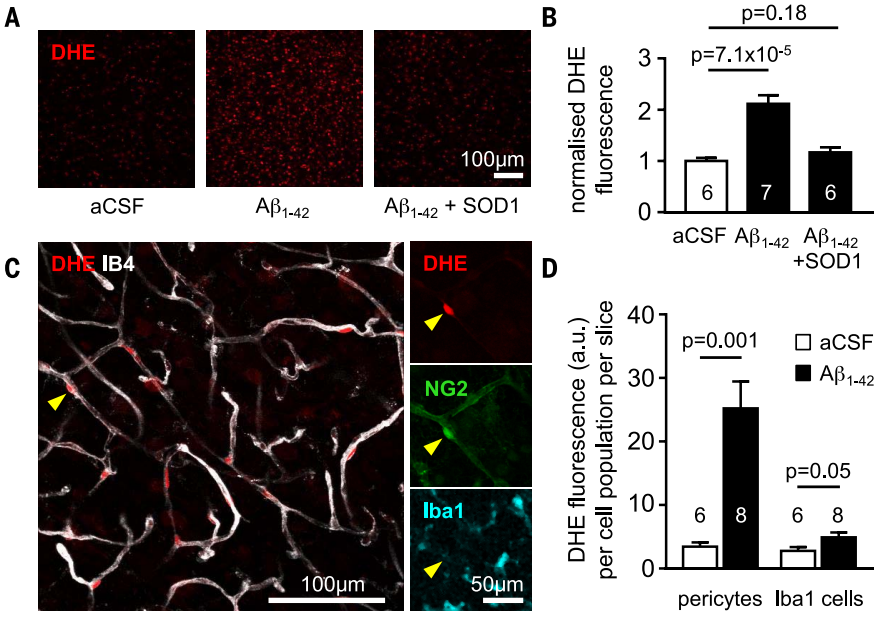



**Fig. 3. Aβ evokes ROS generation in pericytes.** (**A**) Fluorescence images of dihydroethidium (DHE)–loaded rat cortical slices incubated in control aCSF or aCSF containing $A\beta_{1-42}$ (72 nM) or $A\beta_{1-42}$ + SOD1 (150 units/ml) for 40 min, showing that Aβ increases ROS level and that this is inhibited by SOD1. (**B**) Fluorescence (normalized to value in aCSF, mean ± SEM) of slices incubated in aCSF ($n$ = 6), $A\beta_{1-42}$ ($n$ = 7), or $A\beta_{1-42}$ + SOD1 ($n$ = 6). (**C**) Left: Image of a cortical slice showing that the brightest DHE-labeled cells are located on $IB_4$-labeled blood vessels (arrowhead). Right: Immunolabeling shows that these cells colocalize with NG2 but not with Iba1, implying that they are pericytes rather than microglia or perivascular macrophages. (**D**) Soma DHE fluorescence [arbitrary units (a.u.), mean ± SEM] from the population of pericytes, or of Iba1-labeled cells, after 40 min in the absence or presence of $A\beta_{1-42}$. Numbers on bars are slices (fluorescence was averaged across three image stacks for each slice).

$P$ = 0.0011) but did not affect diameter when applied alone (changed by 5.7 ± 5.6%, $n$ = 6, $P$ = 0.35 after 1 hour), whereas the NOX2 blocker ebselen (2 μM) reduced the constriction by only 45% ($n$ = 8, $P$ = 0.027); on its own, ebselen had no effect (diameter changed by 1.4% ± 3.8%, $n$ = 9, $P$ = 0.8 after 1 hour). These data suggest that NOX4 in pericytes or endothelial cells (*27*–*29*), rather than NOX2 in immune cells (*28*, *29*), is the NADPH oxidase mainly responsible for generating the ROS that evoke capillary constriction. Data presented in Fig. 3 suggest that the NOX4 producing the ROS is in pericytes.

To confirm that pericytes constrict in response to activation of ET receptors, we applied ET (10 nM) either alone or with a blocker of its type A ($ET_A$) or type B ($ET_B$) receptors. Endothelin-1 evoked a strong (>65%) pericyte-mediated constriction of capillaries ($P = 2 \times 10^{-12}$), which was blocked by the $ET_A$ blocker BQ-123 (1 μM, $P = 2.6 \times 10^{-11}$) but not by the $ET_B$ blocker BQ-788 (1 μM, $P$ = 0.91; Fig. 2H). ET still evoked a constriction in the presence of SOD1 ($P = 1.3 \times 10^{-8}$; Fig. 2I), implying that ET acts downstream of ROS. Use of $H_2O_2$ (1 mM) to generate ROS evoked a constriction ($P = 1.1 \times 10^{-5}$) that was reduced by BQ-123 ($P$ = 0.009; Fig. 2J), which suggests that ROS evoke constriction via $ET_A$ receptor activation. Consistent with the idea that $ET_A$ receptors that generate pericyte contraction are on the pericytes themselves, we found that in pericytes expressing GCaMP5G (see materials and methods), applying ET (10 nM) evoked a rise in intracellular calcium concentration $[Ca^{2+}]_i$, whereas artificial cerebrospinal fluid (aCSF) had no effect (Fig. 2, K and L). These data establish $A\beta_{1-42}$-evoked generation of ROS as being upstream of the elevated level (*30*, *31*) [or potentiated effect (*32*)] of ET that makes pericytes constrict capillaries.

In profound ischemia, pericyte-evoked constriction of capillaries is followed by the pericytes dying necrotically in rigor (caused by an excessive rise of $[Ca^{2+}]_i$), thus maintaining a decreased capillary diameter and a long-lasting decrease of blood flow (*19*). Pericytes also die after accumulating Aβ in AD (*33*). We assessed whether exposure to 1.4 μM soluble $A\beta_{1-42}$ or 100 nM ET for 3 hours had a similar effect on pericyte health by applying propidium iodide to label cells with membranes that had become nonspecifically permeable, as occurs in ischemia (*19*). These treatments did not significantly increase pericyte death on this time scale (Fig. 2M; $P$ = 0.85 for $A\beta_{1-42}$, $P$ = 0.59 for ET).

To assess which cell types generated ROS in response to Aβ, in brain slices we used imaging of the ROS sensor dihydroethidium, which generates fluorescence when oxidized dihydroethidium intercalates into DNA (see materials and methods). $A\beta_{1-42}$ (72 nM, applied for 40 min) evoked an increase in ROS level that was suppressed by the presence of SOD1 (Fig. 3, A and B). Previous work has suggested that ROS can be generated in response to Aβ by resident microglia (*34*) or perivascular macrophages (*35*), but the cells showing the brightest oxidized dihydroethidium fluorescence were located on capillaries, had the morphology of pericytes, and could be labeled for the proteoglycan NG2 (found on pericytes) but not for the immune cell marker Iba1 (ionized calcium-binding adaptor molecule 1) (Fig. 3C), implying that they are pericytes. The ROS signal generated in regions of interest placed over the nuclei of NG2-expressing cells on capillaries (pericytes), or of Iba1-labeled immune cells, was quantified in six image stacks (one stack per slice) from slices not exposed to Aβ (containing a total of 128 pericytes and 238 Iba1-labeled cells) and eight stacks from slices exposed to Aβ (containing 171 pericytes and 270 Iba1-labeled cells). Aβ increased ROS production in pericytes by a factor of 7.28 ($P$ = 0.001) and in immune (Iba1-expressing) cells by a factor of 1.76 ($P$ = 0.05). Taking into account the different numbers and basal ROS production of pericytes and immune cells revealed that Aβ evoked more total ROS generation by pericytes than by immune cells by a factor of 6.4 (Fig. 3D). This is consistent with the data above (Fig. 2, F and G) and suggests that NOX4 in pericytes (*27*–*29*) is the main generator of the ROS involved in constricting capillaries early in the response to Aβ.

To confirm that both pericytes and microglia generate ROS in response to Aβ, in brain slices we fluorescently imaged the level of reduced glutathione (GSH; see materials and methods), which is consumed as it scavenges ROS. Aβ (72 nM for 40 min) reduced the GSH level in pericytes by 20% and in microglia by 55% (fig. S4; $P$ = 0.014 and $P = 2 \times 10^{-29}$, respectively). These changes cannot be converted to ROS synthesis rates because they will be affected by GSH regeneration rate, which may differ in microglia and in pericytes.

## Pericytes constrict capillaries in human cognitive decline patients with Aβ deposition

Because acute exposure to Aβ cannot mimic the slow increase that occurs over decades in human AD patients, we studied rapidly fixed brain cortical biopsy tissue from living patients being investigated for cognitive decline of unknown cause (see materials and methods for demographics, biopsy, and tissue-processing details). Tissue sections were labeled with antibodies recognizing residues 8 to 17 of Aβ and PDGFRβ (Fig. 4, A and B, bottom and top, respectively). Of 13 patients, 7 turned out to have Aβ deposition and 6 did not. Pericytes were readily identifiable from their PDGFRβ labeling. Averaging over 120 to 140 adjacent fields of view (400 μm square in size, randomly placed on each section as a 5 × 4 grid of squares) in tissue from the two types of patient, with the experimenter blinded to the occurrence of Aβ deposition (viewing only the PDGFRβ image

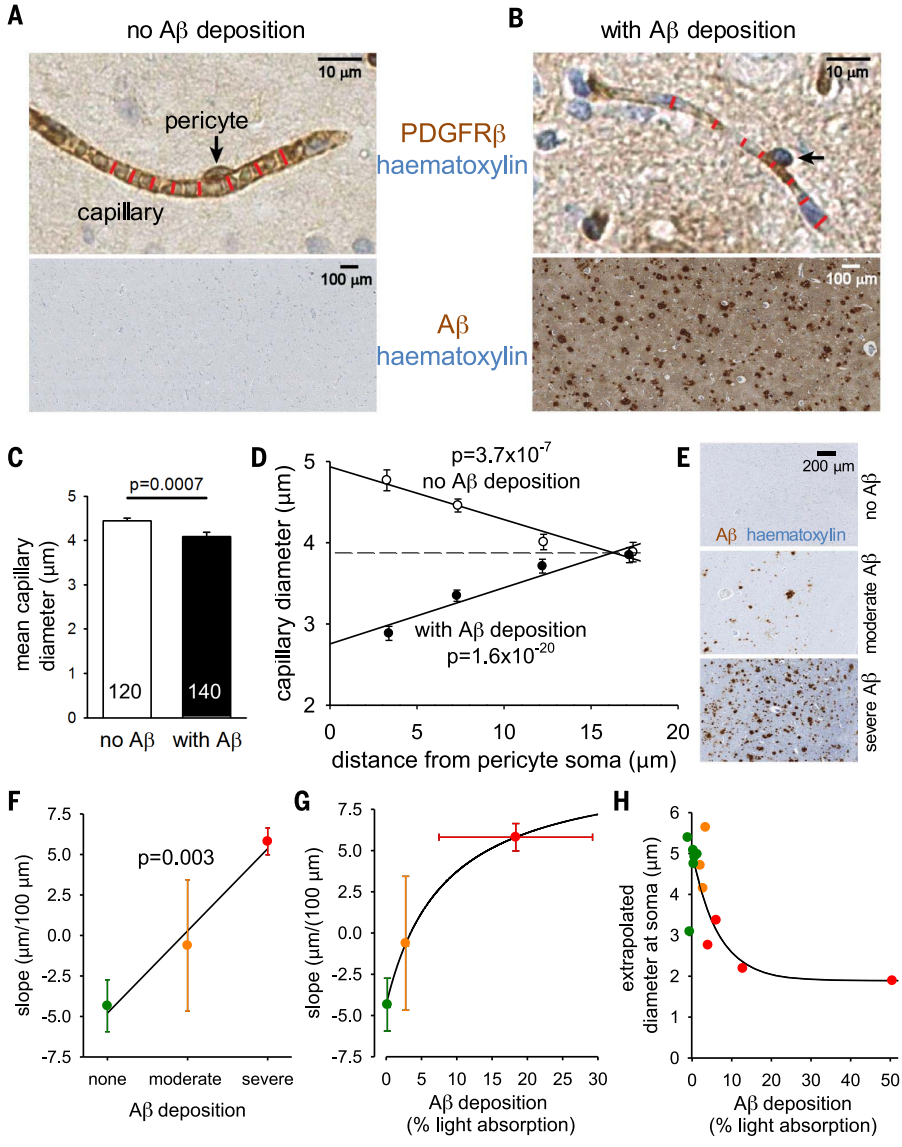



**Fig. 4. Pericyte-mediated capillary constriction occurs in humans with Aβ deposits.** (**A** and **B**) Specimen images of human cortical biopsies, labeled for PDGFRβ (brown in top panels) to show pericytes (arrows), from patients lacking (A) or exhibiting (B) Aβ deposits (brown in bottom panels, hematoxylin counterstain in blue). Red lines indicate capillary diameter. (**C**) Mean (± SEM) diameter of capillaries in patients lacking (3921 diameters measured) or exhibiting (5121 diameters measured) Aβ deposits (numbers of images analyzed shown on bars). (**D**) Dependence of capillary diameter on distance from a visible pericyte soma (in 5-μm bins from 0 to 20 μm, plotted at the mean distance for each bin) for patients lacking or exhibiting Aβ deposits (moderate and severe Aβ deposition pooled together). *P* values assess whether slope of regression line is significantly different from zero. (**E**) Examples of Aβ labeling assessed by the neuropathologist as absent, moderate, or severe. (**F**) Slope of regression lines as in (D) plotted as a function of neuropathologist-rated parenchymal Aβ load for each biopsy ($n$ = 6 biopsies for none, $n$ = 3 for moderate, $n$ = 4 for severe). *P* value compares slope of line with zero. (**G**) Slope of regression lines as in (D) plotted as a function of severity of Aβ deposition measured optically for each biopsy, with subjects grouped by color [defined in (F)] as classified by the neuropathologist. (**H**) Dependence of extrapolated diameter at soma [as in (D)] on severity of Aβ deposition measured optically for each biopsy, with subject points colored as classified by the neuropathologist [defined in (F)]. Lines through data in (F) to (H) show the trends in the data.

channel), we found no significant change in capillary density (12% larger in subjects depositing Aβ, $P$ = 0.56; see materials and methods). However, the mean capillary diameter was reduced by 8.1% ($P$ = 0.0007) in the patients with Aβ deposition (5121 diameters measured) relative to those without Aβ deposition (3921 diameters measured; Fig. 4C).

To assess whether this diameter reduction was a nonspecific effect of AD, or was pericyte-related, we plotted the capillary diameter measurements as a function of the distance from the nearest PDGFRβ-labeled pericyte soma (see materials and methods). In patients with no detectable Aβ deposition, the capillary diameter increased at locations near pericyte somata relative to locations far from the somata (~25% larger, slope of line is significantly less than zero, $P = 3.7 \times 10^{-7}$ for 813 data points from six such patients; Fig. 4D). A similar increase in capillary diameter near somata was previously found in rodent brain capillaries in vivo (*19*) and was attributed to the presence of the soma inducing more growth of the endothelial tube. In contrast, in patients with Aβ deposition, the capillary diameter was significantly reduced near the pericyte somata relative to locations distant from the somata (Fig. 4D; ~30% smaller, slope of line is significantly greater than zero, $P = 1.6 \times 10^{-20}$ for 1313 data points from seven patients), as expected if pericytes cause the capillary constriction by contracting their circumferential processes that are mainly located near their somata. The data shown in Fig. 4D are averaged over all measured pericytes and capillaries (and thus include pericytes on higher–branch order vessels that may be less contractile). For a fixed blood pressure applied at the pial vessels, this average constriction is predicted to reduce flow by ~50% versus what it would be in the absence of constriction (see materials and methods), which is similar to the 42% decrease observed in the gray matter in patients with AD (*3*).

The pericyte soma-specific location of the constriction (Fig. 4D) is consistent with the distribution of circumferential processes relative to pericyte somata (fig. S2) and the fact that exogenous vasoconstrictors constrict capillaries specifically at pericyte locations (fig. S3) (*18*). These data, and the fact that no other cells on capillaries show contractile activity (*18*), imply that it is pericytes that constrict capillaries in human patients depositing Aβ.

## Pericyte constriction of human capillaries increases with Aβ load

The subjects were classified by neuropathologists assessing the Aβ-labeled biopsies as having "no Aβ deposition," "moderate Aβ deposition," or "severe Aβ deposition" in the parenchyma (as diffuse deposits and/or as plaques with central amyloid cores; Fig. 4E). The mean slope for individual patients, from graphs like those in Fig. 4D, for six patients with no Aβ deposition, three patients with moderate deposition, and four patients with severe deposition showed a progressive change from negative (implying a larger capillary diameter at the soma) to positive (implying a smaller diameter at the soma) as the severity of the Aβ deposition increased (Fig. 4F; $P$ = 0.003 compared with a relationship with zero slope). This further supports the idea that Aβ is the cause of the capillary constriction.

To quantify Aβ levels more rigorously, we measured light absorption by the peroxidase product generated by the Aβ antibody, in the region where the vessel diameters were measured

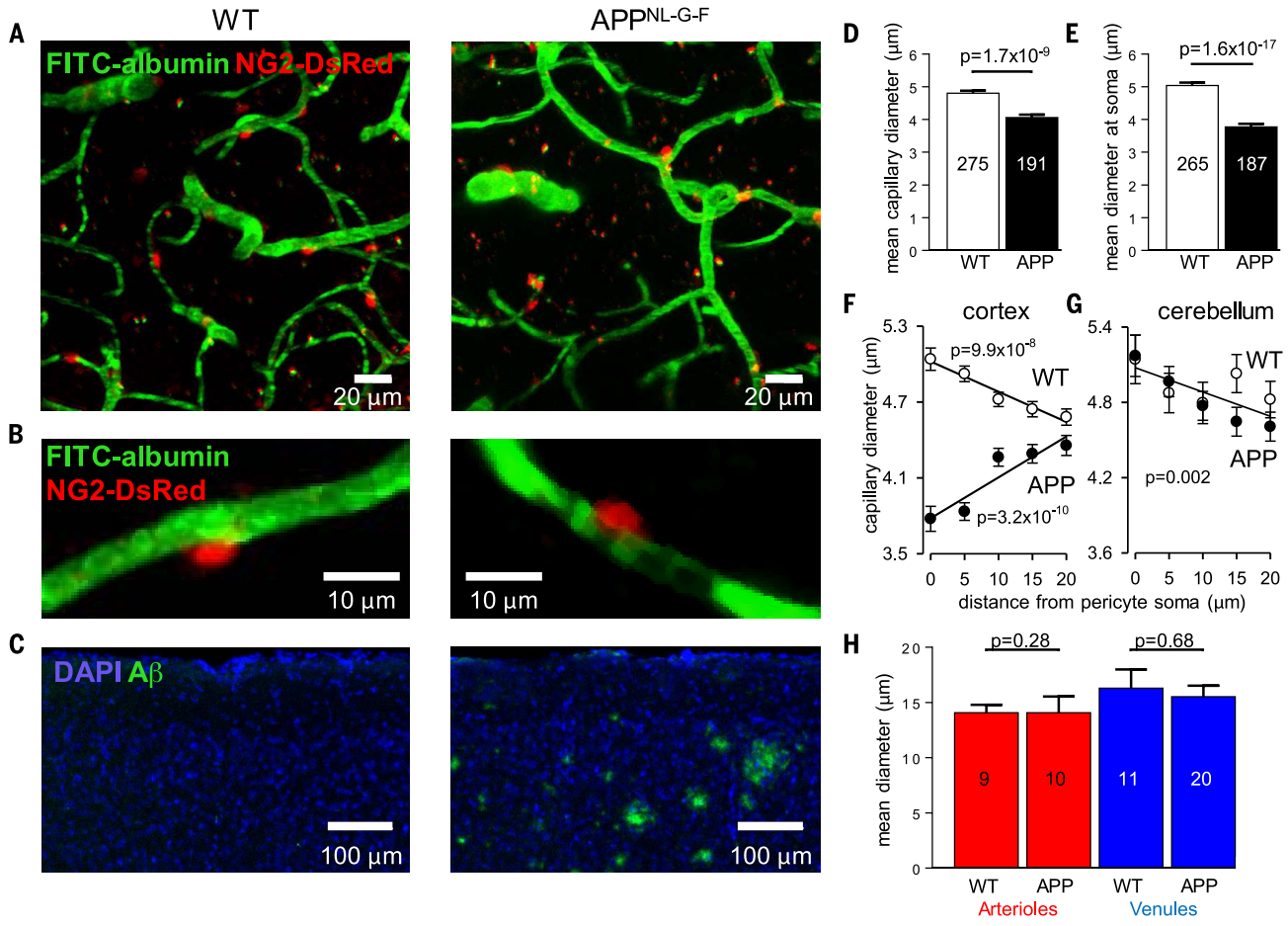



**Fig. 5. Capillaries, but not arterioles or venules, are constricted in AD mice.** (**A**) Specimen images (taken through the dura) of blood vessels in the somatosensory neocortex of wild-type (WT) and homozygous AD ($APP^{NL\text{-}G\text{-}F}$) NG2-DsRed mice, with FITC-albumin (green) in the blood (pericytes are labeled red). (**B**) Examples of single neocortical capillaries and pericytes, showing a larger diameter at the pericyte soma in a WT mouse and constriction of a capillary at the pericyte soma in an AD mouse. (**C**) Images of neocortex labeled for nuclei (DAPI, blue) and for amyloid plaques (green, 82E1 antibody). (**D**) Mean (± SEM) capillary diameter in neocortical layers I to IV in three WT mice (2131 diameters measured; measurements on same capillary were averaged) and four AD mice (1403 diameters measured). Numbers of capillaries are shown on bars. (**E**) Mean neocortical capillary diameter at pericyte somata in three WT and four AD mice (numbers of pericytes on bars). (**F**) Plot of neocortical capillary diameter as a function of distance from pericyte somata shows a smaller diameter at the soma in AD mice and a larger diameter in WT mice (compare with Fig. 4D; each WT mouse studied showed a negative slope for this relationship, and each AD mouse showed a positive slope). (**G**) Plots as in (F) but for the cerebellum, which lacks amyloid plaques, show no constriction near the pericyte somata in the AD mice (regression line is a fit to all data from three WT and three APP mice). (**H**) Mean diameter of neocortical penetrating arterioles and venules in WT and AD mice. Numbers of vessels are shown on bars. Diameters were assessed at depths that did not differ significantly: 158.4 ± 6.7 µm and 131.9 ± 5.0 µm ($P$ = 0.23, Mann-Whitney test) for neocortical capillaries, 142 ± 26 µm and 137 ± 21 µm ($P$ = 0.88) for arterioles, and 85 ± 15 µm and 89 ± 9 µm ($P$ = 0.81) for venules, in WT and AD mice, respectively.

in each biopsy (see materials and methods; although this measure of Aβ may largely reflect the presence of plaques, it is likely that the soluble Aβ concentration correlates with plaque load). Plotting the slopes of graphs like those in Fig. 4D, for each biopsy, as a function of the amount of Aβ deposition again showed a monotonic progression from a negative slope to a positive slope as Aβ deposition increased, but with the change of slope occurring more strongly at low levels of Aβ deposition (Fig. 4G). Similarly, plotting the value of the capillary diameter at the pericyte soma for each biopsy (extrapolated from a straight line fit as in Fig. 4D) as a function of Aβ deposition showed that the diameter was reduced strongly by low levels of Aβ, with smaller increments of constriction as deposition increased (Fig. 4H). This presumably reflects the dose-response curve of Fig. 2D (although it could also reflect increased Aβ production when blood flow is less).

### Pericytes constrict capillaries in vivo in AD mice

To confirm that pericytes constrict capillaries in vivo in AD (i.e., that the constriction seen in human biopsy tissue was not an artifact of fixing the tissue), and to provide a possible framework for future testing of drugs to prevent this constriction, we used in vivo two-photon imaging of layers I to IV of the somatosensory cortex (Fig. 5, A and B) in a mouse model of AD, in which amyloid precursor protein (APP) with a humanized Aβ region with three AD-related mutations ($App^{NL\text{-}G\text{-}F}$) is knocked in (see materials and methods). Comparing four homozygous AD mice and three wild-type mice [age range, postnatal day 119 (P119) to P143, when the AD mice already show plaques; Fig. 5C] revealed that, as in human subjects with and without Aβ deposition, in AD mice the mean capillary diameter was less (Fig. 5D, $P = 1.7 \times 10^{-9}$), and the diameter at pericyte somata was more strongly reduced (Fig. 5E, $P = 1.6 \times 10^{-17}$). A plot of capillary diameter as a function of distance from pericyte somata showed a dilation at the soma in wild-type mice, but a constriction in the AD mice (Fig. 5F), relative to the diameter midway between somata (compare with Fig. 4D).

To check whether capillary constriction was present throughout the brain of AD mice, or occurred only where Aβ levels rise, we imaged capillaries in vivo in the cerebellum—an area that is relatively spared of amyloid plaque pathology in humans, and that had no plaques in our AD mice at P120 to P140 (fig. S5), suggesting lower levels of Aβ oligomers. In the cerebellum, both wild-type and AD mice ($n$ = 3 each) showed a larger capillary diameter near pericyte somata (Fig. 5G; $P$ = 0.002), with no evidence for a constriction in the AD mice. Thus, capillary constriction is associated with Aβ production.

Exogenous Aβ has been reported to constrict isolated penetrating arterioles (*12*) but, at the endogenous level of Aβ produced in the AD mice, arterioles (and venules) were not constricted (Fig. 5H). This may reflect a different response of pericytes and of arteriolar smooth muscle cells to the ET released by Aβ (fig. S6). The lack of arteriole constriction that we observed in the AD mice (Fig. 5H) suggests that capillary constriction is the cause of the decrease in cerebral blood flow that occurs in early AD (*1*).

## Hypoxia is increased in the AD cortex

Our measured capillary constrictions are predicted to decrease cerebral blood flow significantly in AD (see above and materials and methods), as has been observed in human patients and AD mice (*1*, *17*, *21*). Consistent with this, hypoxic tissue labeling by pimonidazole (hypoxyprobe) was increased significantly in vivo in the AD mice (fig. S7).

## Reversal of Aβ-evoked capillary constriction

Prevention or reversal of Aβ-evoked capillary constriction and tissue hypoxia could be a promising therapy in early AD. In brain slices, we investigated two strategies to achieve this (fig. S8), assuming it were possible to target drugs specifically to central nervous system (CNS) capillaries. The first strategy involved combined block of the ROS generator NOX4 (with 0.45 μM GKT137831; Fig. 2F) and of the downstream constricting $ET_A$ receptor (with 1 μM BQ-123; Fig. 2E). This prevented further constriction evoked by Aβ ($P$ = 0.027) but did not reverse the capillary diameter to its baseline value on a 1-hour time scale (Fig. 6A). The second strategy used C-type natriuretic peptide (CNP), which can reverse ET-mediated effects (*36*) by blocking $Ca^{2+}$ release from internal stores and activating myosin light chain phosphatase (fig. S8). Remarkably, CNP (100 nM) reversed the Aβ-evoked capillary constriction ($P$ = 0.029; Fig. 6A).

## Discussion

Genetic evidence strongly implicates Aβ in triggering neuronal damage and cognitive decline in Alzheimer's disease, yet the first change in AD is a decrease of cerebral blood flow (*1*). Our data make five contributions to understanding the vascular effects of Aβ and their role in Alzheimer's disease: (i) Aβ constricts human and rodent capillaries by acting on pericytes; (ii) the mechanism of this constriction involves ROS generation and ET release; (iii) in rapidly fixed biopsies from living human patients with Aβ deposition and cognitive decline, cortical capillaries are constricted by 30% at pericyte locations, which is sufficient to produce a major reduction of cerebral blood flow; (iv) in vivo, in a rodent model of AD, capillaries are constricted by pericytes; and (v) it is in principle possible to reverse the Aβ-evoked capillary constriction. Together, these data imply that the reduction in cerebral blood flow that occurs early in AD results from Aβ-evoked pericyte-mediated constriction of the cerebral capillary bed (Fig. 6B).

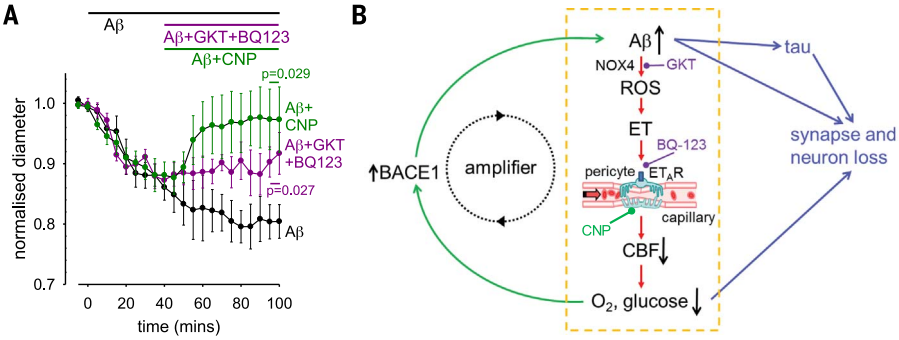



**Fig. 6. Aβ effects on capillaries may amplify the onset of AD and are reversible.** (**A**) Applying GKT137831 (0.45 μM) to block NOX4 and BQ-123 (1 nM) to block $ET_A$ receptors, or applying C-type natriuretic peptide (CNP, 100 nM; see fig. S8), significantly reduced the constriction evoked by Aβ (72 nM, $P$ = 0.027 and 0.029, respectively, corrected for multiple comparisons; data are means ± SEM). (**B**) Summary of our results and their implications. Our data reveal the pathway within the yellow dashed box. Aβ oligomers activate NOX4 in pericytes to generate ROS. These in turn release, or potentiate the constricting effects of, endothelin-1, which acts via $ET_A$ receptors on pericytes on capillaries—the locus (*16*) of the largest component of vascular resistance within the brain parenchyma. Capillary constriction decreases cerebral blood flow and hence the supply of oxygen and glucose to the brain. Green arrows at the left show that this increases the production of Aβ, in part by up-regulating (*13*, *14*) the expression of BACE1, thus forming an amplifying positive feedback loop. Blue arrows at the right show that a rise in Aβ concentration (directly, via downstream tau production, or via the decrease in oxygen and glucose supply) leads to the loss of synapses and neurons. Potential sites for therapeutic intervention are highlighted at the stages of ROS production by NOX4 (GKT), endothelin receptors (BQ-123), and CNP receptors (see also fig. S8).

At low nanomolar concentrations, exogenous soluble $Aβ_{1–42}$ oligomers evoke a constriction of human and rat cortical capillaries, which is mediated by pericytes. Capillaries are the site in the cortical vasculature where most of the resistance to flow is located (*16*), and so may be the major site where Aβ produced within the brain can produce vessel diameter changes that reduce cerebral blood flow. In rodents, the capillary constriction was the result of Aβ evoking the generation in pericytes and microglia of ROS, which evoked a release of ET that acted via $ET_A$ receptors to make pericytes constrict the capillaries. We assume that the $ET_A$ receptors involved are located on the pericytes themselves, because ET raised the $[Ca^{2+}]_i$ in pericytes, but we cannot rule out the possibility that they are on a different cell type. The $EC_{50}$ for the action of $Aβ_{1–42}$, 4.7 nM, is comparable to the concentration of soluble Aβ found in the human AD brain [6 nM, from table 1 of (*37*); note that this brain concentration is higher than the level found in the CSF, which falls during the development of AD as plaques are formed]. Thus, wherever Aβ is produced, or can diffuse, in the AD brain, we would expect all contractile pericytes in that region to constrict capillaries. Indeed, our live human biopsy and in vivo mouse imaging data show that the endogenous level of Aβ reached in AD is sufficient to constrict capillaries. However, some aging humans accumulate Aβ and yet do not develop AD; future work could examine whether, in such people, compensation for the vasoconstricting effects of Aβ develops, such as an up-regulation of vasodilatory mechanisms.

Throughout this work, pericytes were identified by their morphology (spatially isolated cells located outside capillaries) as confirmed by $IB_4$ labeling, or by antibody labeling for their characteristic marker PDGFRβ, or by expression of dsRed under the NG2 promoter. Although arterioles (recognized as being surrounded by rings of abutting smooth muscle cells) have also been reported to be constricted by exogenous Aβ (*12*), in AD mice we found no constriction of arterioles. This may be because, at the ET level reached during AD pathology, ET constricts capillaries but has opposing dilating and constricting effects on arterioles, mediated by different types of ET receptor, which may approximately cancel out. Further work assessing the level of Aβ reached in the AD mice, and the relative affinity of the constricting $ET_A$ and dilating $ET_B$ receptors, will be needed to test this idea.

Three results demonstrate that the effects of Aβ on pericytes that we have demonstrated are pathologically relevant in AD. First, analyzing the diameter of capillaries in biopsies from living human patients with cognitive decline, who either had or lacked Aβ deposition, showed that

Alzheimer's pathology leads to capillary constriction specifically at pericytes. Second, the magnitude of the capillary constriction in human dementia patients increased with the severity of Aβ deposition and is predicted to produce a decrease of cerebral blood flow (~50%) similar to the 42% seen in AD patients (*3*). Capillary constriction by pericytes may explain why some capillaries become occluded by neutrophils in AD (*38*), but neutrophil block of 2% of capillaries, as observed, was predicted to reduce blood flow by only 5% (*38*). Finally, in a mouse model of AD, in vivo imaging showed that cerebral capillaries were constricted at pericyte locations, whereas arterioles and venules were unaffected.

Both the reduction of basal blood flow produced by Aβ and a reduction in the blood flow increase normally produced by neuronal activity (*39*), which may also reflect the constricting action of Aβ on pericytes, will decrease the energy supply to the brain. This in turn increases Aβ production by up-regulating β-amyloid–converting enzyme (BACE1, also called β-secretase 1) (*13*, *14*). Consequently, the pericyte-mediated capillary constriction evoked by Aβ may act as an amplifying mechanism in a positive feedback loop (Fig. 6B), increasing the levels of Aβ and downstream hyperphosphorylated tau, which ultimately lead to the loss of synapses and neurons.

These data suggest several potential therapeutic approaches for early AD, based on the mechanisms generating pericyte constriction. Aβ-evoked generation of ROS by NOX4 in pericytes might be targeted. Indeed, overexpression of SOD1 in APP-overexpressing mice abolishes the lethal effects of the APP overexpression (*26*, *40*). Another approach might be to try to reduce ET release [presumably from brain cells expressing ET strongly, i.e., endothelial cells, microglia, or pericytes (*28*, *29*)] or to block the effects of ET on its $ET_A$ receptors on CNS pericytes. In a proof-of-concept experiment, a combination of a NOX4 blocker and an $ET_A$ blocker prevented further Aβ-evoked constriction (and could conceivably reverse the existing constriction given sufficient time), whereas CNP, which acts via two separate pathways downstream of ET (fig. S8), was able to reverse the constriction in the maintained presence of Aβ. These therapeutic approaches could be tested by targeting drugs to CNS pericytes in the mouse model of AD, which also shows the pericyte-mediated constriction of capillaries. Finally, our scheme (Fig. 6B) prompts the question of what fraction of the damage to synapses and neurons in AD reflects direct actions of Aβ and downstream tau, and what fraction is a consequence of the decrease of energy supply that Aβ produces by constricting capillaries.

## Materials and methods

### *Human brain slices*

The work on fresh living human brain tissue received ethical approval from the National Health Service (REC number 15/NW/0568) and all patients gave informed consent. During neurosurgical operations for tumor treatment, apparently normal cortical tissue that was removed (to gain access to the tumor), which would otherwise have been discarded, was placed in ice-cold brain slicing solution containing 93 mM *N*-methyl-D-glucamine (NMDG) chloride, 2.5 mM KCl, 30 mM $NaHCO_3$, 10 mM $MgCl_2$, 1.2 mM $NaH_2PO_4$, 25 mM glucose, 0.5 mM $CaCl_2$, 20 mM HEPES, 5 mM Na ascorbate, 3 mM Na pyruvate, and 1 mM kynurenic acid (to block glutamate receptors, so as to prevent excitotoxic damage to neurons during the slicing; the experimental solution lacked kynurenic acid, as described below). This solution was oxygenated by gassing with 95% $O_2$/5% $CO_2$ and transported in less than 15 min to the laboratory. Tissue was cut into 200-μm sections and the slices were incubated at 34°C in the same solution for 10 min, and then incubated at room temperature until used in experiments in a similar solution (*41*) with the NMDG-Cl, $MgCl_2$, and $CaCl_2$ replaced by 92 mM NaCl, 1 mM $MgCl_2$, and 2 mM $CaCl_2$. Each patient's tissue typically generated ~2 brain slices. When sufficient tissue was present, histological examination of the slices using hematoxylin and eosin by neuropathologists was used to assess tumor infiltration into the nominally normal tissue. This revealed that some slices showed no infiltration by the tumor, whereas others did. Aβ was applied only to slices that showed no tumor infiltration. Pericyte responses to noradrenaline and glutamate as documented in Fig. 1 were observed whether or not there was tumor infiltration.

### *Rodent brain slices*

Experiments used P21 Sprague-Dawley rats or transgenic mice (as described below) of either sex. All animal procedures were carried out in accordance with EU and UK regulations. Cerebral cortical slices (300 μm thick) were prepared (*18*) and stored as for human slices.

### *Extracellular solution*

Human and rodent brain slices were superfused at 3 to 4 ml/min with aCSF solution containing 124 mM NaCl, 2.5 mM KCl, 26 mM $NaHCO_3$, 1 mM $MgCl_2$, 1 mM $NaH_2PO_4$, 10 mM glucose, 2 mM $CaCl_2$, and 1 mM Na-ascorbate. This solution was gassed with 20% $O_2$/75% $N_2$/5% $CO_2$, which produces a physiological level of oxygen in the slice near the capillaries being imaged (*19*). Mechanism-blocking drugs were superfused for 5 to 15 min before applying Aβ or ET.

### *Imaging capillaries in brain slices*

Healthy capillaries [<10 μm in diameter, mean diameter 5.61 ± 0.03 μm ($n$ = 299) in rat and 5.08 ± 0.33 μm ($n$ = 12) in human, with no rings of arteriolar smooth muscle around them] were selected as described (*41*) and regions of them were imaged, which were in focus in a single image plane over at least 30 μm along the length of a capillary and which exhibited a candidate pericyte with a bump-on-a-log morphology (Figs. 1E and 2A). A CCD camera was used to capture images 100 μm square during superfusion of drugs. An analyst blinded to the time and identity of drug application measured capillary diameter from the resulting movies by placing a line across the lumen on magnified images using Metamorph software. In some experiments, pericytes were identified prior to imaging by incubating slices for 30 min in $IB_4$ (10 μg/ml) conjugated to Alexa 488 or 568 (ThermoFisher I21411 or 121412), which binds to α-D-galactose residues in the basement membrane generated by pericytes and endothelial cells, and outlines pericytes (*41*). This also allowed two-photon imaging (using a Zeiss LSM710 microscope, excitation wavelength 800 nm) of the endothelial tube and the pericytes on it (Fig. 2B).

### *Oligomerizing Aβ and assessing the form and concentration of Aβ applied*

The method used to generate oligomeric Aβ preparations was modified from that previously described (*42*). Synthetic $Aβ_{1–42}$ (Bachem H-1368.1000), $Aβ_{1–40}$ (Bachem H-1194.1000), and scrambled $Aβ_{1–42}$ (Bachem H-7406.1000) were suspended in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP; 52527, Sigma) at 1 mM, vortexed to obtain a homogeneous solution, and aliquoted to microcentrifuge tubes. The HFIP was removed by overnight evaporation and the Aβ was completely lyophilized via a Speed-Vac. The Aβ peptide films were stored desiccated at −80°C until further processed (within 2 weeks). The peptide films were then resuspended at a nominal 5 mM in DMSO, bath-sonicated for 10 min, and vortexed for 30 s. To form Aβ oligomers, this solution was diluted to a nominal 100 μM with phosphate-buffered saline (PBS), vortexed for 15 to 30 s, and incubated at 4°C for 24 hours. Immediately before use, the oligomeric preparations were centrifuged at 14,000*g* for 10 min at 4°C (to remove any fibrils that might be present) and the supernatants were further diluted to the final experimental concentrations (quantified below) with extracellular solution.

Quantification of Aβ peptide concentration was performed using a Pierce BCA protein assay kit (ThermoFisher 23227), calibrated against a known concentration of bovine serum albumin, taking into account the different chromophoric development of albumin and Aβ peptides by multiplying by a factor of 1.51 (*43*, *44*). This showed that the amount of the molecule remaining as soluble monomers and oligomers (i.e., not undissolved or removed as fibrils) was 28.7 ± 2.9% ($n$ = 4) of the nominal concentration added for $Aβ_{1–42}$, 39.9 ± 1.5% ($n$ = 4) for $Aβ_{1–40}$, and 43.6 ± 2.3% ($n$ = 3) for scrambled $Aβ_{1–42}$. Concentrations stated in the text have been corrected for these factors and are given based on the monomeric molecular weight. It was not possible to make pure monomeric preparations of $Aβ_{1–42}$ or $Aβ_{1–40}$.

The Aβ oligomeric preparations were analyzed via SDS-PAGE using 10 to 20% Tris-glycine gels (EC61352BOX, Invitrogen). Samples of Aβ peptides (50 μg) were added to Tris-glycine SDS sample buffer (LC2676, Invitrogen). Equal volumes of each sample (10 μl) were loaded onto gels along with SeeBluePlus2 (Invitrogen) prestained molecular weight markers and electrophoretically separated at 100 V. Gels were stained for total protein using a SilverXpress Silver Staining kit (LC6100, Invitrogen) according to the

manufacturer's protocol. $A\beta_{1–42}$ and $A\beta_{1–40}$ formed monomers and oligomers, whereas scrambled $A\beta_{1–42}$ formed mainly monomers (Fig. 1G). Using densitometry, we estimated that for Fig. 1G the percentages of $A\beta_{1–42}$, $A\beta_{1–40}$, and scrambled $A\beta_{1–42}$ present as monomers (defined as molecular weight 2.5 to 6.5 kDa) were 48%, 39%, and 89%, respectively; the percentages as dimers (MW 6.5 to 11.5 kDa) were 11%, 46%, and 11%; the percentages as trimers (MW 11.5 to 15.5 kDa) were 22%, 6%, and 0%; and the percentages as tetramers (MW 15.5 to 20.5 kDa) were 19%, 4%, and 0%. Thus, the measured $EC_{50}$ of 4.7 nM for the effect of $A\beta_{1–42}$ on constriction in Fig. 1D, which was calculated based on the monomeric molecular weight, would become approximately 4.7 × 0.19 = 0.9 nM if only the tetramer was active.

### Immunohistochemistry of non-biopsy tissue

Human and rat brain slices were fixed in 4% paraformaldehyde (PFA) for 1 hour, washed three times in PBS, then blocked in 10% goat serum/0.5% Triton X-100 in PBS. Primary antibodies for PDGFRβ (Santa Cruz, sc432, 1:200) or α-SMA (Santa Cruz, CGA7, 1:200) or Aβ (IBL, 82E1, 1:500) were applied overnight, followed (after washing in PBS) by application overnight of Alexa Fluor 647 or 633 conjugated secondary antibodies (ThermoFisher, A-21245, A-21070, A-21050, 2 μg/ml). Slices were then washed once in PBS containing DAPI nuclear stain (1:50,000) for 10 min and then washed again in PBS. After mounting, slices were imaged on a Zeiss LSM700 confocal microscope.

### Imaging pericyte $[Ca^{2+}]_i$

Experiments were carried out on acute cortical brain slices from P44 to P88 mice, of either sex, generated by crossing tamoxifen-inducible NG2-$Cre^{ERT2}$ mice (*45*) with floxed GCaMP5G-IRES-tdTomato mice (JAX 024477). Coexpression of the genetically encoded $Ca^{2+}$ indicator GCaMP5G and the morphological marker tdTomato (driven by the CAG promoter after Cre-mediated recombination) was induced by oral gavage of tamoxifen (1 mg per 10 g body weight) for four consecutive days (starting from P23). Brain slices (300 μm thick) were prepared from 21 days after the first tamoxifen administration, as described above for human and rat brain slices. Cortical capillary pericytes, identified in the tdTomato channel from their bump-on-a-log somatic morphology and processes wrapped around capillaries, were imaged using a two-photon microscope (Zeiss LSM 710 or 780) with the two-photon laser (Ti:sapphire Mai Tai DeepSee, Spectra Physics) tuned to 940 nm. Images were acquired with a 20×/1.0 NA water immersion objective (W Plan-Apochromat, Zeiss). Laser power was 5 to 20 mW in the focal plane. Emitted fluorescence was spectrally divided by a 555-nm dichroic mirror and collected by GaAsP detectors. Two-photon image stacks (50 to 200 μm × 50 to 200 μm × 20 to 40 μm; 150 to 300 nm pixel size, 2 μm z-step size, 1.58 to 2.55 μs pixel dwell time) were acquired every 30 s and processed using FIJI (ImageJ). Image stacks were first projected at maximum intensity in the z-dimension, and both channels were co-registered to correct for movement artifacts using the FIJI plugin Multistackreg. Percentage changes in the GCaMP5G fluorescence of regions of interest (ROIs) drawn around pericyte somata were calculated. Oligodendrocyte precursor cells (OPCs), identified on the basis of a star-like morphology and weaker baseline fluorescence, and arteriolar smooth muscle cells were excluded from the analysis.

### Assessing pericyte death

This was carried out as described (*19*). Briefly, brain slices (250 μm thick) were incubated at 36° ± 1°C in a multiwell plate, with 95% $O_2$/5% $CO_2$ blown gently at the surface, in aCSF, or aCSF with oligomerized $A\beta_{1–42}$ or ET added. All extracellular solutions contained $IB_4$ (*41*) to label the basement membrane (ThermoFisher I21411, 10 μg/ml), and hence to label pericytes that are enveloped by this (Fig. 1B), and 7.5 μM propidium iodide to label cells with membranes that had become nonspecifically permeable (*19*). After 3 hours of incubation, slices were fixed in 4% PFA for 2 hours, washed three times with PBS for 15 min each, mounted in DAKO medium, and imaged on a confocal microscope. To avoid counting cells killed by the slicing procedure, quantification of the percentage of pericytes that were dead excluded cells within 20 μm of the slice surface.

### Imaging ROS production

Cellular production of ROS in brain slices was visualized through the $O_2$•-specific oxidation of dihydroethidium to ethidium, which binds to the DNA and RNA of $O_2$•-producing cells (*46*). Rat cortical slices (250 μm thick) were incubated in aCSF or in aCSF containing $A\beta_{1–42}$ (72 nM) or $A\beta_{1–42}$ + SOD1 (150 units/ml) at 34°C. Dihydroethidium (DHE, 8 μM, Cayman, 104821) was added to all solutions immediately before use to avoid auto-oxidation of the dye. No preincubation with DHE was used, so as to limit the intracellular accumulation of oxidized product. After 40 min, the slices were quickly rinsed in PBS, mounted, and immediately imaged using a confocal microscope. A single image stack was acquired at the middle of each slice and for Fig. 3B the fluorescence intensity of the maximum intensity projections was measured using ImageJ. For establishing the identity of ROS-producing cells, slices were fixed in 4% PFA for 20 min and immunostained for NG2 (Millipore AB5320, 1:200) and Iba1 (Synaptic Systems 234006, 1:200). Alexa 488-$IB_4$ (ThermoFisher I21411, 10 μg/ml) was added with the secondary antibodies to also label blood vessels. In maximum intensity projections of z-stacks, ROIs were then drawn around the nuclei of pericytes (NG2-expressing cells on capillaries) and Iba1-expressing immune cells (microglia and perivascular macrophages) and the DHE signal within each ROI was measured in ImageJ. For each z-stack, a mean intensity per pericyte or Iba1-labeled cell, and the total intensity per population of pericytes or Iba1-labeled cells, were calculated, and analysis was performed using the z-stack as the statistical unit.

For glutathione imaging, rat brain slices were incubated with Aβ and fixed as described above, then incubated with 10 mM *N*-ethylmaleimide (NEM) for 4 hours at 4°C and washed thoroughly with PBS. The sections were additionally immunolabeled with a GSH-NEM antibody (Millipore MAB3194, 1:500), which is specific to this adduct, allowing quantification of reduced glutathione after reaction with NEM (*47*). After confocal imaging, ROIs were drawn around the soma of $IB_4$-labeled pericytes and Iba1-expressing immune cells as above and the total fluorescence signal for GSH-NEM was quantified for each cell and averaged over cells.

### Human biopsy data

Diagnostic brain biopsies, comprising cortex and subcortical white matter, were performed as part of routine clinical investigation at the National Hospital for Neurology and Neurosurgery, Queen Square, London, to exclude treatable causes of neurological symptoms that patients showing cognitive decline had presented with. All patients gave informed consent for the biopsy. The use of human tissue samples was licensed by the National Research Ethical Service, UK (University College London Hospitals NRES license for using human tissue samples, project ref 08/0077). The storage of human tissue was licensed by the Human Tissue Authority, UK (License #12054).

Biopsies (volume typically 1 $cm^3$) were all from the right frontal lobe. The biopsies were fixed in 10% buffered formalin less than 30 min after the resection, for a minimum of 12 hours. The formalin-fixed tissue was dehydrated through graded alcohols and embedded in paraffin wax, from which 4-μm-thick sections were cut for routine hematoxylin and eosin staining and a panel of immunohistochemical stains. As part of the diagnostic workup, the sections were immunostained for Aβ with immunoperoxidase-labeled antibody 6F3D (DAKO, 1:50), and for this study in addition with antibody against PDGFRβ (RD systems, MAB1263, 1:20) to label pericytes. This was performed on a Roche Ventana Discovery automated staining platform following the manufacturer's guidelines, using biotinylated secondary antibodies and streptavidin-conjugated horseradish peroxidase and diaminobenzidine as the chromogen. The extent of parenchymal Aβ deposition was assessed semiquantitatively as absent, moderate, or severe by a neuropathologist. In addition, to objectively quantify Aβ deposition, the images of the immunoperoxidase label for Aβ were imported into ImageJ and split into red, green, and blue channels. Then, the light intensity in the blue channel (which gave best distinction of the immunoperoxidase label from the background tissue hematoxylin labeling) was measured in the region of the biopsy where diameters were measured, normalized by the intensity in a region of the section showing no visible Aβ label and converted to a percentage of light absorbed by the Aβ. Normalizing by the intensity in a (tissue-free) region without any tissue absorption gave values that were 5.8 ± 0.5% larger, which did not materially change the form of the

graphs. Although this measure of Aβ may largely reflect the presence of plaques, it is likely that the soluble Aβ concentration correlates with plaque load (*48*).

The mean age of patients without Aβ deposition was 50.5 ± 5.5 ($n$ = 6, 4 women and 2 men), and of those with Aβ deposition was 62.1 ± 4.2 ($n$ = 7, 4 women and 3 men, not significantly different, $P$ = 0.11). Regressing mean capillary diameter against age from all patients, or from the patients lacking Aβ deposition, showed that there was no significant dependence on age ($P$ = 0.5 and $P$ = 0.82, respectively).

Images were analyzed to assess capillary diameter with the experimenter blinded to the level of Aβ deposits (i.e., viewing only the PDGFRβ channel; the condition of the tissue was sometimes worse for patients with Aβ deposition, but it was not possible to unambiguously decide whether the patient had Aβ deposition without viewing the Aβ channel). A standard 5 × 4 grid of 20 squares (each with sides 400 μm long) was superimposed on each image, and all capillaries with clearly demarcated endothelial walls visible in each square had their diameter measured. The image squares were treated as the experimental unit for statistical analysis. Analysis of the diameter as a function of distance from the nearest visible pericyte used a subset of all the measured diameters, because often no pericyte was visible on some short capillary segments. The total number of measurable capillary segments (within the 5 × 4 grid) per subject was not significantly different ($P$ = 0.56) between subjects depositing Aβ (732 ± 96) and subjects not depositing Aβ (654 ± 88), suggesting no detectably greater loss of capillaries in the subjects depositing Aβ.

### Experiments in vivo on AD mice

AD mice, in which APP with a humanized Aβ region containing three AD-related mutations (*App*$^{NL\text{-}G\text{-}F}$) is knocked in (*49*) to avoid artifacts associated with overexpressing APP, were crossed with NG2-DsRed mice in which pericytes express DsRed (*19*). Mice aged ~4 months (P119 to P143, not significantly different for wild-type and AD, $P$ = 0.13) were anesthetized using urethane (1.55 g/kg given in two doses 15 min apart). Adequate anesthesia was ensured by confirming the absence of a withdrawal response to a paw pinch. Body temperature was maintained at 36.8° ± 0.3°C and eyes were protected from drying by applying polyacrylic acid eye drops (Dr. Winzer Pharma). The animal was secured in a stereotaxic frame and lidocaine/prilocaine (AstraZeneca) was applied topically prior to exposing the skull. A custom-built headplate was then attached to the skull using superglue to create a sealed well filled with HEPES-buffered aCSF (140 mM NaCl, 10 mM HEPES, 2.5 mM KCl, 1 mM $NaH_2PO_4$, 10 mM glucose, 2 mM $CaCl_2$, and 1 mM $MgCl_2$) during imaging. A craniotomy of approximately 3 mm diameter was performed over the right primary somatosensory cortex, immediately caudal to the coronal suture and approximately 2 to 6 mm laterally from the midline, or over the right cerebellar hemisphere for imaging cerebellar vessels. The dura was left intact to reduce perturbation of the brain. During imaging, the headplate was secured under the objective on a custom-built stage.

Cortical or cerebellar vessel diameter was recorded using two-photon microscopy of the intraluminal dyes Cascade Blue dextran (MW 10 kDa, Invitrogen, D1976, 1.25 mg in 100 μl of saline given i.v.) or albumin–fluorescein isothiocyanate conjugate (FITC-albumin, Sigma, A9771, 1 mg in 100 μl of saline given retro-orbitally). Two-photon excitation was carried out using a Newport-Spectraphysics Ti:sapphire MaiTai laser pulsing at 80 MHz, and a Zeiss LSM710 microscope with a 20× water immersion objective (NA 1.0). Fluorescence was evoked using a wavelength of 920 nm for DsRed, 820 nm for FITC-albumin, and 800 nm for Cascade-Blue. The mean laser power under the objective did not exceed 35 mW. Penetrating arterioles >10 μm in size were identified by the typical ring shape of vascular smooth muscle cells expressing DsRed in NG2-DsRed (wild-type or APP$^{NL\text{-}G\text{-}F}$) mice. Image stacks were taken in 2-μm depth increments across layers I to IV of the cortex (up to 400 μm deep from the cortical surface). To measure vessel diameter, a line was drawn in ImageJ across the vessel perpendicular to its axis and the width of the intraluminal dye fluorescence was measured, either manually or using an automated routine fitting a Gaussian function to the fluorescence profile in ImageJ and calculating the full width at quarter-maximum of the peak fluorescence intensity (which gave results insignificantly different from the manual measurement).

### Assessing hypoxia in vivo with pimonidazole

Hypoxia was assessed in vivo using the HypoxyprobePlus (HP2-100, Hypoxyprobe Inc.) kit following the manufacturer's instructions. After anesthesia induction with 3% isoflurane in air, animals were switched to 1.5% isoflurane in air and pimonidazole HCl (60 mg/kg) was injected intraperitoneally. Four hours after pimonidazole injection, animals were transferred to urethane anesthesia (1.55 g/kg) and killed by perfusion fixation. Brains were extracted and kept in paraformaldehyde for 24 hours prior to sectioning for immunohistochemistry using the FITC-conjugated antibody provided in the kit, which recognizes conjugates of pimonidazole with protein SH groups in hypoxic cells.

### Statistics

Data are presented as means ± SEM. Data normality was assessed with Shapiro-Wilk or D'Agostino-Pearson omnibus tests. Comparisons of normally distributed data were made using two-tailed Student $t$ tests. Equality of variance was assessed with an F test, and heteroscedastic $t$ tests were used if needed. Data that were not normally distributed were analyzed with Mann-Whitney tests. $P$ values were corrected for multiple comparisons using a procedure equivalent to the Holm-Bonferroni method (for $N$ comparisons, the most significant $P$ value is multiplied by $N$, the 2nd most significant by $N$ – 1, the 3rd most significant by $N$ – 2, etc.; corrected $P$ values are significant if they are less than 0.05). Assessment of whether the slope of linear regressions differed significantly from zero was obtained using the $t$ statistic for the slope. $P$ values comparing vessel diameters in the absence and presence of drugs were calculated for the last data point in each graph shown, or for an exposure time of 45 to 60 min if no graph is shown. An estimate of the sample size needed for a typical experiment is as follows: For a control response of 100%, a response standard deviation of 10%, a response in a drug of 70% (30% inhibition), a power of 80% and $P$ < 0.05, fewer than six vessels are needed in each of the control and drug groups (www.biomath.info/power/ttest.htm). The exact numbers depend on the drug effect size and standard error of the data.

### Calculation of effect of vessel constriction on flow

We assume that pericytes are regularly spaced on capillaries at an interval of 2$L$. For flow governed by Poiseuille's law, the resistance of a segment of capillary of length $L$ (from a pericyte soma to midway between two pericytes) and radius $r_1$ is given by

$$\frac{kL}{{r_1}^4}$$

where $k$ is a constant. If Aβ-induced pericyte contraction reduces the capillary diameter from a value of $r_1$ at the midpoint between pericytes to $r_2$ near the pericyte soma (see Fig. 4, A, B, and D), then, if this reduction is linear with distance, the resistance of the capillary segment from the soma to the midpoint is given by

$$\frac{kL({r_1}^2 + r_1 r_2 + {r_2}^2)}{3{r_1}^3 {r_2}^3}$$

so the factor by which the resistance is altered (relative to that with a uniform diameter $r_1$) is

$$\frac{\left[1 + \frac{r_1}{r_2} + \left(\frac{r_1}{r_2}\right)^2\right]\frac{r_1}{r_2}}{3}$$

Thus, with Aβ deposition, the 30% pericyte constriction reported at pericyte somata in Fig. 4D will increase the resistance by a factor of 2.1 relative to a situation with the capillary having a uniform diameter equal to that measured far from the pericyte somata (~3.9 μm in Fig. 4D), and the 27% increase in diameter at the soma in subjects without Aβ deposition will decrease the resistance to 0.63 of the value with a uniform capillary. Taking the ratio of these changes leads to the conclusion that the capillary constriction occurring with Aβ deposition will increase the capillary resistance by a factor of 3.4 (relative to the condition with no Aβ deposition). Because the capillaries provide 57% of the total vascular resistance in the brain parenchyma (*16*), and because the diameter of arterioles and venules is

not changed (Fig. 5H), it follows that if the pressure is fixed at the pial end of penetrating arterioles and venules, then cerebral blood flow will be decreased by 58%, calculated as (43% + 57%)/[43% + (3.4 × 57%)]. In reality, the flow reduction could be greater than this because Poiseuille's law does not apply for small capillary diameters for which the effective blood viscosity increases as the diameter decreases below 10 μm (*50*). Note that the data in Fig. 4 were averaged over all visible pericytes in the images, and so they already take account of the fact that the contractility of capillary pericytes decreases for higher branch orders of capillary (*19*).

## REFERENCES AND NOTES


1. Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, J. M. Mateos-Pérez, A. C. Evans, Alzheimer's Disease Neuroimaging Initiative, Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nat. Commun.* **7**, 11934 (2016). doi: 10.1038/ncomms11934; pmid: 27327500
2. S. Love, J. S. Miners, Cerebrovascular disease in ageing and Alzheimer's disease. *Acta Neuropathol.* **131**, 645–658 (2016). doi: 10.1007/s00401-015-1522-0; pmid: 26711459
3. I. Asllani *et al*., Multivariate and univariate analysis of continuous arterial spin labeling perfusion MRI in Alzheimer's disease. *J. Cereb. Blood Flow Metab.* **28**, 725–736 (2008). doi: 10.1038/sj.jcbfm.9600570; pmid: 17960142
4. J. Kang *et al*., The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733–736 (1987). doi: 10.1038/325733a0; pmid: 2881207
5. J. Hardy, D. Allsop, Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends Pharmacol. Sci.* **12**, 383–388 (1991). doi: 10.1016/0165-6147(91)90609-V; pmid: 1763432
6. D. J. Selkoe, The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487–498 (1991). doi: 10.1016/0896-6273(91)90052-2; pmid: 1673054
7. T. Kimura, T. Hashimura, T. Miyakawa, Observations of microvessels in the brain with Alzheimer's disease by the scanning electron microscopy. *Jpn. J. Psychiatry Neurol.* **45**, 671–676 (1991). pmid: 1800815
8. J. C. de la Torre, T. Mussivan, Can disturbed brain microcirculation cause Alzheimer's disease? *Neurol. Res.* **15**, 146–153 (1993). doi: 10.1080/01616412.1993.11740127; pmid: 8103579
9. T. Thomas, G. Thomas, C. McLendon, T. Sutton, M. Mullan, β-Amyloid-mediated vasoactivity and vascular endothelial damage. *Nature* **380**, 168–171 (1996). doi: 10.1038/380168a0; pmid: 8600393
10. Z. Suo *et al*., Soluble Alzheimers β-amyloid constricts the cerebral vasculature in vivo. *Neurosci. Lett.* **257**, 77–80 (1998). doi: 10.1016/S0304-3940(98)00814-3; pmid: 9865931
11. R. Deane *et al*., RAGE mediates amyloid-β peptide transport across the blood-brain barrier and accumulation in brain. *Nat. Med.* **9**, 907–913 (2003). doi: 10.1038/nm890; pmid: 12808450
12. H. H. Dietrich, C. Xiang, B. H. Han, G. J. Zipfel, D. M. Holtzman, Soluble amyloid-β, effect on cerebral arteriolar regulation and vascular cells. *Mol. Neurodegener.* **5**, 15 (2010). doi: 10.1186/1750-1326-5-15; pmid: 20388225
13. X. Sun *et al*., Hypoxia facilitates Alzheimer's disease pathogenesis by up-regulating BACE1 gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18727–18732 (2006). doi: 10.1073/pnas.0606298103; pmid: 17121991
14. X. Zhang *et al*., Hypoxia-inducible factor 1α (HIF-1α)-mediated hypoxia increases BACE1 expression and β-amyloid generation. *J. Biol. Chem.* **282**, 10873–10880 (2007). doi: 10.1074/jbc.M608856200; pmid: 17303576
15. K. Niwa *et al*., Aβ-peptides enhance vasoconstriction in cerebral circulation. *Am. J. Physiol. Heart Circ. Physiol.* **281**, H2417–H2424 (2001). doi: 10.1152/ajpheart.2001.281.6.H2417; pmid: 11709407
16. I. G. Gould, P. Tsai, D. Kleinfeld, A. Linninger, The capillary bed offers the largest hemodynamic resistance to the cortical blood supply. *J. Cereb. Blood Flow Metab.* **37**, 52–68 (2017). doi: 10.1177/0271678X16671146; pmid: 27780904
17. R. B. Nielsen *et al*., Capillary dysfunction is associated with symptom severity and neurodegeneration in Alzheimer's disease. *Alzheimers Dement.* **13**, 1143–1153 (2017). doi: 10.1016/j.jalz.2017.02.007; pmid: 28343848
18. C. M. Peppiatt, C. Howarth, P. Mobbs, D. Attwell, Bidirectional control of CNS capillary diameter by pericytes. *Nature* **443**, 700–704 (2006). doi: 10.1038/nature05193; pmid: 17036005
19. C. N. Hall *et al*., Capillary pericytes regulate cerebral blood flow in health and disease. *Nature* **508**, 55–60 (2014). doi: 10.1038/nature13165; pmid: 24670647
20. D. Attwell, A. Mishra, C. N. Hall, F. M. O'Farrell, T. Dalkara, What is a pericyte? *J. Cereb. Blood Flow Metab.* **36**, 451–455 (2016). doi: 10.1177/0271678X15610340; pmid: 26661200
21. E. Gutiérrez-Jiménez *et al*., Disturbances in the control of capillary flow in an aged $APP^{swe}$/PS1ΔE9 model of Alzheimer's disease. *Neurobiol. Aging* **62**, 82–94 (2018). doi: 10.1016/j.neurobiolaging.2017.10.006; pmid: 29131981
22. W. L. Klein, G. A. Krafft, C. E. Finch, Targeting small Aβ oligomers: The solution to an Alzheimer's disease conundrum? *Trends Neurosci.* **24**, 219–224 (2001). doi: 10.1016/S0166-2236(00)01749-5; pmid: 11250006
23. J. Attems, F. Lintner, K. A. Jellinger, Amyloid β peptide 1-42 highly correlates with capillary cerebral amyloid angiopathy and Alzheimer disease pathology. *Acta Neuropathol.* **107**, 283–291 (2004). doi: 10.1007/s00401-004-0822-6; pmid: 14986026
24. L. Alarcon-Martinez *et al*., Capillary pericytes express α-smooth muscle actin, which requires prevention of filamentous-actin depolymerization for detection. *eLife* **7**, e34861 (2018). doi: 10.7554/eLife.34861; pmid: 29561727
25. J. Palmer, S. Love, Endothelin receptor antagonists: Potential in Alzheimer's disease. *Pharmacol. Res.* **63**, 525–531 (2011). doi: 10.1016/j.phrs.2010.12.008; pmid: 21193044
26. C. Iadecola *et al*., SOD1 rescues cerebral endothelial dysfunction in mice overexpressing amyloid precursor protein. *Nat. Neurosci.* **2**, 157–161 (1999). doi: 10.1038/5715; pmid: 10195200
27. J. Kuroda *et al*., Nox4 is a major source of superoxide production in human brain pericytes. *J. Vasc. Res.* **51**, 429–438 (2014). doi: 10.1159/000369930; pmid: 25612841
28. Y. Zhang *et al*., An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014). doi: 10.1523/JNEUROSCI.1860-14.2014; pmid: 25186741
29. A. Zeisel *et al*., Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018). doi: 10.1016/j.cell.2018.06.021; pmid: 30096314
30. J. Luo, P. Grammas, Endothelin-1 is elevated in Alzheimer's disease brain microvessels and is neuroprotective. *J. Alzheimers Dis.* **21**, 887–896 (2010). doi: 10.3233/JAD-2010-091486; pmid: 20634595
31. J. C. Palmer, R. Barker, P. G. Kehoe, S. Love, Endothelin-1 is elevated in Alzheimer's disease and upregulated by amyloid-β. *J. Alzheimers Dis.* **29**, 853–861 (2012). doi: 10.3233/JAD-2012-111760; pmid: 22330820
32. D. Paris *et al*., Vasoactive effects of Aβ in isolated human cerebrovessels and in a transgenic mouse model of Alzheimer's disease: Role of inflammation. *Neurol. Res.* **25**, 642–651 (2003). doi: 10.1179/016164103101201940; pmid: 14503019
33. N. B. Hamilton, D. Attwell, C. N. Hall, Pericyte-mediated regulation of capillary diameter: A component of neurovascular coupling in health and disease. *Front. Neuroenergetics* **2**, 5 (2010). doi: 10.3389/fnene.2010.00005; pmid: 20725515
34. V. Della Bianca, S. Dusi, E. Bianchini, I. Dal Prà, F. Rossi, β-amyloid activates the O-2 forming NADPH oxidase in microglia, monocytes, and neutrophils. A possible inflammatory mechanism of neuronal damage in Alzheimer's disease. *J. Biol. Chem.* **274**, 15493–15499 (1999). doi: 10.1074/jbc.274.22.15493; pmid: 10336441
35. L. Park *et al*., Brain perivascular macrophages initiate the neurovascular dysfunction of Alzheimer Aβ peptides. *Circ. Res.* **121**, 258–269 (2017). doi: 10.1161/CIRCRESAHA.117.311054; pmid: 28515043
36. K. Špiranec *et al*., Endothelial C-type natriuretic peptide acts on pericytes to regulate microcirculatory flow and blood pressure. *Circulation* **138**, 494–508 (2018). doi: 10.1161/CIRCULATIONAHA.117.033383; pmid: 29626067
37. B. R. Roberts *et al*., Biochemically-defined pools of amyloid-β in sporadic Alzheimer's disease: Correlation with amyloid PET. *Brain* **140**, 1486–1498 (2017). doi: 10.1093/brain/awx057; pmid: 28383676
38. J. C. Cruz Hernández *et al*., Neutrophil adhesion in brain capillaries reduces cortical blood flow and impairs memory function in Alzheimer's disease mouse models. *Nat. Neurosci.* **22**, 413–420 (2019). doi: 10.1038/s41593-018-0329-4; pmid: 30742116
39. L. Park *et al*., Aβ-induced vascular oxidative stress and attenuation of functional hyperemia in mouse somatosensory cortex. *J. Cereb. Blood Flow Metab.* **24**, 334–342 (2004). doi: 10.1097/01.WCB.0000105800.49957.1E; pmid: 15091114
40. G. A. Carlson *et al*., Genetic modification of the phenotypes produced by amyloid precursor protein overexpression in transgenic mice. *Hum. Mol. Genet.* **6**, 1951–1959 (1997). doi: 10.1093/hmg/6.11.1951; pmid: 9302276
41. A. Mishra *et al*., Imaging pericytes and capillary diameter in brain slices and isolated retinae. *Nat. Protoc.* **9**, 323–336 (2014). doi: 10.1038/nprot.2014.019; pmid: 24434801
42. M. P. Lambert *et al*., Vaccination with soluble Aβ oligomers generates toxicity-neutralizing antibodies. *J. Neurochem.* **79**, 595–605 (2001). doi: 10.1046/j.1471-4159.2001.00592.x; pmid: 11701763
43. A. Jan, D. M. Hartley, H. A. Lashuel, Preparation and characterization of toxic Aβ aggregates for structural and functional studies in Alzheimer's disease research. *Nat. Protoc.* **5**, 1186–1209 (2010). doi: 10.1038/nprot.2010.72; pmid: 20539293
44. AmideBio, www.amidebio.com/wp-content/uploads/2016/07/Abeta_Quantitation_Protocol.pdf.
45. W. Huang *et al*., Novel NG2-CreERT2 knock-in mice demonstrate heterogeneous differentiation potential of NG2 glia during development. *Glia* **62**, 896–913 (2014). doi: 10.1002/glia.22648; pmid: 24578301
46. L.-J. Wu *et al*., The voltage-gated proton channel Hv1 enhances brain damage from ischemic stroke. *Nat. Neurosci.* **15**, 565–573 (2012). doi: 10.1038/nn.3059; pmid: 22388960
47. S. J. Won, J. E. Kim, G. F. Cittolin-Santos, R. A. Swanson, Assessment at the single-cell level identifies neuronal glutathione depletion as both a cause and effect of ischemia-reperfusion oxidative stress. *J. Neurosci.* **35**, 7143–7152 (2015). doi: 10.1523/JNEUROSCI.4826-14.2015; pmid: 25948264
48. M. P. Murphy, H. LeVine 3rd, Alzheimer's disease and the amyloid-β peptide. *J. Alzheimers Dis.* **19**, 311–323 (2010). doi: 10.3233/JAD-2010-1221; pmid: 20061647
49. T. Saito *et al*., Single *App* knock-in mouse models of Alzheimer's disease. *Nat. Neurosci.* **17**, 661–663 (2014). doi: 10.1038/nn.3697; pmid: 24728269
50. A. R. Pries, T. W. Secomb, P. Gaehtgens, J. F. Gross, Blood flow in microvascular networks. Experiments and simulation. *Circ. Res.* **67**, 826–834 (1990). doi: 10.1161/01.RES.67.4.826; pmid: 2208609


## ACKNOWLEDGMENTS


We thank N. Bazargani, B. De Strooper, M. Ford, N. Fox, A. Gibb, J. Hardy, J. Kittler, D. Kullmann, M. Rice, P. Salinas, and A. Silver for comments on the manuscript, and F. Kirchhoff for NG2-$Cre^{ERT2}$ mice. **Funding:** Supported by European Research Council (BrainPower and BrainEnergy) and Wellcome Trust Investigator Awards (099222/Z/12/Z) (D.A.), the National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre (S.B. and Z.J.), a UCL Sea and Currents grant (A.M. and D.A.), a Leonard Wolfson Experimental Neurology Centre PhD studentship (R.N.), a BBSRC LIDo PhD studentship (N.K.), a Chulabhorn Royal Academy PhD studentship (C.H.), a Wellcome Trust 4-year PhD studentship (P.I.), an EMBO fellowship (T.P.), a Lundbeck Foundation fellowship (L.K.), and Deutsche Forschungsgemeinschaft Sino-German joint project KI 503/14-1 and DFG SFB894 (W.H.). **Author contributions:** R.N. carried out experiments for Fig. 1, Fig. 2, A to E and H to J, Fig. 4, and fig. S6; N.K. carried out the work for Fig. 5 and figs. S2 and S7, and contributed to Fig. 6A; C.H. carried out the work for Fig. 2, F and G, Fig. 6A, and figs. S1 and S3, and analyzed AD mouse data; P.I. carried out the ROS work for Fig. 3 and fig. S4 and plaque imaging for Fig. 5 and fig. S5; A.M. analyzed capillary diameters for Fig. 1 and all of Fig. 2 except panels F and G; Z.J. and A.R.-L. identified biopsy patients and optimized PDGFRβ labeling for Fig. 4; V.K. assessed the oligomeric nature of the Aβ (Fig. 1G) and contributed to work on AD mice; T.P. and L.K. performed [$Ca^{2+}$] imaging for Fig. 2, K and L; C.M. performed imaging for Fig. 2B; H.G. performed and analyzed pericyte death experiments for Fig. 2M; W.H. made NG2-$Cre^{ERT2}$ mice; T.S. and T.C.S. provided AD knock-in mice; S.B. provided facilities for the biopsy work in Fig. 4; H.S. performed the neurosurgery to obtain the tissue for Fig. 1; R.N. and D.A. conceived the project, analyzed the data, and wrote the first draft of the manuscript; and all other authors commented on the manuscript. **Competing interests:** None. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/eaav9518/suppl/DC1
Figs. S1 to S8
References (*51*, *52*)


7 November 2018; resubmitted 10 April 2019
Accepted 4 June 2019
Published online 20 June 2019
10.1126/science.aav9518

RESEARCH ARTICLE

QUANTUM SIMULATION

# String patterns in the doped Hubbard model

**Christie S. Chiu[1], Geoffrey Ji[1], Annabelle Bohrdt[2,1,3], Muqing Xu[1], Michael Knap[2,3], Eugene Demler[1], Fabian Grusdt[1,3], Markus Greiner[1]*, Daniel Greif[1]**

**Understanding strongly correlated quantum many-body states is one of the most difficult challenges in modern physics. For example, there remain fundamental open questions on the phase diagram of the Hubbard model, which describes strongly correlated electrons in solids. In this work, we realize the Hubbard Hamiltonian and search for specific patterns within the individual images of many realizations of strongly correlated ultracold fermions in an optical lattice. Upon doping a cold-atom antiferromagnet, we find consistency with geometric strings, entities that may explain the relationship between hole motion and spin order, in both pattern-based and conventional observables. Our results demonstrate the potential for pattern recognition to provide key insights into cold-atom quantum many-body systems.**

Quantum superposition describes quantum systems as simultaneously realizing different configurations. Such behavior is believed to be at the heart of phenomena in strongly correlated quantum many-body systems, which cannot be described by single-particle or mean-field theories. An intriguing consequence of the superposition principle is the existence of hidden order in correlated quantum systems: Although every individual configuration is characterized by a particular pattern, the average over these configurations leads to an apparent loss of order. By contrast, instantaneous projective measurements have the potential to reveal these underlying patterns.

One notable example of a system with hidden order is the one-dimensional (1D) Fermi-Hubbard model at strong coupling (*1*, *2*). Although 1D chains with additional holes or particles beyond an average of one particle per site (doped) yield average two-point spin correlations that decay more rapidly with distance than chains with an average of one particle per site (half-filled), this magnetic ordering can be revealed by accounting for the fluctuating positions across individual configurations of the additional dopants within each chain. The apparent loss of magnetic order is in fact hidden order, hidden by the dopants and their varying positions (*3*, *4*). Although direct detection of this hidden string order remains inaccessible in solids, experiments with ultracold atoms enable projective measurements, or "snapshots," and generally can provide access to such structures (*5*). In particular, quantum gas microscopy (*6*) enables site-resolved imaging and access to correlators that have been constructed to reveal the hidden order (*7*).

The hidden order in 1D is well understood, but the physics of the 2D Hubbard model is fundamentally more complex because of an intricate interplay between spin and charge degrees of freedom; as a result, formulating an appropriate correlation function to search for hidden order becomes considerably more challenging. The 2D Hubbard model is believed to capture the rich physics of high-temperature superconductivity and other phases (*8*–*10*) such as the strange metal, stripe, antiferromagnet (AFM), or pseudogap phase, but a unified understanding of these phenomena is still lacking. For example, the behavior of individual dopants in an AFM is not agreed upon, including whether hidden string order is present and dopants hide magnetic correlations by shifting the positions of a string of spins. Quantum gas microscopy, however, provides a perspective that goes beyond the framework of two- or multipoint correlations. Hidden string order can be searched for directly within individual snapshots of the quantum mechanical wave function, where quantum fluctuations are resolved.

Here we perform a microscopic study of the hole-doped Fermi-Hubbard model and report indications of string patterns in 2D over a wide doping range. Our measurements use ultracold fermions in an optical lattice down to the lowest currently achievable temperatures, where at low doping AFM correlations extend across the system size (*11*). We identify string patterns in individual projective measurements and compare them with predictions from microscopic theoretical approaches.

[1]Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA. [2]Department of Physics and Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany. [3]Munich Center for Quantum Science and Technology (MCQST), Schellingstr. 4, D-80799 München, Germany.
*Corresponding author. Email: greiner@physics.harvard.edu

## Candidate theories for the doped Hubbard model

We study the Fermi-Hubbard model, which is defined by the Hamiltonian

$$\hat{\mathcal{H}} = -t \sum_{\sigma=\uparrow,\downarrow} \sum_{\langle \mathbf{i},\mathbf{j} \rangle} (\hat{c}^{\dagger}_{\mathbf{i},\sigma} \hat{c}_{\mathbf{j},\sigma} + \text{h.c.}) + U \sum_{\mathbf{j}} \hat{c}^{\dagger}_{\mathbf{j},\uparrow} \hat{c}_{\mathbf{j},\uparrow} \hat{c}^{\dagger}_{\mathbf{j},\downarrow} \hat{c}_{\mathbf{j},\downarrow} \quad (1)$$

(see Fig. 1A). The first term describes tunneling of amplitude $t$ of spin-½ fermions $\hat{c}_{\mathbf{j},\sigma}$ with spin $\sigma$ between adjacent sites $\mathbf{i}$ and $\mathbf{j}$ of a two-dimensional square lattice. The second term includes on-site interactions of strength $U$ between fermions of opposite spin. We consider the strongly correlated regime, where $U \gg t$ and doubly occupied sites are energetically costly.

The Fermi-Hubbard model is well understood when the band is half-filled at an average of one particle per site (Fig. 1B). For temperatures $T \ll J$, where $J = 4t^2/U$ is the superexchange coupling, AFM correlations appear. Although these magnetic correlations are finite-ranged at nonzero temperatures, sufficiently cold finite-size systems can have AFM order across the entire system (*11*).

Much less is known about the doped Fermi-Hubbard model. However, it is understood that dopant delocalization for kinetic energy minimization competes with spin interactions in the background AFM. Experiments on the cuprates have also shown that at temperatures $T < J$ and between 10 and 20% doping, the pseudogap phase crosses over to the strange metal, located above the superconducting dome (*9*). These two metallic phases (pseudogap and strange metal) defy a description in terms of conventional quasiparticles and still lack a unified theoretical understanding.

Although phenomenological, numeric, and mean-field (MF) approaches have provided key insights in the past, quantum gas microscopy is naturally suited to assess microscopic theoretical approaches. One such theory is Anderson's resonating valence bond (RVB) picture (*12*), which considers trial wave functions of free holes moving through a spin liquid composed of singlet coverings. We consider one particular class of RVB wave functions that have been studied extensively, called π-flux states (*13*). They stem from an MF density matrix $\hat{\rho} = \hat{\mathcal{P}}_{\text{GW}} e^{-\hat{\mathcal{H}}_{\text{MF}}/k_{\text{B}}T} \hat{\mathcal{P}}_{\text{GW}}$, where $k_{\text{B}}$ is Boltzmann's constant, $\hat{\mathcal{P}}_{\text{GW}}$ is the Gutzwiller projection, and $\hat{\mathcal{H}}_{\text{MF}}$ is the quadratic Hamiltonian of itinerant fermions on a square lattice with a Peierls phase of π per plaquette [see section 6.1 of (*14*) for details]. Snapshots of the trial state in the Fock basis can be obtained by Monte-Carlo sampling, with temperature $T$ as a free fit parameter (*15*).

A second microscopic approach that we examine is the geometric-string theory (*16*), where AFM order at half-filling is hidden in doped states via hole motion. This theory extends earlier work (*17*–*19*) and establishes a relationship between the AFM parent state at half-filling and the strongly correlated quantum states at finite doping. Here, holes move through the parent AFM by displacing

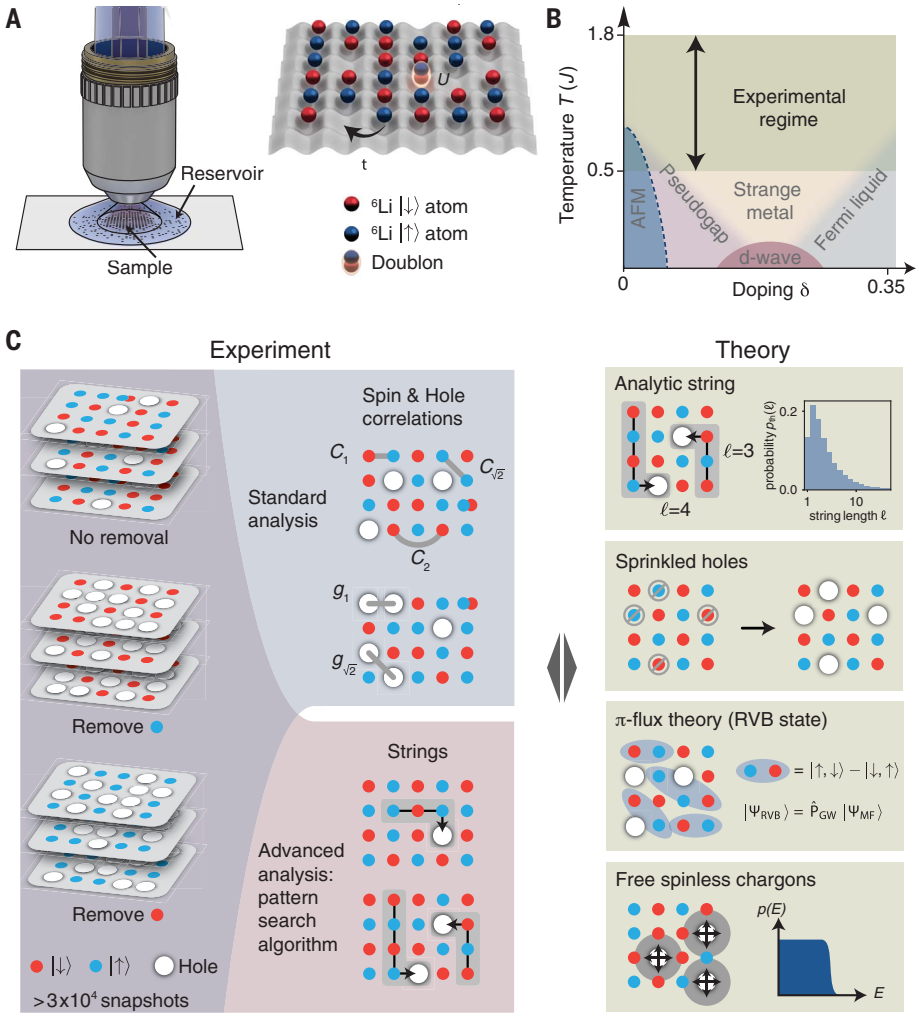



**Fig. 1. Quantum simulation of the Hubbard model.** (**A**) Quantum gases trapped in optical lattices realize the Hubbard model with tunable on-site interaction $U$ and nearest-neighbor hopping $t$. Quantum gas microscopy enables site-resolved readout of the quantum state. (**B**) Schematic of the conjectured phase diagram of the finite-size 2D Hubbard model with the experimentally accessed regime (green shading). (**C**) Outline of experimental observables used and theoretical models evaluated. We evaluate theories using both standard observables and pattern-recognition–based observables using snapshots of the quantum state.

each spin along its trajectory by one lattice site, whereas the AFM quantum state remains otherwise unmodified; this is the frozen-spin approximation (*20*). The delocalization of each hole can then be described as a superposition state of hole trajectories, or geometric strings, whose lengths $\ell$ depend on the strength of AFM correlations and the ratio of the kinetic energy $t$ to the superexchange $J$. For any given temperature, a distribution function $p_{\mathrm{th}}(\ell)$ of string lengths can be obtained by sampling a Boltzmann distribution of string states (Fig. 1C).

We directly assess these microscopic theoretical approaches with a quantum gas microscope, which provides projective measurements of the quantum mechanical wave function for the doped Hubbard model in the parity-projected Fock basis. Our experimental setup consists of a balanced two-component gas of fermionic lithium in the lowest band of a square optical lattice (*21*), with $U/t$ set to 8.1(2). We selectively image one of the spin states or the total atom distribution (*22*). Entropy redistribution with a digital micromirror device enables a disk-shaped homogeneous system of ~80 sites with temperatures as low as $T/J = 0.50(4)$ (*11*). We alter the local chemical potential to dope the system, maintaining independent temperature control [section 7.1 of (*14*)]. We determine the doping from the single-particle occupation density and temperature from the nearest-neighbor spin correlator, both by comparing to numerics [section 2 of (*14*)].

## Pattern recognition of geometric strings

We design a pattern recognition algorithm for geometric strings that we apply to real-space snapshots where doublons and one of the two spin states have been removed (Fig. 2A). Because geometric strings describe a relationship between doped and half-filled AFMs, we search for string-like patterns in the deviation between snapshots of the doped Hubbard model and an approximation to the AFM, an exact checkerboard. For each image, we take the set of sites that deviate and extract string patterns using the following rules: (i) every string pattern is a connected subset of sites forming a path without branching points; (ii) each site can be part of only one string pattern; (iii) longer string patterns are favored; and (iv) every string pattern must have at one end a site that is detected as empty, and therefore consistent with having a hole on that site. We discuss alternate algorithms in section 3.4 of (*14*).

We find that this algorithm is indeed sensitive to hole doping. Figure 2B shows string-pattern length distributions $p^{\delta}(\ell)$ over pattern lengths $\ell$, averaged over experimental data at temperatures between $0.50(4)J$ and $0.70(3)J$. As the sample is doped from half-filling to a doping $\delta$ of 10.0(8)%, the number of string patterns increases across the entire range of lengths. The appreciable distribution of string patterns $p^{0}(\ell)$ detected at half-filling reflects the deviation of a quantum AFM from our checkerboard approximation and therefore should be considered as a baseline level. This baseline can be reproduced through Heisenberg quantum Monte Carlo simulation [see section 3.3.3 of (*14*)] and is largely caused by the finite temperature and underlying SU(2) symmetry of the system. We have lessened these contributions by reducing the analysis region to a diameter of seven sites and postselecting on the staggered magnetization. In section 3.3 of (*14*), we show that results are robust to the choice of postselection scheme and that the limited detection of one of the spin states causes only an overall factor decrease in string patterns detected.

Next, we compare our experimental results to the simulation results of three microscopic models. We make predictions by producing artificial images and evaluating them with our string pattern detection algorithm, such that the detection is common to experiment and theoretical simulation. Beginning with the analytic string model, we generate images by randomly placing a number of holes into actual experimental images taken at half-filling, then randomly propagating each hole according to the analytically generated string-length histogram (see Fig. 1C) and appropriately displacing the spins along the hole's path. Note that this approach preserves the SU(2) symmetry of the system. The resulting string-pattern length distribution agrees with experimental data [see Fig. 2B for 10% doping], even though the theory has no free parameters.

To verify whether our measured signal simply results from the introduction of holes rather than changes to the spin background, we next compare our experimental result with simulations where holes are artificially and randomly placed ("sprinkled") into experimental data taken at half-filling, equivalent to placing one-site-long strings. The associated string-pattern length distribution $p_s^{\delta}(\ell)$ fails to explain the experimental results, revealing the nontrivial interplay of spin and charge degrees of freedom in the 2D doped Hubbard

model. Last, we compare our experimental result to π-flux states by fitting the nearest-neighbor spin correlator for an effective temperature and producing simulated images at 10% doping, and find quantitative agreement with experiment at short pattern lengths, but a deficit at long lengths.

We repeat the measurements for a sample heated before lattice loading to investigate temperature effects. Figure 2C shows experimental data at half-filling and at 10.1(8)% doping, along with the simulated prediction, averaged over samples at temperatures between 1.3(1)$J$ and 1.8(1)$J$. In contrast to colder temperatures, there is no statistically significant deviation between the experimental data with and without hole doping; $p^{0.1}(\ell) \approx p^{0}(\ell)$. For these temperatures, spin ordering is so weak that the resulting string patterns may mask additional effects from doping. These deviations appear to set an upper bound on the density of detectable string patterns (Fig. 2E); we therefore plot the pattern length distribution for high-temperature and half-filling as a reference for the cold temperature datasets in Fig. 2B (gray dash-dotted line).

In Fig. 2D and its inset we plot, respectively, the relative and absolute differences between the pattern-length histograms in the doped and undoped cases; these differences are shown for both the cold and hot datasets used in Fig. 2, B and C. For the undoped case, we use the sprinkled string-pattern length distribution $p_s^{\delta}(\ell)$ to account for any deviation from the half-filling distribution resulting from the introduction of holes. Although the absolute difference does not recover the exact analytic string distribution (Fig. 1C), which can be attributed to the imperfect detection of the pattern recognition algorithm, for cold temperatures it does assume a qualitatively similar distribution. Notably, at 10.0(8)% doping we find more than three times as many length-9 patterns as there are at half filling, reflecting the large impact of holes in an AFM spin background.

Focusing on the cold dataset, we now examine the relationship between doping and the number of detected string patterns (Fig. 3A). In this string-pattern count, we omit patterns of one or two sites to avoid contributions from quantum fluctuations such as doublon-hole pairs or spin-exchange processes. The string-pattern count increases with doping and saturates at about 16% doping. This saturation is consistent with a high density of strings and overlapping or adjacent strings scrambling spin order such that pattern detection becomes insensitive to additional strings. The continued agreement between geometric strings and experiment in both the string-pattern count and the absolute difference $p^{\delta}(\ell) - p_s^{\delta}(\ell)$ suggests that the increase in number of string states is sufficient to explain the experimental data.

The experimental string-pattern count is significantly larger than that of the sprinkled-hole simulation; nonetheless, there is an increase in detected string patterns owing to the additional holes. The string-pattern count from π-flux states shows considerably better agreement with experimental data than with sprinkled holes, exhibiting only a slight excess of string patterns at low doping and a deficit at high doping. The largest deviations occur at low doping, which may be related to the absence of long-range order at zero temperature in π-flux states at half-filling.

The average string-pattern length quantifies the size of the region around the hole where the spin pattern is distorted by the string (Fig. 3B). The observed values are comparatively small, influenced by the large contributions from quantum fluctuations at half-filling. The average string-pattern length does not change substantially with doping, consistent with spatially isolated patterns; however, at larger dopings, we observe a slight decrease in average length that coincides with the observed saturation in the string count. This behavior is captured by the geometric-string model for low and intermediate doping. At high doping, the theory exhibits shorter average string lengths than the experiment, which may result from high–string-density effects such as string-string interactions, which are not included in the theory.

We compare these results to a dataset in which geometric strings are not expected to occur. This dataset consists of experimental images taken at various temperatures at half-filling with sprinkled holes to match each desired doping level [for

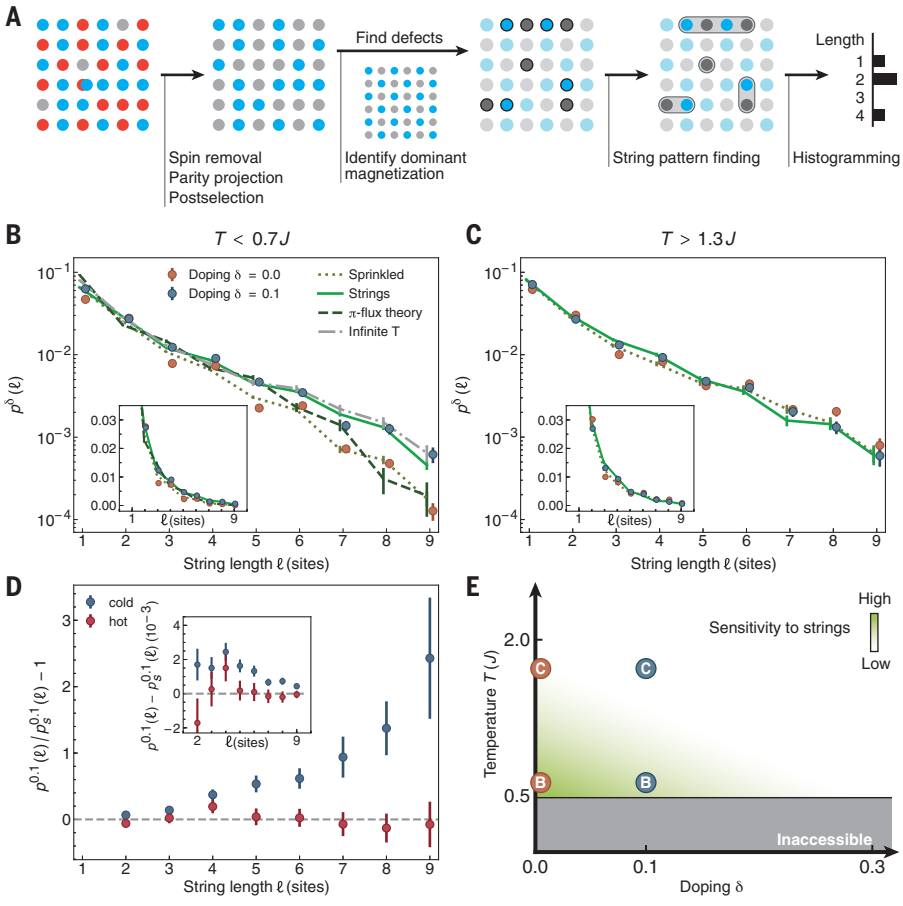



**Fig. 2. Measurement of string-pattern length histograms from site-resolved snapshots.** (**A**) Schematic explanation of the string-pattern identification algorithm (see text). (**B** and **C**) Change in string-pattern length histograms upon doping to 10% and comparison with simulated models at 10% doping, for temperatures below (B) and above (C) the superexchange energy $J$. The observable is only sensitive to doping in the colder dataset, and simulated strings seem to fit the doped experimental result best. Points have been slightly offset horizontally for readability and insets plot the same data on linear-linear axes. (**D**) Relative and absolute (inset) difference between doped and undoped pattern-length histograms, highlighting temperature-dependent sensitivity. The sprinkled-hole result is used for the undoped case as it accounts for the change in density. (**E**) Regions of the phase diagram examined in (B) and (C). The string-pattern observable has sensitivity at temperatures below $J$ and below intermediate doping. In (B), (C), and (D), histograms are normalized by the number of lattice sites analyzed and error bars represent 1 SEM from more than 5500 (half-filling, cold), 3500 (doped, cold), 2900 (half-filling, hot), and 4600 (doped, hot) images.

details, see section 3.5 of (*14*)]. Temperatures are chosen to match the measured staggered magnetization and capture the observed loss of AFM order. Notably, the average string-pattern length reveals that this loss through heating occurs in a fundamentally different way than through doping. For all nonzero doping, the temperature-based dataset exhibits shorter average string-pattern lengths than the experimentally measured doping dataset. As doping increases, the average length monotonically decreases. Alternatively, we match the nearest-neighbor spin correlator instead of the staggered magnetization and find an even greater distinction between the doped and temperature-based datasets.

We better understand the role of temperature in string-pattern detection by observing how the string-pattern count varies with temperature at fixed doping. For 10% doping, we plot the difference between the experiment and sprinkled-hole string-pattern counts (Fig. 3C), which are plotted separately in the inset. At our lowest temperatures, the difference is greatest. This high sensitivity is consistent with the greatest spin ordering for the parent AFM at low temperatures, accompanied by a relatively large string-pattern count from the experimental data. The difference decreases steadily with increasing temperature, predominantly owing to the increase in the sprinkled-hole string-pattern count from decreased spin ordering in the parent AFM, vanishing around $T = J$.

## Spin correlations and staggered magnetization

An accurate microscopic framework for the Fermi-Hubbard model should also be able to predict more conventional observables such as two-point correlation functions, which have been used with quantum gas microscopes to quantify spin and charge order (*22*–*25*). To that end, we measure the sign-corrected spin-spin correlation function for displacements $|\mathbf{d}| = d$, averaged over all sites $\mathbf{i}$ in the system and all experimental realizations

$$C_s(|\mathbf{d}|) \equiv (-1)^{||\mathbf{d}||}\frac{\langle \hat{S}^z_{\mathbf{i}}\hat{S}^z_{\mathbf{i}+\mathbf{d}}\rangle - \langle \hat{S}^z_{\mathbf{i}}\rangle\langle \hat{S}^z_{\mathbf{i}+\mathbf{d}}\rangle}{S^2} \quad (2)$$

where $\hat{S}_{\mathbf{i}}z$ is the spin-$S$ operator on site $\mathbf{i}$, $S = 1/2$, and $||\mathbf{d}||$ denotes the $L^1$ norm of $\mathbf{d}$, by measuring charge correlations in experimental realizations with and without spin removal (*22*). Thanks to the sign correction $(-1)^{||\mathbf{d}||}$, positive correlator values indicate AFM ordering. Figure 4A shows the nearest neighbor, diagonal next-nearest neighbor, and straight next-nearest neighbor spin correlators [$C_s(1)$, $C_s(\sqrt{2})$, and $C_s(2)$, respectively] as a function of doping at $T = 0.65(4)J$. At half-filling, $C_s(1)$ is substantially larger than both $C_s(\sqrt{2})$ and $C_s(2)$, owing to a strong admixture of spin singlets on adjacent sites (*26*). As the system is doped, all correlators exhibit a reduction in magnitude. $C_s(1)$ remains positive for all experimentally realized doping values, whereas $C_s(\sqrt{2})$ exhibits a statistically significant sign change around 20% doping. These features have been observed in experiment (*22*, *24*, *27*) and numerics (*24*), and are good benchmarks for the evaluation of theoretical models.

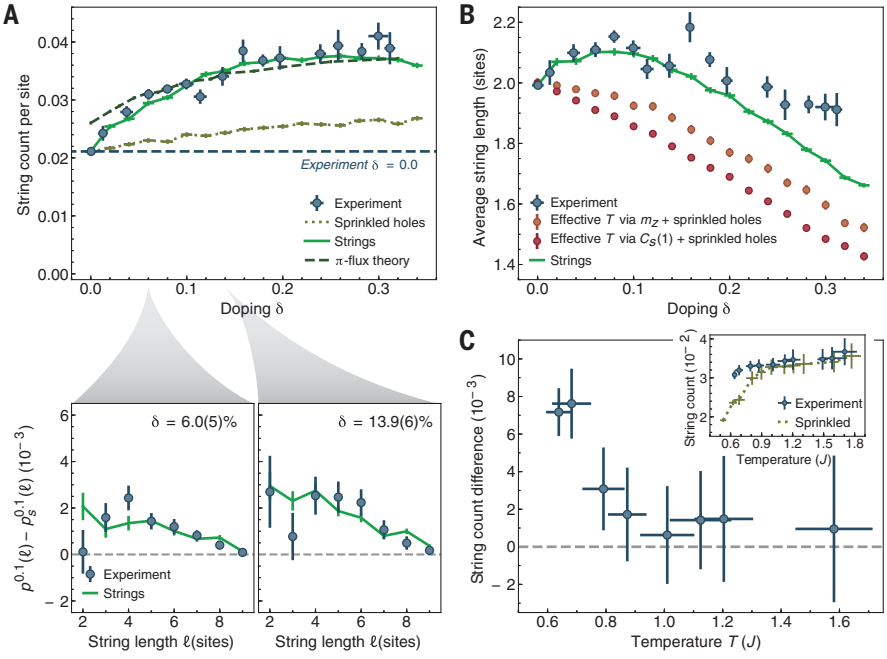



**Fig. 3. Detailed examination of the detected string patterns upon doping.** (**A**) (Top) Total number of string patterns exceeding length 2, normalized by the system size, as a function of doping. Although the string model and sprinkled-hole simulation both agree with experiment at half-filling by construction, already at low doping the string model performs significantly better than sprinkled holes. The string model is quantitatively accurate across a larger doping range than for π-flux states, but both are in greater agreement with experiment than the sprinkled-hole simulation. (Bottom) Although the absolute difference between doped and sprinkled-hole pattern-length histograms increases with doping, the shape remains roughly invariant. (**B**) Average string-pattern length versus doping. Doped AFMs exhibit longer-length string patterns compared to heated AFMs, even when the staggered magnetization or nearest-neighbor spin correlator is equal and holes are sprinkled in to equate doping levels (see text). (**C**) Total string count at 10% doping as a function of temperature, with corresponding sprinkled-hole string count subtracted. Sensitivity to strings decreases with temperature owing to decreased order in the parent AFM as seen in the sprinkled string count (inset). In (A) and (B), error bars on the doping are calculated as in (*14*), section 2.1. All other error bars represent 1 SEM. The figure is based on more than 24,800 experimental realizations.

We make predictions for spin correlations from ensembles of non-postselected images with sprinkled holes, geometric strings, or π-flux states. By construction, at half-filling the predictions of sprinkled holes and the string model are the same as those of experimental half-filling data. Away from half-filling, sprinkled holes underestimate the decrease of the correlators because the model fails to account for the disruption of AFM order as the system is doped. By contrast, beginning at intermediate doping values, the string model overestimates the decrease of $C_s(1)$, which could stem from backaction of the background state after string-state formation. However, it explains the decrease of $C_s(\sqrt{2})$ and $C_s(2)$ on a quantitative level. The π-flux model performs well and accurately predicts $C_s(1)$ and $C_s(2)$ far from half-filling but fails to predict the sign change of $C_s(\sqrt{2})$ at intermediate doping, even when the fitted temperature is varied. The sign change of $C_s(\sqrt{2})$ is an interesting qualitative feature that is predicted and can be explained by the string model. As a direct result of spins being displaced by one site when a string passes through, $C_s(1)$ is mixed into $C_s(\sqrt{2})$. Because $C_s(1)$ reflects opposite spin alignment from $C_s(\sqrt{2})$, this mixing results in a sign change once the contribution of $C_s(1)$ exceeds that of the original correlation strength at some critical doping.

Cold-atom experiments provide access to full-counting statistics (FCS) because of their ability to project and measure an entire quantum system at once (*11*). We measure the FCS of the staggered magnetization operator

$$\hat{m}^z = \frac{1}{N}\sum_{\mathbf{i}}(-1)^{||\mathbf{i}||}\frac{\hat{S}^z_{\mathbf{i}}}{S} \quad (3)$$

for system size $N$ across all experimental realizations as we dope the system (Fig. 4B). As expected, the staggered magnetization distribution narrows, reflecting the finite-size crossover from the AFM-ordered phase (*11*). The sprinkled-hole simulation does not exhibit a major change in the distribution as the system is doped, as it fails to account for holes disrupting the AFM order. By contrast, both π-flux states and geometric strings demonstrate reasonable agreement with the experimentally

measured distribution function across all dopings. Across all observables considered, both of these theories perform quite well, especially in comparison to the sprinkled-holes simulation and the naïve phenomenological models detailed in section 5 of (*14*). However, we find the sign change of $C_s(\sqrt{2})$ to be a key qualitative feature that is captured only by geometric strings.

### Antimoment correlations

All observables studied in this work thus far have focused on the spin sector of the Hubbard model. Next, we examine correlations in the charge sector. At sufficiently low temperatures, one may expect signatures of pairing (*10*, *28*) or stripe phases (*29*, *30*), which lead to hole bunching. However, anticorrelations of the holes, as observed previously at increased temperatures (*24*), are expected in the strongly correlated metallic regime of the Hubbard model. The transition between these two regimes in the Hubbard model phase diagram is not yet fully understood; however, the currently accessible experimental regime allows us to place more accurate bounds on where this transition can occur. We continue to compare experimental results to predictions of π-flux states, but do not compare to predictions of the geometric string theory because it approximates that charges are uncorrelated. Rather, because each string is associated with a single hole, correlation functions of holes can reveal possible interactions and correlations between geometric strings, should they exist.

In our experiment, doubly occupied sites appear as empty when imaged and the exact hole correlation is not directly accessible; rather, we measure "antimoment" correlations $C_h(|\mathbf{d}|)$ at a distance $|\mathbf{d}|$, which include contributions from doublon-doublon and doublon-hole correlations:

$$C_h(|\mathbf{d}|) \equiv \Big(\langle(1-\hat{n}_{s,\mathbf{i}})(1-\hat{n}_{s,\mathbf{i}+\mathbf{d}})\rangle - \langle(1-\hat{n}_{s,\mathbf{i}})\rangle\langle(1-\hat{n}_{s,\mathbf{i}+\mathbf{d}})\rangle\Big) \quad (4)$$

where $\hat{n}_{s,\mathbf{i}}$ is the single particle occupation on site $\mathbf{i}$. Note that this correlator is identical to the moment correlator. At half-filling, numerics indicate positive antimoment correlations at the percent level for nearest neighbors, dominated by positive doublon-hole correlations (*24*). Doublon-hole pairs beyond nearest neighbors become increasingly unlikely; therefore, to avoid the effects of doublon-hole pairs, we focus on correlations at distances greater than 1. We find the nearest-neighbor antimoment correlator at half-filling to be weaker than predicted according to numerics, which may be caused by imperfect imaging fidelity. However, this effect only weakens the magnitude of the antimoment correlators measured; we therefore focus on qualitative conclusions from the experimental data.

Figure 5A shows the antimoment correlation for 3% (top) and 19% doping (bottom) at a temperature $T = 0.65(4)J$. Whereas holes appear uncorrelated close to half-filling, at larger doping qualitatively different behavior appears. We find statistically significant antimoment anticorrelations out to distances over two sites, reflecting

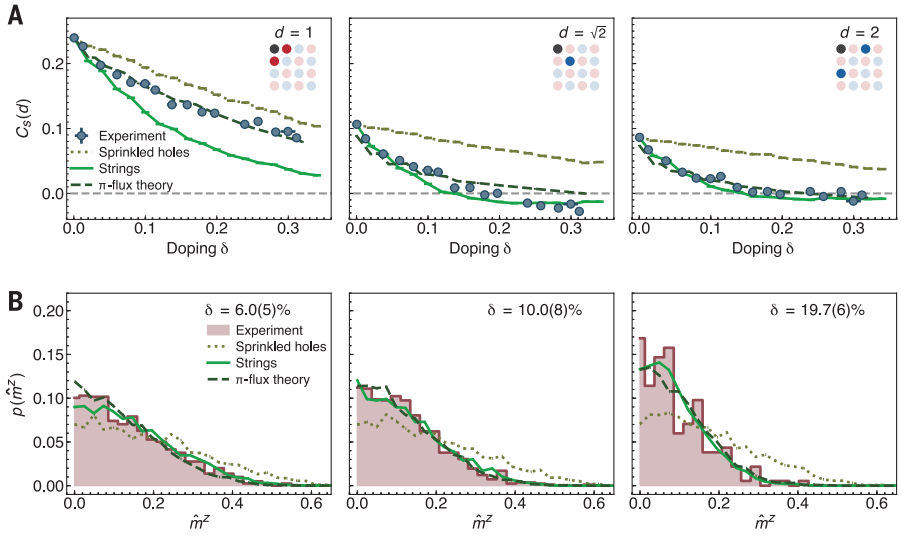



**Fig. 4. Spin correlations and staggered magnetization.** (**A**) Decay of nearest-neighbor (left), diagonal next-nearest-neighbor (center), and straight nearest-neighbor (right) spin-spin correlation functions upon doping. The π-flux theory most quantitatively explains $C_s(1)$, but only the string model captures the sign change of $C_s(\sqrt{2})$. In all three cases, sprinkled holes overestimate the spin correlations. Doping error bars are calculated as in (*14*), section 2.1; all other error bars represent 1 SEM. (**B**) Full counting statistics of the staggered magnetization for doping values of 6.0(5)% (left), 10.0(8)% (center), and 19.7(6)% (right). Both π-flux states and geometric strings show reasonable agreement, whereas sprinkled holes do not. The figure is based on more than 29,900 experimental realizations at average temperature $T = 0.65(4)J$.

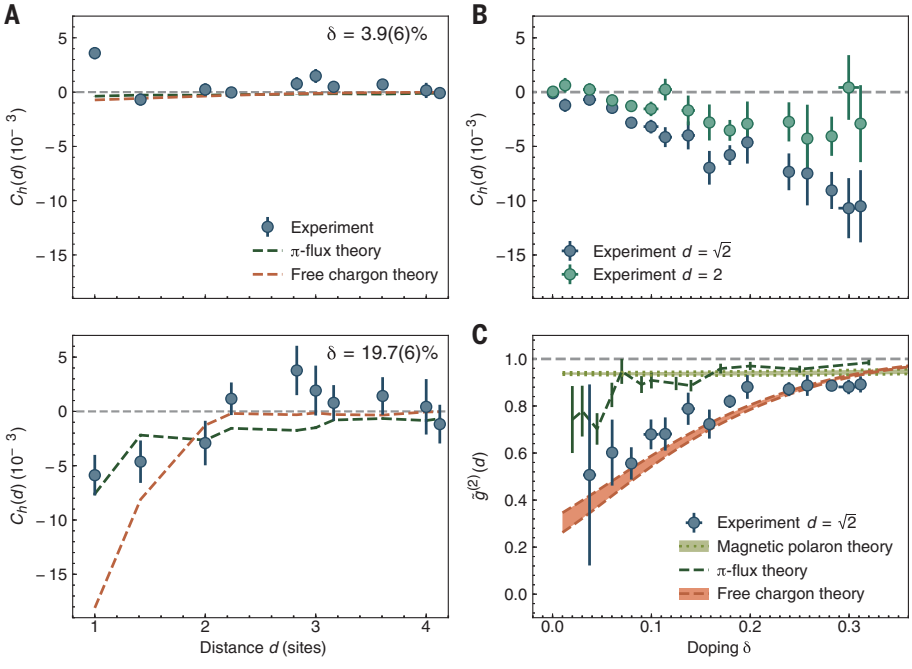



**Fig. 5. Observation of hole antibunching.** (**A**) Antimoment correlation function for weak (top) and strong (bottom) doping. The correlation functions are different up to a distance of $d = 2$. (**B**) Diagonal next-nearest neighbor and straight next-nearest neighbor antimoment correlators versus doping. At both distances negative correlations grow with doping. (**C**) Normalized antimoment correlator at $d = \sqrt{2}$ versus doping. The experimental result cannot be explained by the π-flux or a point-like magnetic polaron theory (see text), but instead matches a free fermionic chargon theory. In (B) and (C), error bars on the doping are calculated as in (*14*), section 2.1. All other error bars represent 1 SEM. The figure is based on more than 9900 experimental realizations at an average temperature $T = 0.65(4)J$.

hole-hole repulsion in this regime. Microscopically, such repulsive interactions can arise from the existence of a low-lying bound state of two holes (*31*). Here we do not consider geometric-string theory or sprinkled holes because both introduce uncorrelated holes by construction. Additionally, in the comparison to π-flux states, we do not include doublon-hole pairs to avoid unintended artifacts in the antimoment correlator. For reference, we plot the predicted hole-hole correlation function for a phenomenological model of spinless fermionic chargons with nearest-neighbor hopping of strength $t$ and temperatures between $0.5J$ and $0.7J$ (*32*). Here, strong anti-correlations result from Pauli repulsion between the fermionic chargons, but qualitatively similar behavior is expected for bosonic chargons with hard-core interactions. We find that both theories qualitatively describe the experimental result.

The emergence of this repelling behavior can be characterized by plotting the antimoment correlation as a function of doping for $d = \sqrt{2}$ and $d = 2$ (Fig. 5B). Beyond the intermediate doping regime, negative correlations appear at distances of $\sqrt{2}$ and 2, suggesting a growth of hole-hole repulsion with doping. Furthermore, the presence of antimoment correlations between sites of differing sublattices at $d = 1$ evidences against holes tunneling preferentially between sites of one sublattice, as predicted by theories of pointlike magnetic polarons with a dispersion minimum at $(\pi/2, \pi/2)$ in the Brillouin zone (*33*–*36*).

Finally, we plot a normalized $g^{(2)}(d = \sqrt{2})$ to account for the difference between doped holes and holes in doublon-hole pairs and quantify the relative fraction of doped holes that are anti-correlated:

$$\tilde{g}^{(2)}(|\mathbf{d}|) \equiv \frac{C_h(\mathbf{d})}{\delta^2} + 1 \qquad (5)$$

for doping δ (Fig. 5C). This rescaling allows direct comparison to the $g^{(2)}$ function for theories without doublon-hole pairs. The number of free holes is too small for doping below 5% to make statistically significant statements about the behavior of holes in this regime. In the geometric-string theory, we assume that chargons (dressed dopants) are completely uncorrelated with each other, but because of their fermionic statistics, Pauli blocking should actually introduce anti-correlations that have not yet been included in our analyses. We first consider a description of these chargons as pointlike magnetic polarons, where the known dispersion relation of the dressed hole (*37*) is used to define a tight-binding hopping model of the polaron. Figure 5C shows that our data are incompatible with this model, which predicts significantly weaker hole-hole anticorrelations. Similar behavior is predicted by the π-flux theory, which models the doped holes as pointlike objects moving in a quantum spin liquid of singlets.

Next, we examine a picture of free chargons, motivated by considering magnetic polarons with a finite extent that results from the spinon-chargon bound state predicted by geometric-string theory. At sufficiently large chargon density, or doping, the chargons are expected to interact, and their hard-core character will introduce anti-correlations. In this regime, geometric strings are also expected to overlap substantially and modify the dispersion relation of the chargons to be independent of spinons, yielding spinless chargons. We find that the experimental results demonstrate consistency with these free chargons, in agreement with earlier theoretical work in the strange-metal regime (*32*, *38*). This extension to the geometric string theory may be able to explain the deviations from experiment seen in other observables at high doping, but additional analyses are required.

## Conclusions and outlook

The string-pattern–based observables introduced here complement established observables such as correlation functions or full counting statistics. Across the observables considered, we find better agreement with experimental data between both the geometric-string theory and π-flux states, as compared to sprinkled holes.

At intermediate doping values, we find evidence for hole-hole repulsion. Although signatures of other phases such as stripe phases, incommensurate spin order, or nematic fluctuations have not yet been observed in this system, they are predicted to emerge at lower temperatures.

The ideas presented can be extended to other real-space patterns—for example, patterns that reflect the underlying physics of other candidate microscopic theories for the doped Hubbard model. Moreover, machine learning techniques could be used to directly compare sets of raw experimental atom distributions to theoretical models without the need for intermediate observables (*39*). This class of techniques is highly promising as quantum simulations of the Hubbard model continue to probe lower temperatures within the pseudogap and strange-metal phases, but can also be applied to spatially resolved studies of quenches across phase transitions (*40*), dynamical phase transitions (*41*), and higher-order scattering processes (*42*). Possible extensions of our work include systems with anisotropic spin interactions (*16*) or doped SU($N$) spin models (*43*).

## REFERENCES AND NOTES


1. E. H. Lieb, F. Y. Wu, *Phys. Rev. Lett.* **20**, 1445–1448 (1968).
2. F. Woynarovich, *J. Phys. Chem.* **15**, 85–96 (1982).
3. M. Ogata, H. Shiba, *Phys. Rev. B Condens. Matter* **41**, 2326–2338 (1990).
4. H. V. Kruis, I. P. McCulloch, Z. Nussinov, J. Zaanen, *Phys. Rev. B* **70**, 075109 (2004).
5. M. Endres *et al*., *Science* **334**, 200–203 (2011).
6. C. Gross, I. Bloch, *Science* **357**, 995–1001 (2017).
7. T. A. Hilker *et al*., *Science* **357**, 484–487 (2017).
8. V. J. Emery, *Phys. Rev. Lett.* **58**, 2794–2797 (1987).
9. P. A. Lee, N. Nagaosa, X.-G. Wen, *Rev. Mod. Phys.* **78**, 17–85 (2006).
10. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
11. A. Mazurenko *et al*., *Nature* **545**, 462–466 (2017).
12. P. W. Anderson, *Science* **235**, 1196–1198 (1987).
13. X.-G. Wen, P. A. Lee, *Phys. Rev. Lett.* **76**, 503–506 (1996).
14. See supplementary materials.
15. C. Gros, *Ann. Phys.* **189**, 53–88 (1989).
16. F. Grusdt *et al*., *Phys. Rev. X* **8**, 011046 (2018).
17. L. N. Bulaevskii, É. L. Nagaev, D. I. Khomskii, *J. Exp. Theor. Phys.* **27**, 836–838 (1968).
18. W. F. Brinkman, T. M. Rice, *Phys. Rev. B* **2**, 1324–1338 (1970).
19. P. Béran, D. Poilblanc, R. B. Laughlin, *Nucl. Phys. B* **473**, 707–720 (1996).
20. F. Grusdt, Z. Zhu, T. Shi, E. Demler, *SciPost Physics* **5**, 057 (2018).
21. M. F. Parsons *et al*., *Phys. Rev. Lett.* **114**, 213002 (2015).
22. M. F. Parsons *et al*., *Science* **353**, 1253–1256 (2016).
23. M. Boll *et al*., *Science* **353**, 1257–1260 (2016).
24. L. W. Cheuk *et al*., *Science* **353**, 1260–1264 (2016).
25. P. T. Brown *et al*., *Science* **357**, 1385–1388 (2017).
26. E. V. Gorelik *et al*., *Phys. Rev. A* **85**, 061602 (2012).
27. J. Koepsell *et al*., Imaging magnetic polarons in the doped Fermi-Hubbard model. arXiv:1811.06907 [cond-mat.quant-gas] (2018).
28. V. J. Emery, S. A. Kivelson, *Nature* **374**, 434–437 (1995).
29. S. R. White, D. J. Scalapino, *Phys. Rev. Lett.* **80**, 1272–1275 (1998).
30. J. Zaanen, O. Y. Osman, H. V. Kruis, Z. Nussinov, J. Tworzydlo, *Philos. Mag. B Phys. Condens. Matter Stat. Mech. Electron. Opt. Magn. Prop.* **81**, 1485–1531 (2001).
31. C. Chin, R. Grimm, P. S. Julienne, E. Tiesinga, *Rev. Mod. Phys.* **82**, 1225–1286 (2010).
32. R. K. Kaul, Y. B. Kim, S. Sachdev, T. Senthil, *Nat. Phys.* **4**, 28–31 (2007).
33. C. L. Kane, P. A. Lee, N. Read, *Phys. Rev. B Condens. Matter* **39**, 6880–6897 (1989).
34. S. Sachdev, *Phys. Rev. B Condens. Matter* **39**, 12232–12247 (1989).
35. G. Martinez, P. Horsch, *Phys. Rev. B Condens. Matter* **44**, 317–331 (1991).
36. J. Liu, X. Sun, D. L. Lin, T. F. Georgy, *J. Phys. Condens. Matter* **4**, 5301–5308 (1992).
37. M. Brunner, F. F. Assaad, A. Muramatsu, *Phys. Rev. B* **62**, 15480–15492 (2000).
38. S. Sachdev, D. Chowdhury, *Prog. Theor. Exp. Phys.* **2016**, 12C102 (2016).
39. A. Bohrdt *et al*., *Nat. Phys.* (2019).
40. H. Bernien *et al*., *Nature* **551**, 579–584 (2017).
41. J. Zhang *et al*., *Nature* **551**, 601–604 (2017).
42. L. Feng, J. Hu, L. W. Clark, C. Chin, *Science* **363**, 521–524 (2019).
43. C. Honerkamp, W. Hofstetter, *Phys. Rev. Lett.* **92**, 170403 (2004).
44. C. S. Chiu *et al*., Data for "String patterns in the doped Hubbard model." Harvard Dataverse (2019).



## ACKNOWLEDGMENTS

We thank M.Kanász-Nagy for Heisenberg QMC code. We thank A. Hébert, S. Sachdev, Z.-Y. Weng, and J. Zaanen for insightful discussions. **Funding:** We acknowledge support from AFOSR grant nos. FA9550-14-10035 and FA9550-16-10323; DoD NDSEG; the Gordon and Betty Moore Foundation EPIQS program and grant no. 6791; NSF GRFP and grant nos. PHY-1506203, PHY-1734011, and DMR-1308435; ONR grant no. N00014-18-1-2863; SNSF; Studienstiftung des deutschen Volkes; and the Technical University of Munich–Institute for Advanced Study, funded by the German Excellence Initiative and the European Union FP7 under grant agreement 291763, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC-2111–390814868, through DFG grant no. KN1254/1-1, and DFG TRR80 (Project F8). **Author contributions:** C.S.C., G.J., M.X., and D.G. performed the experiment. C.S.C. and A.B. curated the data and, along with F.G., developed the pattern finding algorithm. C.S.C., G.J., A.B., M.X., and D.G. analyzed the data. F.G. and E.D. developed the geometric-string theory. A.B. and F.G., together with M.K. and E.D., developed the theory simulations. M.G. and D.G. supervised the work. All authors contributed to the interpretation of the results and writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All experimentally measured site-resolved atom distributions and analysis code are available (*44*).


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/251/suppl/DC1
Supplementary Text
Figs. S1 to S10
References (*45*–*57*)

26 September 2018; accepted 5 June 2019
10.1126/science.aav3587

REPORT

METASURFACES

# Temporal color mixing and dynamic beam shaping with silicon metasurfaces

Aaron L. Holsteen[1], Ahmet Fatih Cihan[2], Mark L. Brongersma[1]*

**Metasurfaces offer the possibility to shape optical wavefronts with an ultracompact, planar form factor. However, most metasurfaces are static, and their optical functions are fixed after the fabrication process. Many modern optical systems require dynamic manipulation of light, and this is now driving the development of electrically reconfigurable metasurfaces. We can realize metasurfaces with fast (>$10^5$ hertz), electrically tunable pixels that offer complete (0- to 2π) phase control and large amplitude modulation of scattered waves through the microelectromechanical movement of silicon antenna arrays created in standard silicon-on-insulator technology. Our approach can be used to realize a platform technology that enables low-voltage operation of pixels for temporal color mixing and continuous, dynamic beam steering and light focusing.**

Many optical applications require active manipulation of light beams, including communication networks (*1*), beam steering (*2*), dynamic holography, imaging, and sensing (*3*). To facilitate their high-speed and low-power operation, it is highly desirable to reduce the size and weight of the dynamic optical components in such systems. For example, compact, deformable lenses are already available that can actively be tuned with electrowetting (*4*), mechanical (*5*), and thermal (*6*) approaches. Despite many advances, their tuning speeds have been limited to a few tens of hertz, and such nonplanar, refractive optical elements still take up quite a bit of space because of their principle of operation. Spatial light modulators (SLMs) using liquid crystals provide speeds higher by an order of magnitude. They are composed of a two-dimensional array of pixels that can locally control the phase imparted on light waves. This enables a greater versatility of dynamic optical functions, including active focusing, steering, and holography. However, the use of liquid crystals results in undesirable polarization-dependent behavior and limited numerical apertures. Subwavelength control over the phase with a new tuning approach is needed to further increase functionality, operating speed, and numerical apertures. For this reason, the prospect of realizing dynamic flat optics with mutable metasurfaces has stimulated a notable research effort (*7*–*18*). Tunable metasurface lenses and lens-doublets have already been realized and typically use strain or thermal (*19*) tuning of the entire metasurface (*10*, *20*) or mechanical movement between metasurface lenses (*21*, *22*). However, these approaches are not amenable to achieve arbitrary optical function, and movement of the metasurface as a whole tends to be slow. Ideally, one would realize active, individually addressable metasurface pixels from tunable optical antennas capable of continuously changing their light-scattering phase by a complete 2π and offering large amplitude modulation. A wide range of electrical, mechanical, chemical, phase-change, magnetic, and thermal external stimuli have been explored, but such performance has thus far been elusive. In

[1]Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA 94305-4045, USA. [2]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: brongersma@stanford.edu

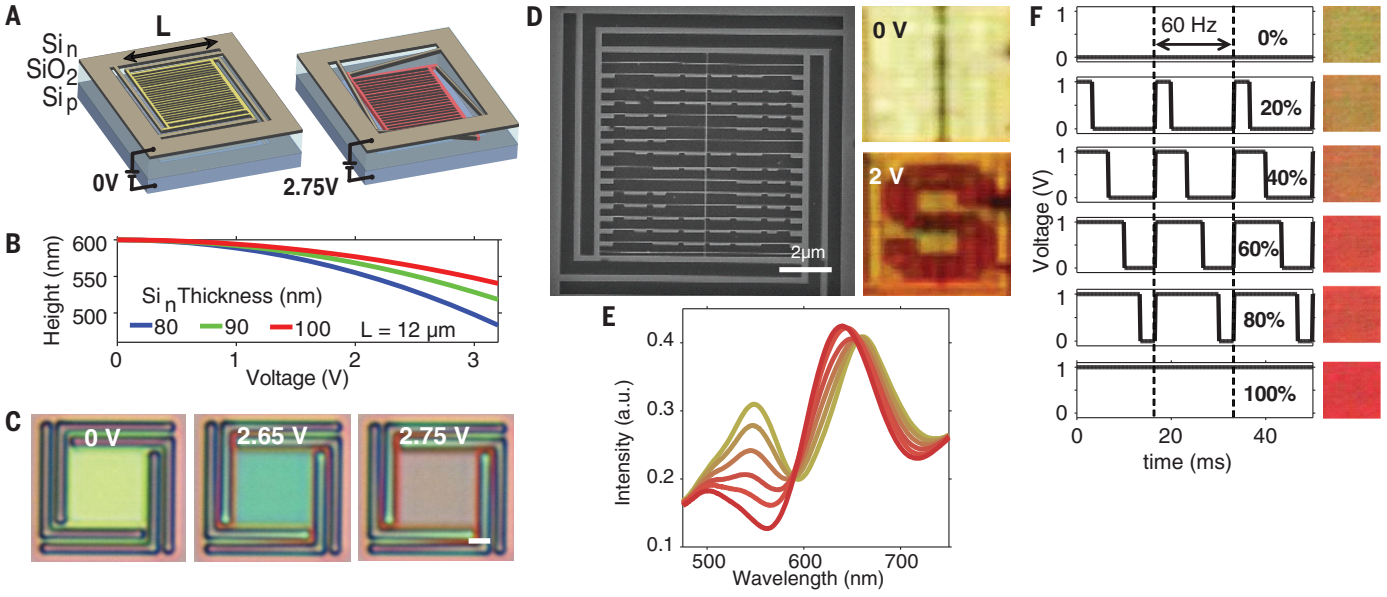



**Fig. 1. Temporal color control from suspended silicon metasurfaces.** (**A**) Schematic metasurface devices at 0-V (left) and a 2.75-V (right) bias. (**B**) Finite-element mechanical simulation for a pixel size of 12 μm. (**C**) Bright-field light scattering from a metasurface with 100-nm-wide nanowires spaced at 300 nm. (**D**) An active metasurface patterned with a fishbone structure to realize (left) structural color tuning of a character (letter "S") as shown in bright-field scattering images taken at biases of (top) 0 V and (bottom) 2 V. (**E** and **F**) Temporally mixed-color spectra and time-averaged color spectra and pixels. The data was taken by using different duty cycles in the voltage modulation. Scale bars, 2 μm.

this work, we realized such pixels by leveraging recent work that demonstrated broad spectral tuning of the structural color of individual Si antennas with micro-electrical mechanical systems (MEMS) technology (*23*).

In our metasurface devices, we use high-index silicon (Si) nanostructures. Such structures support strong Mie-type optical resonances that afford effective local control over light scattering (*24*). When arranged in subwavelength patterns, such resonators can create flat optical elements that can perform optical functions that are not achievable with conventional glass-based optical components (*25*–*27*). We used Si-on-insulator (SOI) technology to suspend different Mie resonator–based metasurfaces above Si substrates to allow for mechanical tuning of various optical functions (Fig. 1A). Finite-element mechanical simulations (COMSOL Multiphysics) show that the electrostatic, capacitive coupling between suspended array structures and Si base can be used to move ~100-nm-thick metasurface layers by hundreds of nanometers at low actuation voltages (Fig. 1B) (*28*). For a metasurface made from Si nanobeams (100 by 100 nm in cross section), one can use the coupling of the Mie resonances supported by beams and the Fabry-Pérot modes that exist between the substrate and the array to create vibrant colors in white light reflection. These colors can be tuned across the visible spectrum (Fig. 1C and fig. S1) (*29*).

Pixels with tunable structural-color patterns can also be created. A scanning electron microscopy (SEM) image of suspended fishbone structures with narrow, nonresonant support beams that hold larger Mie-resonant blocks is shown in Fig. 1D. These Mie resonant structures were designed to tune color with height and allow the appearance of a desired character. Temporal color mixing can also be achieved in pixels with distinct resonant scattering peaks in the yellow and red spectral ranges (Figs. 1, E and F). By applying 0- and 1-V bias voltages at different duty cycles, color can be mixed temporally. Such pixels could be arrayed to realize low-power, reflective displays.

The ability to dynamically control the phase of scattered light further expands the possibilities for our metasurfaces. We can control the phase of scattered light over a broad range by capitalizing on the interplay between the Mie and Fabry-Pérot resonances supported by this system (Fig. 2). This point is illustrated with simulations of the reflection phase pickup for 600-nm-wavelength transverse electric (TE)–polarized light reflected from a Si beam array with 100-nm-thick beams at different heights (fig. S2) (*29*). The Mie resonators can also be engineered to achieve very different dependencies of the reflection phase pickup on the resonator height. The relative scattering phase shifts for arrays with different beam widths can approach 2π just by moving the metasurface up and down above the Si substrate (Fig. 2B). This facilitates the design of actively tunable gradient metasurfaces capable of dynamic wavefront shaping. For the design of such metasurfaces, we assume that each nanowire

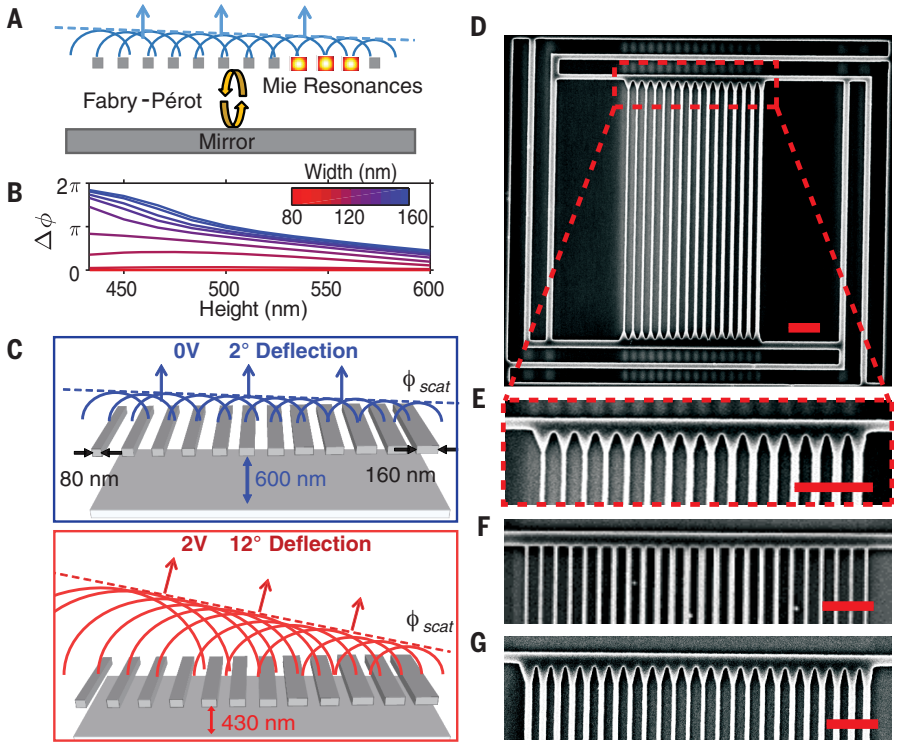



**Fig. 2. Operational principle of suspended Si metasurfaces.** (**A**) Resonance interaction between the Fabry-Pérot cavity and the nanowire Mie resonances afford spatial control over the reflection phase. (**B**) FDTD-simulated phase delay of TE-polarized light scattered from a suspended nanowire array with a periodicity of 300 nm and thickness of 100 nm relative to an array with 80-nm widths. (**C**) Schematic of an individual device at (top) 600 nm and (bottom) 430 nm array heights above the substrate, showing the scattered wave redirection akin to phased antenna arrays. (**D**) SEM of an entire beam-steering device. (**E** to **G**) Close ups of nanowire array designed to realize (E) beam steering, (F) a concave lens, and (G) a convex lens. Scale bars, 1 μm.

in a metasurface serves as an individual antenna whose height above the surface controls the local scattering phase of light. In this case, the realization of the metasurfaces with a gradient in their beam width affords continuous and active steering as the height of a metasurface is raised or lowered uniformly.

Shown in Fig. 2C is a schematic of a device composed of a periodic nanowire array with a thickness of 100 nm and widths linearly graded from 80 to 160 nm. This device can continuously steer 600 nm light with an electrical bias of a few volts (Fig. 3). Both beam steering and tunable cylindrical lenses can be made with width-graded metasurface gratings (Fig. 2, D to G).

Shown in Fig. 3, A and B, is the beam steering performance of light scattered from the active metasurface device shown in Fig. 2D. In order to remove the direct reflection from the substrate, we used crossed polarizers to examine light scattered from the active metasurface (fig. S3) (*29*). When this device is illuminated with 600-nm light, the diffracted beam can continuously be redirected from 2° to 12° off the sample normal by applying a bias of up to 3.2 V. The confocally collected beam profile immediately above the device is illustrated in Fig. 3C, clearly showing the tilt in the reflected light beam with applied bias. Fourier imaging of the far-field light scattering shows the steering performance in angular space (Fig. 3D). Smaller pixels with a steeper phase progression can be made within a single 100-nm SOI layer to achieve larger steering angles, although we find continuous beam steering requires 12 or more nanobeams (fig. S4) (*29*).

The possible application of dynamic metasurfaces critically depends on the achievable speeds. The response times of MEMS-based devices are generally dependent on the stiffness and mass of the suspended structure. For our metasurfaces, there are trade-offs between the achievable response time, the actuation distance, and driving voltage. To quantify the speed, we modulated the height of our metasurfaces with a sinusoidal bias and measured the ratio of the scattered power between the maximum and minimum deflection peaks (Fig. 3D). In a frequency sweep, the fundamental mode resonance of a typical metasurface device at 400 kHz is clearly visible (Fig. 3E). This frequency is consistent with the fundamental mechanical resonance mode as obtained through numerical simulation (fig. S5) (*29*). This demonstrates that these active metasurfaces can be operated in ambient conditions up to 1 MHz

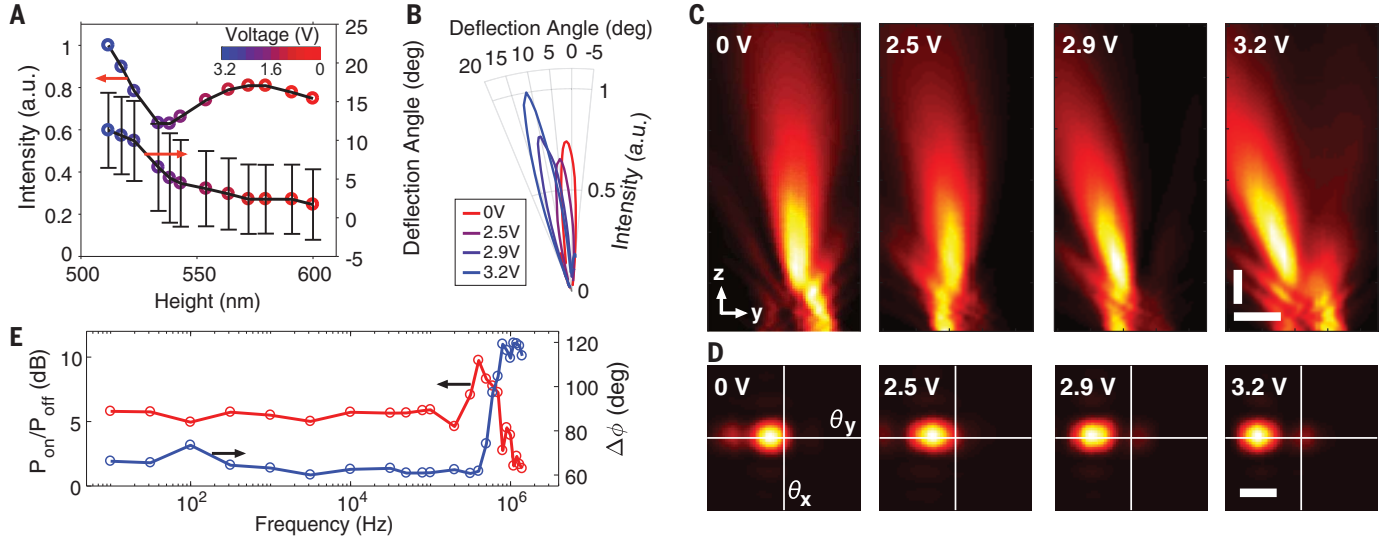



**Fig. 3. Beam steering from suspended Si metasurfaces.** (**A**) Measured intensity and beam steering from an active metasurface nanowire array. The error bars indicate the full width at half maximum of the beam steering shown in (**B**) the polar representation of the steered light collected in the Fourier plane. (**C**) Intensity profiles for the continuously steered beam above the active metasurface nanowire array. Scale bars, 3 μm. (**D**) Fourier images of the beam steering showing the angular deflection of the beam from the surface normal shown at the origin. Scale bar, 8°. (**E**) The frequency response of the beam steering device showing (red) the power ratio of light steered off the peak intensity point at 0 V and (blue) the relative phase between the output modulation and input driving voltage.

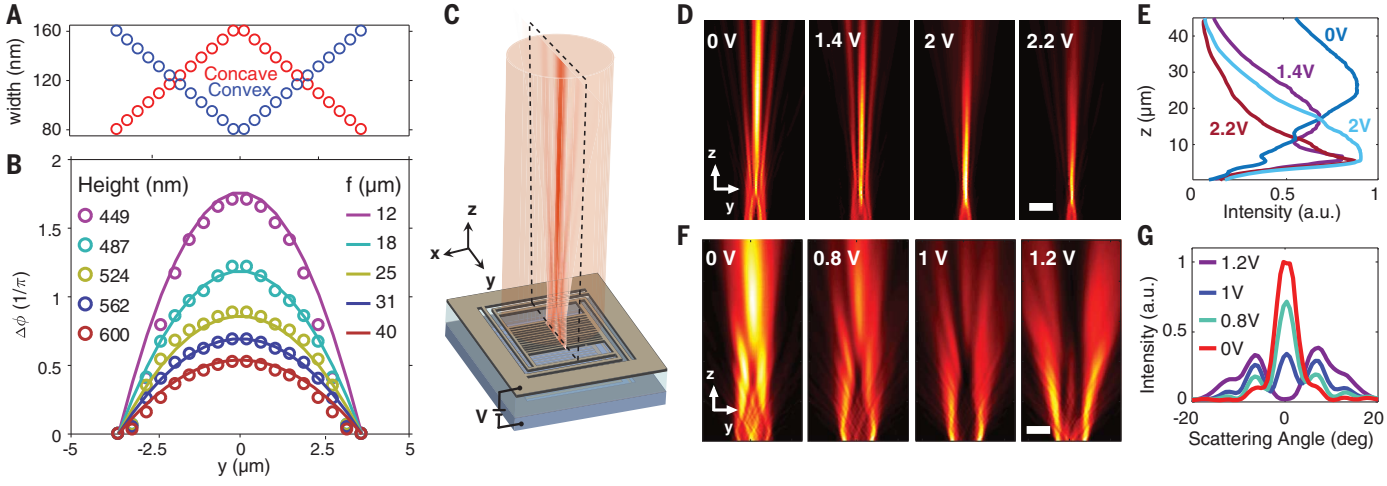



**Fig. 4. Tunable light focusing from suspended Si metasurfaces.** (**A**) Spatial arrangement of the nanowire widths in the concave (red) and convex (blue) active metasurface lenses. (**B**) Simulated relative scattering phase from the concave active metasurface lens at various heights with overlaid hyperbolic phase profiles corresponding to various lens focal lengths. (**C**) Schematic of metasurface lens showing the cross-sectional plane used in (D) and (F). (**D** and **F**) Cross sections along the *y*-*z* plane of the measured intensity above the (D) concave and (F) convex tunable active metasurface lenses. (**E**) Integrated intensity of the focal profiles along the *z* axis for the concave lens. (**G**) Angular scattering captured in the Fourier plane for the convex lens. Scale bars, 5 μm.

and have response times suitable for display applications, light detection and ranging (LIDAR), and optical switching for global optical interconnects (*30*). These devices are known to be very robust against cycling fatigue (*31*), and we have run these devices for more than $10^9$ cycles without noticeable degradation. The experimental setups for measuring the beam steering performance of the light scattered from the active metasurface are provided in fig. S3, A to C.

To demonstrate the flexibility in realizing different, dynamically tunable phase profiles, we also realized tunable lenses. By linearly grading the beam widths of a metasurface from 80 to 160 nm, the phase profile of a cylindrical lens can be generated. Both concave and convex lenses can be realized, depending on whether the beam width increases or decreases size toward center of the optical element (Fig. 4A). An easily achievable 300-nm spacing results in a 40-μm focal length *f*. Shown in Fig. 4, B and C, are the achievable hyperbolic phase profiles along the *y* direction given by $\phi(y) = \pi + \frac{2\pi}{\lambda}(f - \sqrt{y^2 + f})$ for various focal lengths *f*. Finite time domain (FDTD)–simulated scattering phases from the device at various suspended heights show that the lens can easily accommodate changes in focal length of several tens of micrometers (Fig. 4B).

When the intensity of 600 nm scattered light immediately above the active metasurface device

is measured with confocal microscopy, the focal profile from the tunable lens is evident. The lens focal length can be tuned experimentally from 26 to 5 μm by applying a bias of up to 2.2 V in a fabricated device (Fig. 4, D and E). If a convex phase profile is designed into the lens, light defocusing can be realized (Fig. 4, F and G). The far-field Fourier images captured from this device show that a normally reflected beam can be swept out to two side lobes peaked at scattering angles ±6° off the device normal. The light focusing results closely match the FDTD simulations for the performance of these tunable lenses (fig. S6) (*29*).

We have demonstrated that a mechanical degree of freedom can effectively be used to create a wide range of dynamically tunable optical functions. Key to achieving these functions is broad spectral, amplitude, and phase control over the scattered light with subwavelength spatial control. The compactness of the metasurface pixels also affords operation of these devices at speeds of up to 1 MHz. This makes their use viable for a range of emerging imaging, display, and remote sensing applications.

## REFERENCES AND NOTES


1. M. C. Wu, O. Solgaard, J. E. Ford, *J. Lightwave Technol.* **24**, 4433–4454 (2006).
2. B. Schwarz, *Nat. Photonics* **4**, 429–430 (2010).
3. I. W. Jung *et al.*, *Int. Conf. Opt. MEMS Nanophoton.* **4**, 238–239 (2012).
4. L. Li, D. Wang, C. Liu, Q.-H. Wang, *Opt. Express* **24**, 2931 (2016).
5. S. W. Lee, S. S. Lee, *Appl. Phys. Lett.* **90**, 2005–2008 (2007).
6. S. Y. Lee, H. W. Tung, W. C. Chen, W. Fang, *IEEE Photonics Technol. Lett.* **18**, 2191–2193 (2006).
7. J. Park, J. H. Kang, X. Liu, M. L. Brongersma, *Sci. Rep.* **5**, 15754 (2015).
8. Y. W. Huang *et al.*, *Nano Lett.* **16**, 5319–5325 (2016).
9. J. Park, J. H. Kang, S. J. Kim, X. Liu, M. L. Brongersma, *Nano Lett.* **17**, 407–413 (2017).
10. S. M. Kamali, E. Arbabi, A. Arbabi, Y. Horie, A. Faraon, *Laser Photonics Rev.* **10**, 1002–1008 (2016).
11. L. Zhu, J. Kapraun, J. Ferrara, C. J. Chang-Hasnain, *Optica* **2**, 255 (2015).
12. A. Komar *et al.*, *Appl. Phys. Lett.* **110**, 071109 (2017).
13. B. Gholipour, J. Zhang, K. F. MacDonald, D. W. Hewak, N. I. Zheludev, *Adv. Mater.* **25**, 3050–3054 (2013).
14. C. Zou *et al.*, *ACS Nano* **10**, 133–141 (2015).
15. Z. Zhu, P. G. Evans, R. F. Haglund Jr., J. G. Valentine, *Nano Lett.* **17**, 4881–4885 (2017).
16. N. Meinzer, W. L. Barnes, I. R. Hooper, *Nat. Photonics* **8**, 889–898 (2014).
17. D. Wang *et al.*, *Sci. Rep.* **5**, 15020 (2015).
18. Q. Wang *et al.*, *Nat. Photonics* **10**, 60–65 (2015).
19. A. Afridi *et al.*, *ACS Photonics* **5**, 4497–4503 (2018).
20. H. S. Ee, R. Agarwal, *Nano Lett.* **16**, 2818–2823 (2016).
21. E. Arbabi *et al.*, *Nat. Commun.* **9**, 812 (2018).
22. T. Roy *et al.*, *APL Photonics* **3**, 021302 (2018).
23. A. L. Holsteen, S. Raza, P. Fan, P. G. Kik, M. L. Brongersma, *Science* **358**, 1407–1410 (2017).
24. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, B. Luk'yanchuk, *Science* **354**, aag2472 (2016).
25. M. Khorasaninejad, F. Capasso, *Science* **358**, eaam8100 (2017).
26. D. Lin *et al.*, *Nano Lett.* **16**, 7671–7676 (2016).
27. A. Arbabi, Y. Horie, M. Bagheri, A. Faraon, *Nat. Nanotechnol.* **10**, 937–943 (2015).
28. COMSOL, COMSOL Multiphysics (COMSOL, 2018).
29. Materials and methods are available as supplementary materials.
30. S. Han, T. J. Seok, N. Quack, B.-W. Yoo, M. C. Wu, *Optica* **2**, 370 (2015).
31. C. L. Muhlstein, S. B. Brown, R. O. Ritchie, *J. Microelectromech. Syst.* **10**, 593–600 (2001).


## ACKNOWLEDGMENTS

We thank P. G. Kik for valuable discussions. **Funding:** This research was supported by the U.S. Air Force Office for Scientific Research grant FA9550-14-1-0389 and the California Metaphotonics Cluster, funded by Samsung Advanced Institute of Technology (SAIT). A.L.H. also acknowledges a National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. **Author contributions:** A.L.H. and M.L.B. conceived the concept of the paper. A.L.H. and A.F.C. fabricated the devices. A.L.H. conducted the optical measurements and simulations. All authors worked on the manuscript. M.L.B supervised the project. **Competing interests:** M.L.B. and A.L.H. are inventors on U.S. patent application US20180299743A1 held and submitted by Stanford University that covers MEMS-actuated high-index optical antennas and metasurfaces for light manipulation and control. The authors declare no other competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary material.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/257/suppl/DC1
Materials and Methods
Figs. S1 to S6

5 April 2019; accepted 20 June 2019
10.1126/science.aax5961

VOLCANOLOGY

# Millennial storage of near-Moho magma

**Euan J. F. Mutch*, John Maclennan, Tim J. B. Holland, Iris Buisman**

**The lower crust plays a critical role in the processing of mantle melts and the triggering of volcanic eruptions by supply of magma from greater depth. Our understanding of the deeper parts of magmatic systems is obscured by overprinting of deep signals by shallow processes. We provide a direct estimate of magma residence time in basaltic systems of the deep crust by studying ultramafic nodules from the Borgarhraun eruption in Iceland. Modeling of chromium–aluminum interdiffusion in spinel crystals provides a record of long-term magmatic storage on the order of 1000 years. This places firm constraints on the total crustal residence time of mantle-derived magmas and has important implications for modeling the growth and evolution of transcrustal magmatic systems.**

Understanding the temporal evolution of magmatic systems from micrometer to crustal scales has become a major focus in volcanology (*1*), driven by important advances in the use of crystal chemistry as a magmatic chronometer (*2*, *3*). Compositional zonation in volcanic crystals has been integrated with diffusion models to show that the time scales of days to years preserved in the crystal record can be linked directly to volcanic monitoring data (*4*–*7*). These short time scales are conceptually linked to magma mixing, remobilization of crystal mush, and magma rise from storage regions before eruption. However, the time scales of magma storage and the behavior of magma plumbing systems during periods of apparent quiescence are not well understood. Estimates of magma and mineral residence times in the crust have ranged from $10^2$ to $10^5$ years (*8*). Parsing these time scales into different crustal levels has proven difficult but is key to understanding transcrustal magmatic systems. Storage time scales in shallow silicic systems can span the aforementioned range in near-solidus conditions of cold storage (*8*, *9*) or in an eruptible high–melt-fraction state (*10*). By contrast, little is known about storage times in the mafic systems that dominate magmatism at mid-ocean ridges, ocean islands, and in the deeper part of arc volcanoes. Direct storage times in the lower crust are notoriously difficult to estimate by diffusion chronometry as shallow and mid-crustal processing obscure lower-crustal signals through crystal equilibration and resorption. We estimate magma residence time in the lower crust by applying diffusion chronometry to observations of Cr number (Cr#) [Cr# = Cr/(Al + Cr)] zoning in spinels contained in ultramafic nodules from the Borgarhraun eruption in Iceland. This eruption provides constraints on magmatic processes operating close to the crust-mantle boundary, called the Moho.

Borgarhraun is an early postglacial (10.5 to 7 thousand years before the present) primitive basalt flow from the Theistareykir System in the Northern Volcanic Zone of Iceland (*11*). Geothermobarometry and melt inclusion studies have shown that the magma crystallized near the Moho (depth ~24 km) from a compositionally diverse set of primary mantle melts (*11*–*13*). Fresh tephra from this eruption contains primitive crystals of olivine ($Fo_{87–92}$, where Fo is forsterite content in mol. %), clinopyroxene, Cr spinel, and plagioclase ($An_{80–90}$, where An is anorthite content in mol. %). Mafic and ultramafic cumulate nodules are abundant with textural characteristics typical of storage and maturation in a crystal mush, such as grain rounding and curved crystal edges (*14*). Wehrlitic nodules show olivine and Cr spinel crystals enclosed in clinopyroxene (Fig. 1), a feature observed in mafic plutonic rocks from layered intrusions (*14*), oceanic sections (*15*), and ophiolites (*16*). The clinopyroxene-hosted olivine and spinel crystals

Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK.
*Corresponding author. Email: ejfm2@cam.ac.uk or em0242@my.bristol.ac.uk

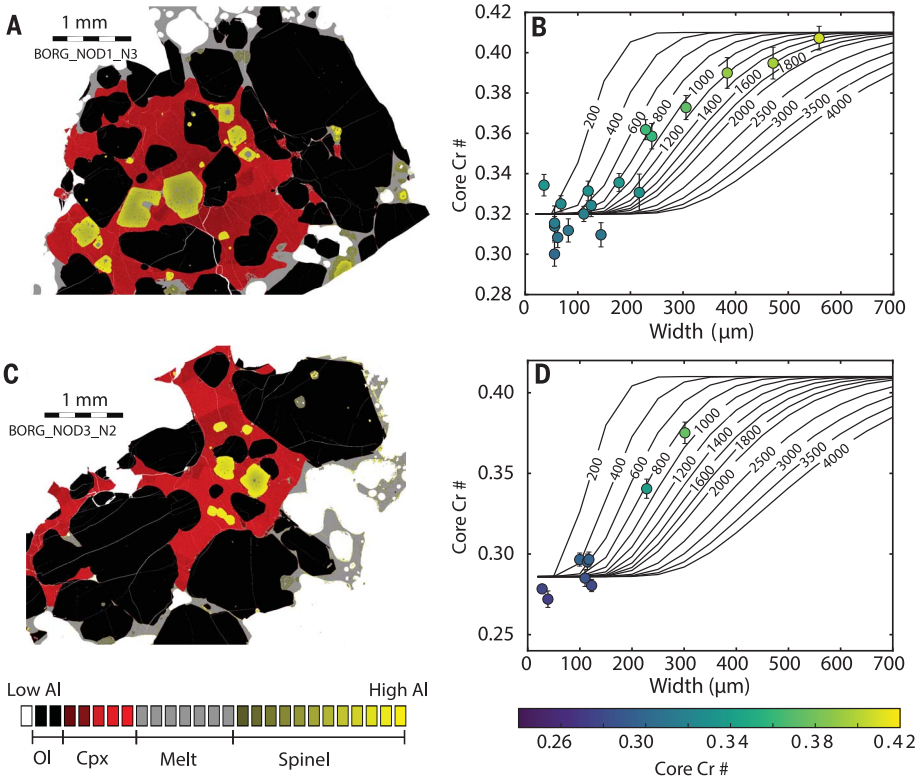



**Fig. 1. Disequilibrium in cumulate nodules from beneath the Icelandic Moho.** (**A** and **C**) Aluminum maps of Borgarhraun wehrlitic nodules with poikilitic texture and (**B** and **D**) accompanying plots that show spinel core Cr# compositions with respect to their measured width. Al maps are colored based on phase: yellow-gray corresponds to Cr spinel (Sp), red corresponds to clinopyroxene (Cpx), black corresponds to olivine (Ol), and gray corresponds to glass (Gl). The black curves are 3D spherical spinel diffusion models (modeled at 1215°C) conducted for different crystal radii and diffusion time scales (years). Error bars correspond to the standard deviation of multiple compositions sampled from each spinel core. (A) and (B) correspond to sample BORG_NOD1_N3 and (C) and (D) to BORG_NOD3_N2.

are associated with small melt pockets (Fig. 1) and are surrounded by thin layers of melt (~2 to 10 μm thick). We measured spinel widths of 20 to 600 μm. Spinels with widths below 200 μm have homogeneous Cr# contents that vary between different nodules. Crystals larger than 200 μm are compositionally zoned with rims similar in composition to the small spinels and cores of higher Cr#. The Cr# of these spinel cores scales with spinel size (Fig. 1). The eradication of compositional heterogeneity in the smaller spinels and the preservation of zonation in the larger spinels are likely to be controlled by a diffusive process. Smaller crystals completely equilibrated with their local environment, whereas larger

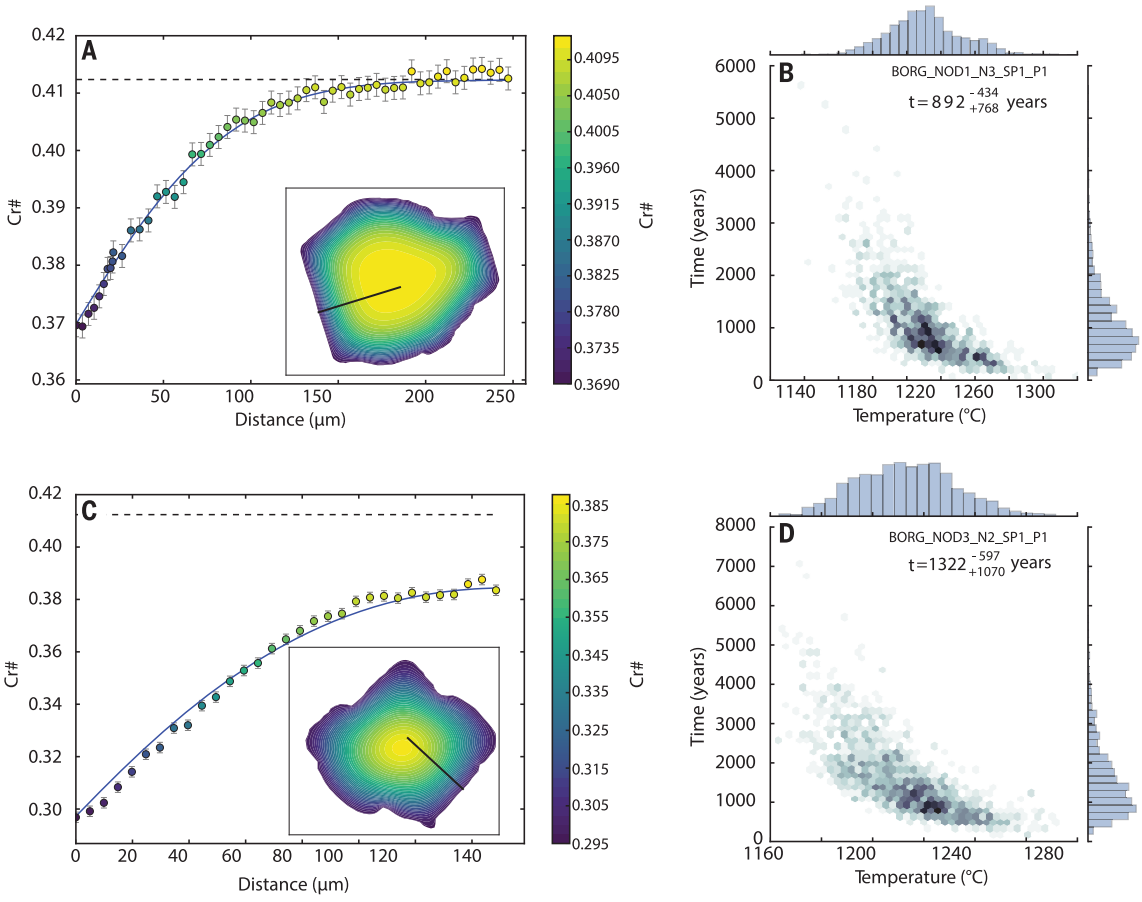



**Fig. 2. Timing diffusive equilibration of Borgarhraun spinel crystals.** (**A** and **C**) 1D Cr# profiles measured by electron microprobe microanalysis across Borgarhraun spinels. Blue curves show the modeled 1D fit from the 2D finite element diffusion model (insets). Dashed lines are model initial conditions. The error bars are 1σ standard deviations based on analytical uncertainties. (**B** and **D**) Temperature-time density plots for posterior distributions from the Bayesian inversion. The quoted time is the median with 1σ uncertainties. (A) and (B) correspond to profile BORG_NOD1_N3_SP1_P1 and (C) and (D) to BORG_NOD3_N2_SP1_P1.

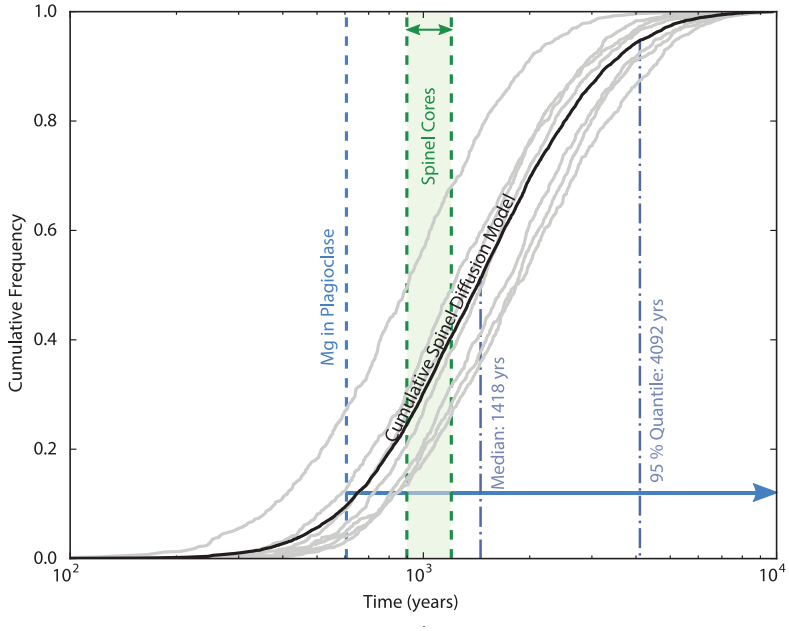



**Fig. 3. Storage time estimates of near-Moho magma.** Cumulative frequency distributions showing the magma storage time estimates from the Bayesian inversion. Gray lines are distributions obtained by modeling individual spinel crystals. The black line is the cumulative frequency distribution of all of the modeled spinel crystals. The dashed blue line and the blue arrow are the minimum and the range of possible residence times based on Mg equilibration in plagioclase at 1215°C, respectively. The light-green region represents mush residence times based on bulk 3D spherical diffusion models shown in Fig. 1.

crystals preserved higher Cr# core compositions in a state of disequilibrium. We observed this effect in a large population of spinels (~100) across seven different nodules, which rules out the role of sectioning effects (figs. S1 and S2).

We expect the crystallization of the spinels and the surrounding clinopyroxene to take place within 1 year of magma emplacement (*17*). We used thermodynamic modeling to account for the spinel Cr# compositions by equilibration over a cooling interval of ~75°C (figs. S3 and S4), similar to the crystallization interval of the Borgarhraun olivines (1350 to 1215°C) (*18*). The spinels were initially encased in clinopyroxene at a high temperature and subsequently equilibrated with the surrounding melt pockets and grain-boundary films following cooling to a lower temperature (1215°C) (fig. S5). We infer very little deformation of the spinels as they retained their equidimensional forms, which precludes stress-directed lattice diffusion of Cr and Al to create the concentric zoning patterns that we observed (*19*).

We cannot resolve the geometry, extent, and interconnectivity of the melt network from two-dimensional (2D) sections of the nodules. However, previous studies have suggested melt pockets and interconnected grain-boundary liquid films play an important role in the final stages of crystallization of large mafic intrusions (*14*, *20*). The melt networks may also have been more extensive before clinopyroxene crystallization, given the important role of the liquid in mush homogenization (*21*). The composition (e.g., Mg# and Cr#) of the melt pockets adjacent to the spinel crystals was different from that of the final carrier liquid (fig. S6), which suggests that these melt networks were isolated from the open magma body. The variability in spinel Cr# between different nodules highlights heterogeneity in the mush and the development of microenvironments for spinel equilibration.

We used the FEniCS (*22*) finite element code to model 2D diffusion of Cr–Al exchange in spinels, combined with a nested sampling Bayesian inversion (*23*) to constrain the timescales of mush storage with robust estimates of uncertainty. We derived diffusivity equations for Cr–Al exchange in Cr spinel that take into account the covariance between the underlying diffusive parameters (*24*, *25*) (figs. S7 and S8 and table S1). We ran the models using Gaussian prior distributions with 1σ uncertainties for temperature (1215 ± 30°C), pressure (0.8 ± 0.14 GPa) (*12*, *18*), and ferric iron contents ($Fe^{3+}/Fe_{total}$) of the system (0.14 ± 0.02) (*26*). We used a multivariate Gaussian prior distribution for the parameters of the Cr–Al interdiffusion coefficient as constrained by its covariance matrix (table S2). We used the equilibration temperature of spinel rim compositions (1215°C) that corresponded to a cooler crystal mush. We assumed the same initial core Cr# of 0.41 for all crystals on the basis of the largest measurable spinel that was most likely to retain its original core composition (Fig. 2). This provided us with a maximum estimate of the storage time scale.

We applied our Bayesian inversion to seven spinel crystals (figs. S9 to S19 and table S3) from three different nodules that converged to magma storage time scales with a median of 1400 years, with 95% of timescales being less than 4100 years (Fig. 3). These storage time scales we determined agreed with 3D spherical spinel diffusion models we conducted at 1215°C for different crystal radii to assess the time scales of diffusive equilibration for spinels of different sizes (Fig. 1). The Mg compositions of Borgarhraun plagioclase macrocryst cores have equilibrated with the mush conditions (figs. S20 to S31). This means the coarsest of these crystals (1200-μm radius) required ~570 years of storage, assuming 3D diffusion in a spherical crystal at 1215°C and a core composition of $An_{90}$ (fig. S32). The time scale of Mg in plagioclase equilibration provides a minimum estimate of the storage time, consistent

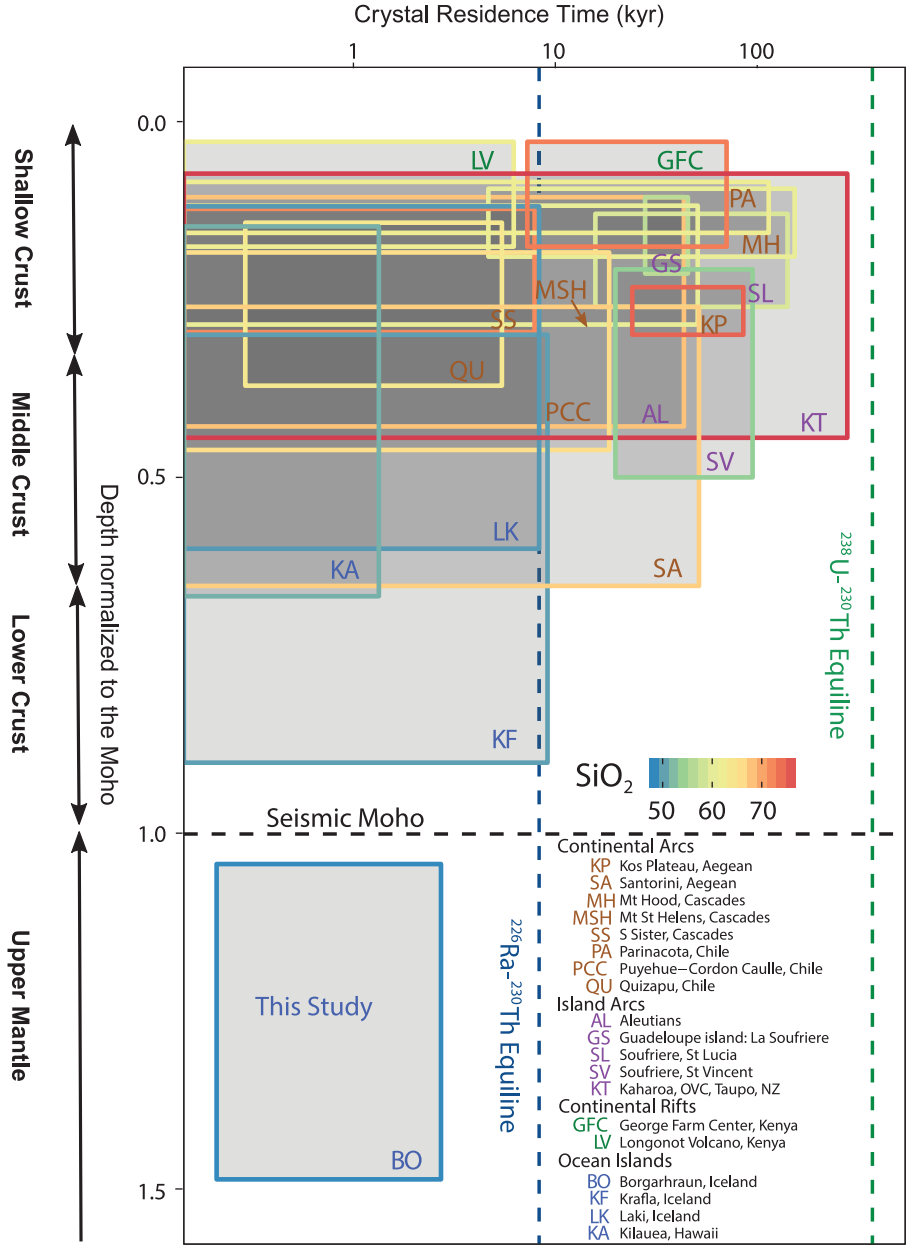



**Fig. 4. Crystal residence times and storage depths for magmatic systems from different tectonic settings.** Shaded boxes represent the range of storage depth (normalized to the Moho) and residence time scale estimates made for individual volcanoes or volcanic systems. Time scale estimates for all eruptions except for Borgarhraun used U-series disequilibria, with residence times being calculated by using the difference between crystallization age and eruption age. The color of the box edges corresponds to the average $SiO_2$ content of the whole rock sample used in the time scale estimates. The color of the symbols corresponds to the tectonic setting. Data and references are in table S4. kyr, thousand years.

with our spinel models that require hundreds to thousands of years of storage.

$^{238}U$-$^{230}Th$ and $^{226}Ra$-$^{230}Th$ crystal residence ages from diverse tectonic settings range from $10^2$ to $10^5$ years (Fig. 4) (*8*). Discrepancies between these estimates and thermally activated diffusion time scales show that magmatic crystals may have complex and protracted histories throughout the crust that are largely independent of the melts that formed them and carried them to the surface for eruption (*8*). This highlights the difficulty in unpicking individual segments of magmatic evolution from bulk crystal ages integrated over zoned crystal populations. Overall, crystal residence time scales in basaltic systems have previously been estimated as hundreds to thousands of years by combining oxygen diffusion durations with U-series observations (*27*, *28*), but these measurements do not provide a spatial context for the depth of magma residence. In contrast, textural and petrological constraints from Borgarhraun unambiguously link the diffusion time scale with near-Moho storage. Olivine-hosted melt inclusions link the crystal cargo to the carrier liquid by concurrent mixing and crystallization (*11*), meaning our results can be interpreted as magmatic storage times. Our diffusion chronometry approach therefore provides an opportunity for relative high temporal and spatial resolution of magmatic storage times in mafic lower-crustal systems (Fig. 4).

Our storage time estimate for basaltic mushes in the lower crust have implications for understanding the time scale of magma transfer through the entire crustal column at a range of tectonic settings. This includes not only mid-ocean ridges and ocean islands (e.g., Iceland and Hawaii) but also oceanic and continental arc settings, where the lower crust is dominated by mafic and ultramafic magmatism, as shown by exposed crustal sections (e.g., the Kohistan arc in Pakistan) (*29*), primitive erupted products (*5*, *30*), and geophysical observations (*31*). There are no direct constraints on the magmatic storage times in the lower crust of arcs. We therefore propose that our results provide an initial guide to expected near-Moho storage times in other mafic systems. For ocean island settings, our results indicate that primary magma can be stored near the Moho for thousands of years after initial mixing and crystallization (*17*). These magmas are rapidly transported over days to the mid- or shallow crust (*7*), where there is further long-term storage and magma processing (*27*, *28*). This impacts our understanding of mantle melt transport, which is based on $^{226}Ra$ excesses, and may lead to downward revision of melting region porosities when millennial storage of magma in the crust is considered (*32*).

Crystal storage in shallow, silicic arc systems can take place over time scales of up to $10^5$ years, which is likely related to different thermal (*1*) and rheological properties when compared with basaltic systems. If the shorter processing times we have found apply to arc systems, then the upper crust may be episodically supplied with relatively young magma batches from the lower mafic crust (Fig. 4). The shorter crystal and magma storage times in the lower crust might reconcile differences between melt production rates in the mantle and the frequency of large silicic eruptions (*1*). In each tectonic setting, punctuated magma storage and rapid transport is likely a prominent feature. Although packets of magma fresh from the mantle melting region could take thousands of years to traverse the crust, storage could accommodate most of this time. This style of magma transfer will play a major role in the movement of heat and volatiles through the crust and controlling the connection between deep and surface geochemical reservoirs, such as the mantle and atmosphere.

## REFERENCES AND NOTES


1. K. V. Cashman, R. S. J. Sparks, J. D. Blundy, *Science* **355**, eaag3055 (2017).
2. F. Costa, R. Dohmen, S. Chakraborty, *Rev. Mineral. Geochem.* **69**, 545–594 (2008).
3. K. M. Cooper, *Philos. Trans. A Math. Phys. Eng. Sci.* **377**, 20180009 (2019).
4. K. Saunders, J. Blundy, R. Dohmen, K. Cashman, *Science* **336**, 1023–1027 (2012).
5. P. Ruprecht, T. Plank, *Nature* **500**, 68–72 (2013).
6. T. H. Druitt, F. Costa, E. Deloule, M. Dungan, B. Scaillet, *Nature* **482**, 77–80 (2012).
7. E. J. F. Mutch, J. Maclennan, O. Shorttle, M. Edmonds, J. F. Rudge, *Nat. Geosci.* **12**, 569–574 (2019).
8. K. M. Cooper, A. J. Kent, *Nature* **506**, 480–483 (2014).
9. A. E. Rubin *et al.*, *Science* **356**, 1154–1156 (2017).
10. M. Barboni *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13959–13964 (2016).
11. J. MacLennan *et al.*, *Geochem. Geophys. Geosyst.* **4**, 8624 (2003).
12. B. Winpenny, J. Maclennan, *J. Petrol.* **52**, 1791–1812 (2011).
13. D. A. Neave, K. D. Putirka, *Am. Mineral.* **102**, 777–794 (2017).
14. M. B. Holness *et al.*, *J. Petrol.* **48**, 1243–1264 (2007).
15. K. M. Gillis *et al.*, *Nature* **505**, 204–207 (2014).
16. J. Korenaga, P. B. Kelemen, *J. Geophys. Res. Solid Earth* **102**, 27729–27749 (1997).
17. J. Maclennan, *J. Petrol.* **49**, 1931–1953 (2008).
18. J. MacLennan, D. McKenzie, F. Hilton, K. Gronvöld, N. Shimizu, *J. Geophys. Res. Solid Earth* **108**, 2007 (2003).
19. K. Ozawa, *Nature* **338**, 141–144 (1989).
20. W. Meurer, M. Meurer, *Contrib. Mineral. Petrol.* **151**, 187–201 (2006).
21. A. Thomson, J. Maclennan, *J. Petrol.* **54**, 745–768 (2012).
22. M. S. Alnæs *et al.*, *Archive of Numerical Software* **3**, 9–23 (2015).
23. F. Feroz, M. Hobson, M. Bridges, *Mon. Not. R. Astron. Soc.* **398**, 1601–1614 (2009).
24. A. Suzuki, A. Yasuda, K. Ozawa, *Phys. Chem. Miner.* **35**, 433–445 (2008).
25. E. S. Posner, J. Ganguly, R. Hervig, *Geochim. Cosmochim. Acta* **175**, 20–35 (2016).
26. O. Shorttle *et al.*, *Earth Planet. Sci. Lett.* **427**, 272–285 (2015).
27. I. N. Bindeman, O. Sigmarsson, J. Eiler, *Earth Planet. Sci. Lett.* **245**, 245–259 (2006).
28. K. M. Cooper, K. W. Sims, J. M. Eiler, N. Banerjee, *Contrib. Mineral. Petrol.* **171**, 54 (2016).
29. O. Jagoutz, M. W. Schmidt, *Chem. Geol.* **298–299**, 79–96 (2012).
30. C. C. Stamper, J. D. Blundy, R. J. Arculus, E. Melekhova, *J. Petrol.* **55**, 1353–1387 (2014).
31. T. Parsons *et al.*, *Geology* **26**, 199 (1998).
32. K. Sims *et al.*, *Geochim. Cosmochim. Acta* **63**, 4119–4138 (1999).


## ACKNOWLEDGMENTS


We thank C. Richardson for helpful advice about FEniCS and M. Edmonds and D. Neave for their comments on an early draft of this manuscript. **Funding:** This research was funded by a NERC studentship awarded to E.J.F.M. (NE/L002507/1). **Author contributions:** The project was initially conceived by E.J.F.M. and J.M. E.J.F.M. developed the diffusion model. E.J.F.M. performed the EPMA and SEM analysis with help from I.B. I.B. conducted the QEMSCAN analysis and processing of QEMSCAN maps. T.J.B.H. conducted the thermodynamic modeling using THERMOCALC. E.J.F.M. did the diffusion modeling. E.J.F.M. and J.M. interpreted the results. E.J.F.M. wrote the first draft of the manuscript, which was revised by all authors. **Competing interests:** The authors declare that they have no competing financial interests. **Data and materials availability:** All data collected and modeled as part of this study are included in the supplementary materials and supplementary data files. The diffusion model code and Bayesian inversion in their current form are available from https://github.com/EuanMutch/Spinel_DFENS.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/260/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S34
Tables S1 to S4
References (*33*–*95*)


21 March 2019; accepted 24 June 2019
10.1126/science.aax4092

MAGNETIC MATERIALS

# Reconfigurable ferromagnetic liquid droplets

Xubo Liu[1,2], Noah Kent[2,3], Alejandro Ceballos[4], Robert Streubel[2], Yufeng Jiang[2,5], Yu Chai[2,4,6], Paul Y. Kim[2], Joe Forth[2], Frances Hellman[2,7], Shaowei Shi[1], Dong Wang[1,8], Brett A. Helms[2,6], Paul D. Ashby[2,6], Peter Fischer[2,3], Thomas P. Russell[1,9,10]*

**Solid ferromagnetic materials are rigid in shape and cannot be reconfigured. Ferrofluids, although reconfigurable, are paramagnetic at room temperature and lose their magnetization when the applied magnetic field is removed. Here, we show a reversible paramagnetic-to-ferromagnetic transformation of ferrofluid droplets by the jamming of a monolayer of magnetic nanoparticles assembled at the water-oil interface. These ferromagnetic liquid droplets exhibit a finite coercivity and remanent magnetization. They can be easily reconfigured into different shapes while preserving the magnetic properties of solid ferromagnets with classic north-south dipole interactions. Their translational and rotational motions can be actuated remotely and precisely by an external magnetic field, inspiring studies on active matter, energy-dissipative assemblies, and programmable liquid constructs.**

Ferromagnetic materials are generally solids with a fixed shape. Reconfigurable magnetic materials are known, such as ferrofluids [dispersions of magnetic nanoparticles (MNPs) in carrier fluids], but they are paramagnetic and lose magnetization once the external magnetic field is removed (*1*, *2*). Ferrofluids exhibit interesting properties and have found use, for example, as magnetic seals, but their inability to retain magnetization limits their broader application. The transformation of a ferrofluid into a ferromagnetic material can be realized by lowering the temperature or increasing the viscosity, at which Brownian motion of the MNPs is suppressed. Here, we show a simple means to effect this transformation by the in situ formation and interfacial jamming of MNP-surfactants.

We immersed an aqueous dispersion of carboxylated 22-nm-diameter MNPs ($Fe_3O_4$-$CO_2H$) in a solution of amine-modified polyhedral oligomeric silsesquioxane (POSS-$NH_2$) in toluene. The POSS-$NH_2$, itself a surfactant, assembles at the interface and electrostatically interacts with the MNPs, anchoring a well-defined number of POSS-$NH_2$ to the MNPs, converting the MNPs into MNP-surfactants. When the droplet shape changes, the interfacial area increases, and additional MNP-surfactants form and assemble at the interface. The droplet proceeds to reshape itself to minimize the interfacial area and, thereby, the free energy of the system, but the MNP-surfactants are compressed and jam, locking in the deformed shape (*3*, *4*) while remaining magnetized even without an external field.

Magnetic hysteresis loops of $Fe_3O_4$-$CO_2H$ ferrofluid droplets (Fig. 1A), measured by a vibrating sample magnetometer, show a saturation magnetization ($M_s$) that depends on the total number of MNPs in the droplets, as well as a vanishing coercive field ($H_c$) and remanent magnetization ($M_r$). By adding POSS-$NH_2$ ligands to the toluene, MNP-surfactants form at the interface. Increasing the concentration of the MNPs in the droplet or decreasing pH increases the coverage of the interface by MNP-surfactants, reducing the interfacial tension (fig. S1, A and B). With sufficient surface coverage, the MNP-surfactants jam, and the ferrofluid droplet transforms into a ferromagnetic liquid droplet. The magnetic hysteresis loops of identical ferrofluid droplets with and without the jammed interfacial assemblies of MNP-surfactants are shown in Fig. 1A. For both, $M_s$ is the same because the total number of MNPs is identical, but $M_r$ ~$1.89 \times 10^{-8}$ A $m^2$ and $H_c$ ~7.2 kA $m^{-1}$ for the ferromagnetic liquid droplet, demonstrating their ferromagnetic character. The jammed, interfacial assemblies of the MNP-surfactants are disordered and have a mechanical rigidity that suppresses thermal fluctuations characteristic of isolated MNPs. The jammed MNPs no longer freely rotate. The spatial separation between adjacent MNP-surfactants is <5 nm, which, combined with the orientation of the dipole magnetization in the MNPs, enhances the thermal stability of the magnetization and transforms the droplet surface into a ferromagnetic layer, similar to a fixed assembly of MNPs (*5*, *6*). When the field is removed, the moment of the ferromagnetic liquid droplet remains until the droplet is exposed to a field exceeding the switching field, whereupon the droplet is remagnetized. Reshaping the droplet by other external fields or reducing the binding energy of the MNP-surfactants will also unjam the MNP-surfactants ($M_r$ and $H_c$ vanish), providing further routes to control the magnetization. This ability to manipulate the magnetization further distinguishes ferromagnetic liquid droplets from ferrofluids and common ferromagnetic materials.

If the MNP-surfactant assembly is not jammed, no hysteresis is observed (Fig. 1B). To produce droplets with an unjammed assembly, the surface coverage of the droplets with the MNP-surfactants is varied from ~7 to ~20% (fig. S1C) by changing the concentrations of the MNPs and POSS-$NH_2$ and the pH. In Fig. 1B, variations in $M_s$ arise from differences in the total number of MNPs in each droplet. With full MNP-surfactant coverage—for example, single droplets of [$Fe_3O_4$-$CO_2H$ MNPs] = 0.5 g $L^{-1}$ at pH 4.5 in toluene containing [POSS-$NH_2$] = 1.0 g $L^{-1}$—the interfacial assembly jams, and a typical ferromagnetic hysteresis loop is seen (Fig. 1C). Hysteresis loops were measured for single droplets with different volumes, and the same droplets sonicated into numerous smaller droplets. This preserves the total volume (summed over all droplets) of the MNP dispersions but increases the surface-to-volume ratio ($S/V$) by two orders of magnitude (fig. S2D). $M_s$ and $M_r$ scale linearly with the total volume (total number of MNPs), whereas $H_c$ remains constant (Fig. 1C). For a given total volume, $M_r$ is independent of $S/V$ with largely varying droplet sizes. The mean separation distance between the dispersed MNPs is ~350 nm, too large for dipolar coupling. For comparison, discs that are 100 nm in diameter and 10 nm thick and have much larger saturation magnetizations are completely uncorrelated when the separation distance is >60 nm (*7*). The MNPs dispersed in the droplet freely diffuse, yet a strong coupling and correlation of the dispersed MNPs to those jammed at the interface is evident, and the liquid droplets behave like solid magnets. Furthermore, the ratio of $M_r/M_s$ for the ferromagnetic liquid droplets is 0.25, independent of droplet volume, which is the same as that for frozen ferrofluids at 4.5 K and fixed assemblies of $Fe_3O_4$ MNPs (*8*–*10*). Consequently, ferromagnetic liquid droplets have a similar energy barrier to overcome during magnetization reversal

[1]Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Physics Department, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. [4]Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. [5]Department of Applied Science and Technology, University of California, Berkeley, Berkeley, CA 94720, USA. [6]The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]Physics Department, University of California, Berkeley, Berkeley, CA 94720, USA. [8]State Key Laboratory of Organic–Inorganic Composites, Beijing University of Chemical Technology, Beijing 100029, China. [9]Polymer Science and Engineering Department, University of Massachusetts, Amherst, MA 01003, USA. [10]WPI–Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan.
*Corresponding author. Email: tom.p.russell@gmail.com

as their frozen or solid counterparts. Therefore, ferromagnetic liquid droplets have the magnetic properties of a solid.

All-liquid printing (*11*–*13*) and microfluidics (*14*) were used to produce ferromagnetic liquid cylinders with a 2:1 aspect ratio (fig. S3). A nonmagnetized ferromagnetic liquid cylinder was transferred to a toluene/$CCl_4$ density gradient in which the cylinder descended until buoyant (fig. S4, A and B). The axis of the cylinder and an insulated solenoid were aligned (Fig. 2A), and a magnetic field of 1 to ~2 kA $m^{-1}$ (Fig. 2B) applied to the solenoid pulled the ferromagnetic liquid cylinder into the solenoid. The cylinder reached the solenoid in 30 s at a speed of ~1.1 × $10^{-4}$ m $s^{-1}$ and stopped after fully entering the solenoid because of the absence of drag forces.

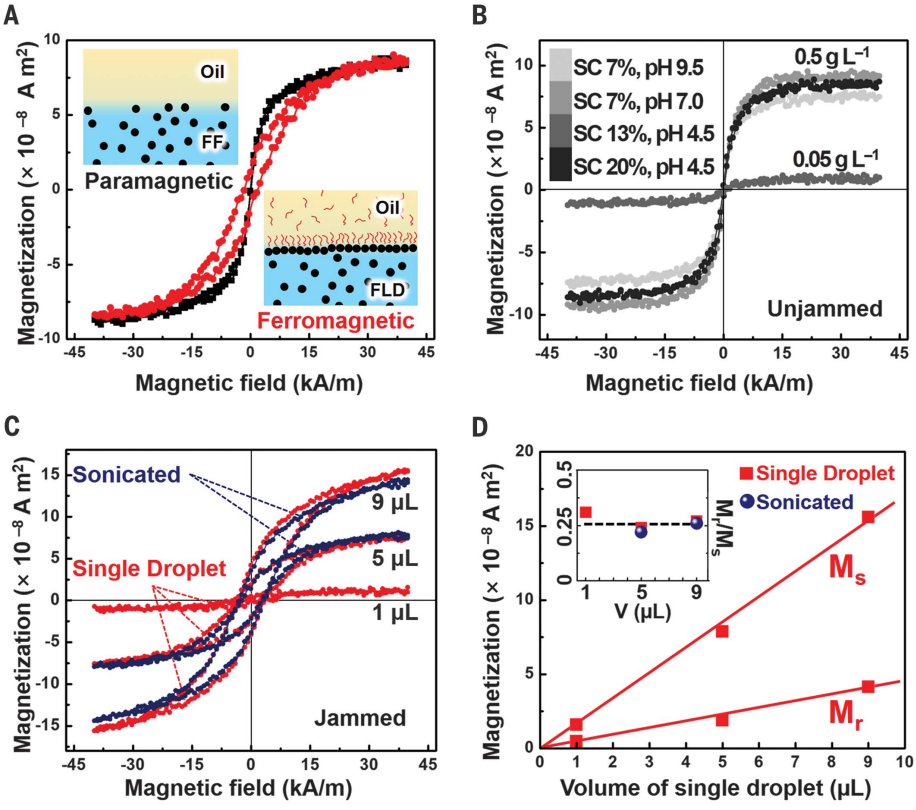



**Fig. 1. Tunable transformation of a paramagnetic FF into an FLD by the interfacial jamming of MNP-surfactants.** (**A**) Magnetic hysteresis loops of droplets with (red line) and without (black line) an interfacial layer of jammed MNP-surfactants measured with a vibrating sample magnetometer. Two schematics of aqueous FF and FLD, containing $Fe_3O_4$-$CO_2H$ MNPs (0.5 g $L^{-1}$) at pH 4.5, immersed in toluene without and with POSS-$NH_2$ ligands (1.0 g $L^{-1}$). (**B**) Hysteresis loops of individual 5-µL aqueous droplets, with 0.5 g $L^{-1}$ and 0.05 g $L^{-1}$ of $Fe_3O_4$-$CO_2H$ MNPs at different pH, immersed in 0.01 g $L^{-1}$ ligand solution. Surface coverage (SC) of droplets is ~7 to ~20% where MNPS assemblies are not jammed. (**C**) Hysteresis loops of single, jammed aqueous droplets with 0.5 g $L^{-1}$ of MNPs at pH 4.5 immersed in a 1.0 g $L^{-1}$ solution of POSS-$NH_2$ in toluene and hysteresis loops of the same system after being sonicated (figs. S1 and S2). (**D**) $M_r$ and $M_s$ of the droplets as a function of droplet volume. In the inset, the remanence ratio $M_r/M_s$ as a function of initial droplet volume (single droplet or droplet sonicated into multiple smaller droplets) remains constant at ~0.25. FF, ferrofluid; FLD, ferromagnetic liquid droplet.

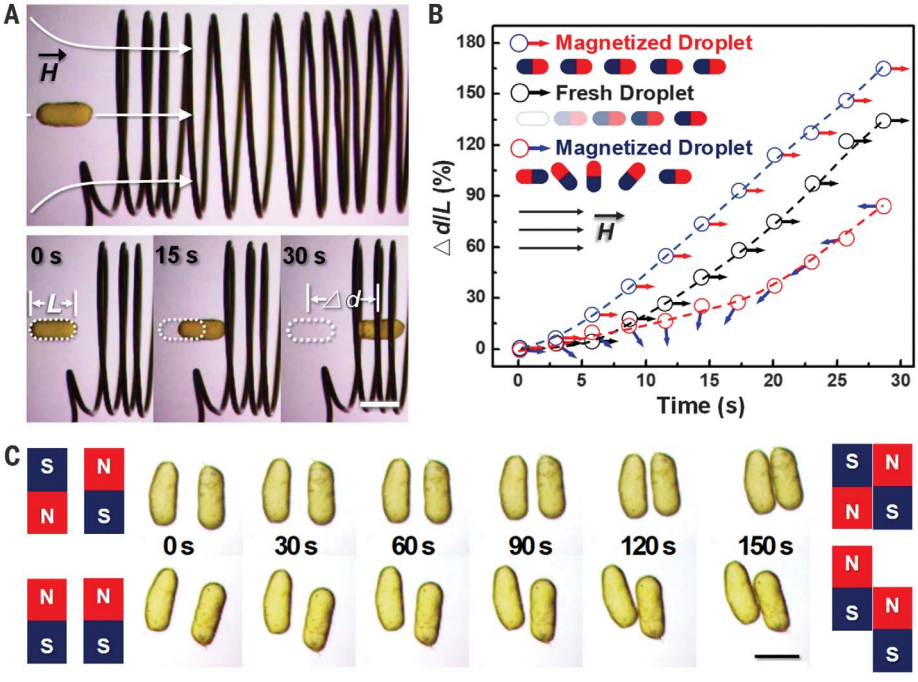



**Fig. 2. Manipulating FLDs with magnetic dipole interactions.** (**A**) A magnetized liquid cylinder is attracted by a magnetic field gradient, generated by the aluminum solenoid. (**B**) Displacement as a function of time for fresh, nonmagnetized liquid cylinder (black arrow), magnetized liquid cylinder with north pole facing the coil (red arrow), and magnetized liquid bar with south pole facing the coil (blue arrow). The ultimate velocity is determined by the field gradient and magnetic moment. The arrows indicate the orientation of the cylinder relative to the initial orientation. (**C**) Dipole interactions, N–S attraction, N–N and S–S repulsion, between two magnetized liquid cylinders. Scale bars, 2 mm.

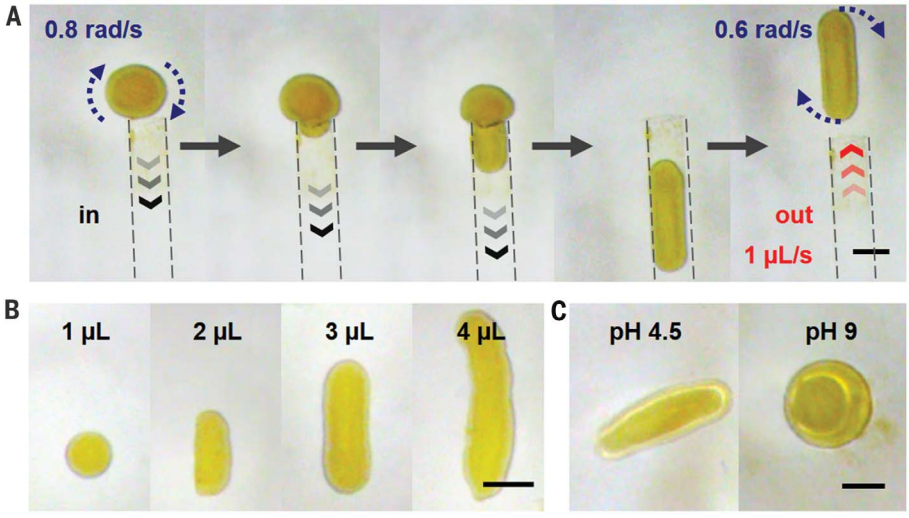



**Fig. 3. Top view of deforming an FLD.** (**A**) Reshaping spherical droplet into the cylinder by mechanically molding it with the glass capillary channel. (**B**) Droplets of different aspect ratios can be formed by using droplets with different volumes. (**C**) Reconfiguration of interfacial jamming and unjamming of MNP-surfactants by tuning the pH (for example, 4.5 and 9, respectively) of aqueous solution in the FLD cylinder. Scale bars, 1 mm.

The displacement, $\Delta d$ (in units of the droplet length, $L$), allows comparison of different ferromagnetic liquid cylinders because the drag force varies linearly with length, which for a fixed cylinder radius corresponds to the volume and $M_s$. Figure 2A shows the location and direction of the magnetic moment of the ferromagnetic liquid cylinder. The cylinder, magnetized only by the solenoid, moved inside the solenoid (movie S1). This ferromagnetic liquid cylinder (now magnetized) was repositioned outside the solenoid, preserving the south-north (S–N) pole orientation. With the same solenoid magnetic field strength (1 to ~2 kA $m^{-1}$), the cylinder now accelerates to the solenoid (Fig. 2B and movie S2). By reversing the field direction of the solenoid (movie S3), the magnetic moment of the ferromagnetic liquid cylinder and the solenoid field are in antiparallel alignment and should repel each other. However, initially a slight attraction is seen because of the free MNPs in the cylinder core; then, the cylinder rotates, aligning the moment of the jammed MNP-surfactants with the solenoid field, and is drawn into the solenoid (Fig. 2B). Considering the low velocity ($v_{max}$ ~1.1 × $10^{-4}$ m $s^{-1}$) and corresponding low Reynolds number ($Re$ ~0.16), the velocity of the ferromagnetic liquid cylinders can be expressed approximately as a function of time: $v(t) = C\frac{ma}{b}(1 - e^{-bt/m})$, where $C$ is a dimensionless constant based on $Re$ and the shape of a liquid droplet, $a$ is the acceleration from the drag force ($a = F_d / m$), $m$ is the mass, and $b$ is the coefficient of viscous friction. The constant velocity at longer times, independent of the initial magnetization configuration (Fig. 2B), implies a constant solenoid field gradient (due to the same $M_s$). Deviations from the linear relation in the early stage originate from viscous drag, the varying field gradient outside the solenoid, and the distinct initial magnetization configuration of the cylinder. Consequently, ferromagnetic liquid cylinders behave like solid magnets with N–N, S–S, and N–S dipole interactions. Figure 2C shows that magnetized liquid cylinders, initially separated by 1 mm, attract each other by such N–S dipole interactions (movie S4).

A distinguishing feature of the ferromagnetic liquid droplets is reconfigurability. A 3-μL spherical ferromagnetic liquid droplet (1.4-mm diameter) was drawn into a 1-mm-diameter glass capillary and then rapidly (after several seconds) ejected (Fig. 3, A and B, and movie S5). This transformed the spherical droplet into a cylinder (aspect ratio of 3:1). The interfacial area increased by 2.5 times, allowing more MNP-surfactants to form and jam, preserving the cylindrical shape. The ferromagnetic liquid droplet retained its ferromagnetic character, as evidenced by the rotation (~0.6 rad $s^{-1}$) in response to a rotating permanent magnetic field. The cylinder rotation is slightly less than that of the spherical droplet (~0.8 rad $s^{-1}$) because of the higher viscous drag force on the cylinder. The shape change can be reversed by tuning the binding energy, as shown in Fig. 3C, in which the pH was increased from 4.5 to 9, allowing the MNP-surfactants to unjam and the droplet shape to revert to spherical. Magnetization is lost, but by decreasing the pH, the MNP-surfactants rejam and the droplet transforms back to a ferromagnetic liquid droplet. Thus, the shape and magnetic state of the ferromagnetic liquid droplets are responsive.

The necessity of the magnetic coupling between the interfacially jammed MNP-surfactants and the dispersed MNPs to generate ferromagnetic liquid droplets is shown by changing the nature of the jammed interfacial assembly. Two sets of ferrofluid droplets at a pH of 4.5 were placed in a mixture of toluene and tetrachloride carbon of equal density containing POSS-$NH_2$ ligands: (i) ferrofluid droplets with carboxyl-functionalized 30-nm (22-nm core) $Fe_3O_4$ MNPs (0.5 g $L^{-1}$) that form ferromagnetic liquid droplets and (ii) a mixture of nonmagnetic sodium carboxymethyl cellulose (CMC-$CO_2Na$) (0.5 g $L^{-1}$) with 30-nm $Fe_3O_4$–polyethylene glycol (PEG) (nonfunctionalized) MNPs (0.5 g $L^{-1}$), in which only the CMC interacts with the POSS-$NH_2$. In both, nanoparticles jam at the interface (Fig. 4A), but in the first case they are ferromagnetic, whereas in the latter they are not ferromagnetic. Ferrofluid droplets of only PEG-functionalized $Fe_3O_4$ MNPs (0.5 g $L^{-1}$) were also placed in the oil. The total number of MNPs in all the droplets was constant. A bar magnet attracts all droplets (Fig. 4A) because all droplets have a ferromagnetic core, but the ferromagnetic liquid droplets are attracted much more strongly (Fig. 4A and movie S6). As shown in Fig. 4B, using a rotating magnet, the spherical ferromagnetic liquids rotate, whereas the unjammed ferrofluid and CMC-jammed ferrofluid droplets do not. The ferromagnetic liquid droplets are also attracted to the center of the magnet, and a dynamically stable pattern forms, balancing a hydrodynamic repulsion against a magnetic attraction, similar to that observed for elastomer discs (*15*, *16*) or ferrofluids (*17*, *18*) containing MNPs. Similar behavior is seen with ferromagnetic liquid cylinders in which the vortex flow in the oil is visualized with an oil-soluble dye (Fig. 4C and movie S7). The separation distance between the ferromagnetic liquid droplets depends on the rotation velocity (fig. S5 and movie S8), as expected. The entire patterned assembly of droplets also rotates in response to the rotating field. Upon stopping the rotating magnet, the ferromagnetic liquid cylinders align along the external field direction (movie S9). Absent of a dipole moment, the droplets with and without the jammed CMC monolayer do not spin and move only in a Brownian manner (movie S10). Therefore, the ferromagnetic liquid droplets can be easily separated, rotated in a controlled manner, and patterned (shown in Fig. 4B), affording a simple strategy for sorting and spatially arranging the ferromagnetic liquid droplets.

In conclusion, we have demonstrated the transformation of a ferrofluid to a ferromagnetic liquid droplet by the interfacial jamming and magnetization of MNP-surfactants. Ferromagnetic liquid droplets have the fluid characteristics of liquids but the magnetic properties of solids. They can be reconfigured while preserving their magnetic properties, and the attractive/repulsive interactions between ferromagnetic liquid droplets can be manipulated. Separation and patterning of ferromagnetic liquid droplets are easily achieved. The formation of ferromagnetic liquid droplets is reversible, and the interfacial assembly of the MNP-surfactants is responsive to external stimuli and provides systems in which translational and rotational motions can be actuated remotely and precisely by an external magnetic field.

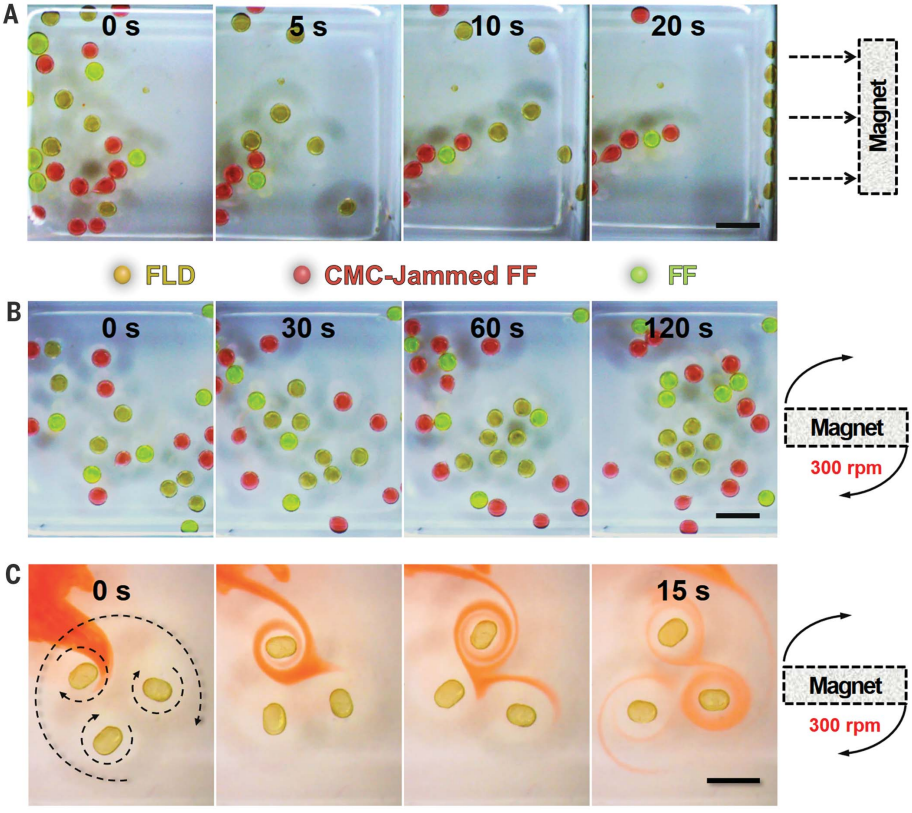



**Fig. 4. Sorting FLDs by using static and rotational magnetic fields.** Mixtures of FLDs comprising a shell of a jammed monolayer of MNP-surfactants and an FF core (dispersed carboxyl-functionalized iron oxide NPs, $Fe_3O_4$-$CO_2H$) (brownish spheres); FF droplet comprising a shell of jammed, nonmagnetic CMC surfactants (CMCSs) encapsulating an aqueous dispersion of PEG-coated iron oxide NPs ($Fe_3O_4$-PEG) (red spheres); FF droplet comprising an aqueous dispersion of $Fe_3O_4$-PEG NPs with no jammed monolayer at the droplet surface (bright green spheres). The separation of FLDs by using (**A**) a static bar magnet on the side of the container and (**B**) a bar magnet rotating under the container. [Rhodamine B in CMC-jammed FF] = 1 g $L^{-1}$, [fluorescein sodium salt in FF] = 1 g $L^{-1}$. (**C**) Visualization of the hydrodynamic vortex flow for an FLD ensemble in a rotating magnetic field using an oil-soluble dye. [Nile red in toluene] = 1 g $L^{-1}$. Scale bars, 3 mm.

## REFERENCES AND NOTES


1. S. Odenbach, *MRS Bull.* **38**, 921–924 (2013).
2. R. E. Rosensweig, "Magnetic fluids" in *Ferrohydrodynamics* (Courier Corporation, 2013), pp. 34–44.
3. M. Cui, T. Emrick, T. P. Russell, *Science* **342**, 460–463 (2013).
4. Z. Zhang *et al.*, *Sci. Adv.* **4**, eaap8045 (2018).
5. A. H. Lu, E. L. Salabas, F. Schüth, *Angew. Chem. Int. Ed.* **46**, 1222–1244 (2007).
6. J. J. Benkoski *et al.*, *J. Am. Chem. Soc.* **129**, 6291–6297 (2007).
7. R. Streubel *et al.*, *Nano Lett.* **18**, 7428–7434 (2018).
8. E. C. Stoner, E. P. Wohlfarth, *Philos. Trans. R. Soc. London A* **240**, 599–642 (1948).
9. W. Luo, S. R. Nagel, T. F. Rosenbaum, R. E. Rosensweig, *Phys. Rev. Lett.* **67**, 2721–2724 (1991).
10. G. F. Goya, T. S. Berquó, F. C. Fonseca, M. P. Morales, *J. Appl. Phys.* **94**, 3520–3528 (2003).
11. X. Liu *et al.*, *Angew. Chem. Int. Ed.* **56**, 12594–12598 (2017).
12. J. Forth *et al.*, *Adv. Mater.* **30**, e1707603 (2018).
13. S. Shi *et al.*, *Adv. Mater.* **30**, 1705800 (2018).
14. A. Toor, S. Lamb, B. A. Helms, T. P. Russell, *ACS Nano* **12**, 2365–2372 (2018).
15. L. Derr, *Proc. Am. Acad. Arts Sci.* **44**, 525–528 (1909).
16. B. A. Grzybowski, H. A. Stone, G. M. Whitesides, *Nature* **405**, 1033–1036 (2000).
17. W. Wang, J. Giltinan, S. Zakharchenko, M. Sitti, *Sci. Adv.* **3**, e1602522 (2017).
18. J. V. I. Timonen, M. Latikka, L. Leibler, R. H. A. Ras, O. Ikkala, *Science* **341**, 253–257 (2013).


## ACKNOWLEDGMENTS


**Funding:** This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division (contract no. DE-AC02-05-CH11231) in the Adaptive Interfacial Assemblies Toward Structuring Liquids program (KCTR16). R.S., A.C., N.K., F.H, and P.F. acknowledge support from U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division (contract no. DE-AC02-05-CH11231) in the Non-equilibrium Magnetic Materials program (MSMAG). Work at the Molecular Foundry (AFM imaging) was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy (contract no. DE-AC02-05CH11231). S.S. was supported by the Beijing NSF (2194083). X.L. was supported by the Beijing Advanced Innovation Center for Soft Matter Science and Engineering at Beijing University of Chemical Technology and China Scholarship Council. **Author contributions:** X.L., N.K., R.S., B.A.H., P.D.A., P.F., and T.P.R. made contributions to the conception and design of the experiments. X.L., N.K., A.C., R.S., Y.J., Y.C., J.F., F.H., S.S., and D.W. performed and supported the experiments. X.L., N.K., R.S., P.Y.K., B.A.H., P.D.A., P.F., and T.P.R. interpreted the data and wrote the manuscript. **Competing interests:** P.D.A. is also associated with Scuba Probe Technologies. **Data and materials availability:** All data are available in the main text or the supplementary materials.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/264/suppl/DC1
Materials and Methods
Figs. S1 to S5
Table S1
Movies S1 to S10
References (*19*, *20*)


31 January 2019; accepted 24 May 2019
10.1126/science.aaw8719

QUANTUM GASES

# Quantum scale anomaly and spatial coherence in a 2D Fermi superfluid

**Puneet A. Murthy[1]*†‡, Nicolò Defenu[2]*†, Luca Bayha[1], Marvin Holten[1], Philipp M. Preiss[1], Tilman Enss[2], Selim Jochim[1]**

**Quantum anomalies are violations of classical scaling symmetries caused by divergences that appear in the quantization of certain classical theories. Although they play a prominent role in the quantum field theoretical description of many-body systems, their influence on experimental observables is difficult to discern. In this study, we discovered a distinctive manifestation of a quantum anomaly in the momentum-space dynamics of a two-dimensional (2D) Fermi superfluid of ultracold atoms. The measured pair momentum distributions of the superfluid during a breathing mode cycle exhibit a scaling violation in the strongly interacting regime. We found that the power-law exponents that characterize long-range phase correlations in the system are modified by the quantum anomaly, emphasizing the influence of this effect on the critical properties of 2D superfluids.**

Symmetries and their violations are fundamental concepts in physics. A prominent type is conformal symmetry, which gives rise to the peculiar effect of scale invariance, where the properties of a system are unchanged under a transformation of scale. For instance, a Hamiltonian $H(x)$ is said to be scale invariant when $H(\lambda x) = \lambda^{\alpha} H(x)$, where $\lambda$ is a scaling factor and $\alpha$ is a real number. Scaling symmetries such as these can be violated by quantum fluctuations, which is a phenomenon known as a quantum anomaly. Such anomalous symmetry breaking is widely discussed in quantum field theory (*1*) and has fundamental implications in the contexts of high-energy physics and phase transitions. However, unambiguous experimental signatures of this effect, particularly in many-body systems, have so far been elusive. Here, we report the direct observation of a quantum anomaly in the dynamics of a two-dimensional Fermi superfluid.

Two-dimensional (2D) systems with contact interactions, $V(x) \propto \delta^2(x)$, are of particular interest in the context of scale-invariance violation because the $\delta^2$ potential does not introduce a characteristic scale to the Hamiltonian. At the classical level, the transformation $x \rightarrow \lambda x$ rescales the interaction potential as $V(\lambda x) \propto \lambda^{-2} V(x)$, exactly the same way as the kinetic energy, and therefore the classical 2D gas is intrinsically scale invariant (*2*, *3*). However, at the quantum mechanical level, this is no longer true because the $\delta^2$ scattering potential supports a two-body bound state for arbitrarily weak attraction (*4*–*6*). The bound state has a characteristic length $a_{2D}$, which describes the expected size of the molecule and introduces a spatial scale. This additional length scale is connected with the binding energy scale $E_B$, and it breaks the scaling relation between interaction and kinetic energy, which leads to a quantum anomaly.

How does this quantum anomaly influence the behavior of 2D systems at macroscopic scales? This question is especially relevant for 2D superfluids, which exhibit algebraic, hence scale-free, decay of phase correlations (*7*, *8*) described by the Berezinskii-Kosterlitz-Thouless (BKT) mechanism. In this context, how does the introduction of a short-distance scale ($a_{2D}$) affect the long-range behavior, such as spatial coherence and transport properties, in 2D superfluids? These questions are at the heart of many-body physics of 2D systems, and answering them may provide insights into the general phenomenology of other lower dimensional systems such as exciton-polariton condensates (*9*) and graphene (*10*, *11*).

In the field of ultracold atomic gases, 2D Bose gases in the weakly repulsive instead of interacting limit are demonstrably scale invariant (*3*, *12*, *13*), suggesting that the bound state plays a negligible role in these systems. However, in 2D Fermi gases, particularly in the strongly interacting regime, the effect of the additional length scale $a_{2D}$ becomes appreciable, for instance in the thermodynamic equation of state (*14*–*19*). On this basis, various theoretical works have predicted a quantitatively pronounced effect of the scale-invariance violation in this regime (*20*–*23*).

In harmonically trapped gases, a notable manifestation of this anomaly is an interaction-induced correction to the collective breathing mode frequency with respect to the noninteracting value (*4*, *20*–*22*, *24*) of twice the trap frequency. Although previous studies on breathing modes found no evidence of such a correction (*25*), the observation of an anomalous frequency shift at low temperatures has been reported in recent experiments (*26*, *27*). However, the relative magnitude of these shifts (~1 to 2%) is several times smaller than the theoretical prediction (~10%), raising questions on the physical relevance of the quantum anomaly for the dynamical properties of 2D Fermi gases. Rather than the breathing mode frequencies, here we explore the spatial coherence properties in momentum space, which reveal the scale-invariance breaking effect that is nearly absent in the position-space density profiles.

In our experiments, we prepared a gas of ~$2 \times 10^4$ $^6$Li atoms in the lowest two hyperfine states, trapped in a highly anisotropic potential and cooled to low temperatures deep in the superfluid phase. The ratio of absolute temperature to the Fermi temperature ($T/T_F$) is in the range of ~0.05 to 0.1. The radial and axial trap frequencies of the harmonic potential are $\omega_r = 2\pi \times 23$ Hz and $\omega_z = 2\pi \times 7.1$ kHz, respectively, corresponding to an aspect ratio $\omega_z/\omega_r \approx 310$. With the relevant thermodynamic scales kept smaller than the axial confinement energy, we ensure that the system is in the kinematically 2D regime. By tuning the interactions between fermions around a Feshbach resonance, we access the 2D Bose-Einstein condensate to the Bardeen-Cooper-Schrieffer (BEC-BCS) crossover region. The interactions in the 2D many-body system are described by a dimensionless parameter $\ln(k_F a_{2D})$, where $k_F$ is the Fermi momentum and $a_{2D}$ is the 2D scattering length obtained from the 3D scattering length (*5*, *6*, *28*). For $\ln(k_F a_{2D}) \ll -1$, we are in the BEC regime, whereas $\ln(k_F a_{2D}) \gg 1$ corresponds to the BCS regime. The strongly correlated regime located between these limits occurs when $1/k_F \sim a_{2D}$. This crossover region exhibits some notable features, such as enhanced critical temperature $T_c$ (*14*) and a large pseudogap region above $T_c$ where pairing is strongly density dependent (*29*).

We investigated the interplay between quantum anomaly and phase correlations by measuring the dynamical evolution of the gas both in position space (i.e., in situ) and in momentum space. Measuring the momentum distribution is particularly important as it encodes information of phase fluctuations in the superfluid. First, we brought the system out of its equilibrium configuration by resonantly modulating the harmonic trapping potential at twice the trap frequency $2\omega_r$ (Fig. 1, A and B). This protocol excites the 2D isotropic breathing mode whereby the gas undergoes periodic cycles of compression and expansion. After a fixed duration (10 cycles), the drive was stopped, and the cloud evolved in the original potential for a variable time *t*. In contrast to previous works, which investigated the frequency of the breathing mode, we focus on how the shapes of the in situ and momentum distributions change within a single breathing cycle. Because the damping rate of the breathing mode is very small (~$0.01\omega_r$) (*26*), the motion is essentially isentropic, which allows the direct probing of scale-invariant behavior.

[1]Physics Institute, Heidelberg University, Heidelberg, Germany. [2]Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: murthyp@phys.ethz.ch (P.A.M.); defenu@thphys.uni-heidelberg.de (N.D.) ‡Present address: Institute for Quantum Electronics, ETH Zurich, Zurich, Switzerland.

To measure the pair momentum distribution $n(k)$, we used a matter wave focusing technique that has been previously demonstrated for 2D gases (*30*, *31*). First, we rapidly ramped the offset magnetic field to the weakly interacting limit of strongly bound dimers. Immediately after the ramp, we switched off the trapping potential and released the sample to ballistically expand in a shallow harmonic potential for a quarter period $T_{exp}/4 = \pi/2\omega_{exp} = 21.8$ ms, where $\omega_{exp}$ is the shallow trap frequency. The $T_{exp}/4$ evolution maps the initial momentum distribution of particles to the spatial distribution. As the time scale of the magnetic field ramp ($\tau_{ramp} \sim 50$ μs) is shorter than the intrinsic time scales of the many-body system, the measured spatial distribution at $t = T_{exp}/4$ reflects, to a very good approximation, the initial momentum distribution of pairs. The strong enhancement of the low-momentum modes in $n(k)$, as seen in Fig. 1D, signals superfluidity in the system as it is related to long-range spatial coherence in the system (*7*, *14*).

In Fig. 1E, we show an example of the measured time-evolution of the in situ $\rho(r,t)$ (orange) and momentum distributions $n(k,t)$ (blue) taken at the interaction parameter $\ln(k_F a_{2D}) \approx 1$. The in situ distribution exhibits periodic compression and expansion at approximately twice the trap frequency ($\omega_B \approx 2\omega_r$), as expected. In contrast, $n(k,t)$ undergoes sharp revivals at twice the rate of $\rho(r)$, i.e., when the cloud size is maximum (outer turning point, $t = t_o$) as well as minimum (inner turning point, $t = t_i$). At intermediate time scales between the turning points, $n(k)$ is broadened. At a qualitative level, this peculiar effect can be understood as a consequence of the oscillation of the hydrodynamic velocity field at any point $(x, y)$, $\mathbf{v_B} \propto \sin(\omega_B t)[x\hat{\mathbf{e}}_\mathbf{x} + y\hat{\mathbf{e}}_\mathbf{y}]$. During the breathing cycle, $\mathbf{v_B}$ vanishes at the two turning points. At the intermediate points, the nonzero value of $\mathbf{v_B}$ manifests in a broadened momentum distribution, whereas the in situ profile shows a monotonous variation between the inner and outer turning points. We provide a more detailed description of the effect using variational Gross-Pitaevskii computations in (fig. S1) (*32*). A similar effect has recently been predicted for the 1D Bose gas in the Tonks-Girardeau regime using scale-invariant dynamics (*33*) and has also been experimentally observed in the weakly interacting regime (*34*).

From these dynamical measurements, the occurrence and violation of scale invariance can be studied by comparing the in situ and momentum-space distributions at different points in time. To illustrate this point, let us consider the time evolution of a scale-invariant gas in a harmonic potential. Naturally, the presence of a trapping potential introduces a length scale and thus explicitly breaks scale invariance. However, as pointed out in (*3*), the special case of a 2D harmonic potential has an inherent SO(2,1) symmetry that restores scaling behavior. Consequently, the harmonically trapped scale-invariant gas displays predictable dynamics with the time-dependent many-body wave function being given in terms of the equilibrium wave function according to

$$\psi(X,t) = \frac{1}{\lambda^N}\psi(X/\lambda, t=0)\times \exp\left(i\frac{m\dot{\lambda}}{2\hbar\lambda}X^2\right)\exp\left(i\theta(t)\right) \quad (1)$$

where $X = (\vec{x}_1, \vec{x}_2, ..., \vec{x}_N)$ are the $2N$ position coordinates of the many-body system, $m$ is the particle mass, $\theta(t)$ is an overall phase, and $\lambda(t)$ is the time-dependent scale factor that obeys the Ermakov-Milne equation (*32*). From the full wave function (Eq. 1), one obtains the evolution of the in situ density and the momentum distribution

$$\rho(\mathbf{r},t) = \frac{1}{\lambda^2}\rho\left(\frac{\mathbf{r}}{\lambda}, t=0\right) \quad (2)$$

$$n(\mathbf{k},t) = \lambda^2\int W\left(\lambda\mathbf{k} + \frac{m}{\hbar}\dot{\lambda}\lambda\mathbf{r}, \mathbf{r}, t=0\right)\mathrm{d}^2\mathrm{r} \quad (3)$$

in terms of the Wigner function $W(\mathbf{k},\mathbf{r},t)$. The in situ density is completely self-similar (Eq. 2), i.e., the density at any time $t$ can be rescaled to its initial form using a single scaling factor $\lambda(t)$. When $\dot{\lambda} = 0$, the momentum distribution $n(k,t)$ also displays self-similar scaling with the inverse factor $\lambda^{-1}$. For the breathing modes, $\dot{\lambda} = 0$ at the two turning points. Therefore, a comparison of the in situ and momentum distributions at the inner and outer turning points can be used as a proxy to study scale invariance.

We measured the dynamically evolving in situ and momentum distributions for various interaction parameters across the BEC-BCS crossover. In Fig. 2, we show $\rho(r)$ (left column) and $n(k)$ (right column) at the inner and outer turning points for interaction strengths $\ln(k_F a_{KD}) = -1.5$, 1, 1.3, 1.5, and 2, which correspond to the strongly interacting crossover region. The open diamonds and filled circles represent the distributions at successive inner ($t_{i,1}$, $t_{i,2}$) and outer ($t_{o,1}$, $t_{o,2}$) turning points within a breathing cycle (Fig. 1E). The scaling to the inner turning point distribution is performed using the mean of the successive outer turning point distributions. In the in situ distributions, the outer turning point profile $\rho(r,t_o)$ (blue) can be collapsed onto the inner turning point profile $\rho(r,t_i)$ using a global scaling factor $0 < \lambda < 1$. The rescaling is represented by the dashed black curves in Fig. 2, where $\rho_{sc}(r) = \lambda^{-2}\rho(r/\lambda,t_o)$; the scaling factor $\lambda$ is obtained by minimizing the mean square deviation between the inner profile and the rescaled outer profile. The measured and rescaled distributions coincide within the systematic and statistical uncertainties of the measured density, which is ~5% (*14*).

In momentum space, the inverse scaling factor $\lambda^{-1}$ should collapse the inner and outer turning-point distributions if the system were scaling invariant. This condition is satisfied to a good approximation both in the BEC $[\ln(k_F a_{2D}) \lesssim -1.5]$ and BCS $[\ln(k_F a_{2D}) \gtrsim 2]$ regimes (Fig. 2, A and E). In these regimes, the difference between the scaling factor obtained independently for the $k$-space distributions and the inverse in situ scaling factor is <2%. However, in the crossover region around $\ln(k_F a_{2D}) \sim 1.3$, we find a notable discrepancy between the measured $n(k,t_i)$ at the inner turning point and the rescaled distribution $n_{sc}(k)$ obtained using the inverse in situ scaling factor. In fact, we expect $n(k,t_i)$ to always be broader

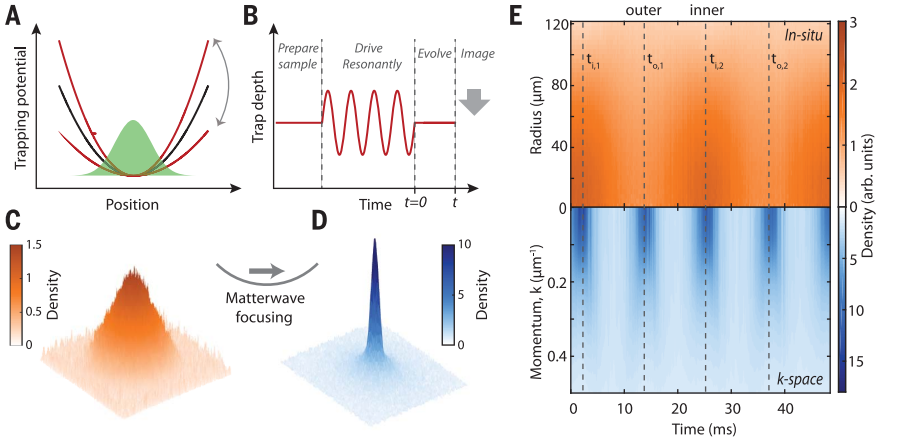



**Fig. 1. Dynamics of a 2D fermionic superfluid in position and momentum space.** (**A** and **B**) We prepare a 2D Fermi gas well below the superfluid critical temperature ($0.05 < T/T_F < 0.1$) (*14*). The isotropic breathing mode is excited by resonantly modulating the harmonic trap at twice the trap frequency. Once the drive is stopped, the breathing oscillations continue for a variable time $t$, at which point we measure (**C**) the in situ density distribution $\rho(r,t)$ and (**D**) the pair momentum distribution $n(k,t)$ using a matter wave focusing technique. (**E**) Example of azimuthally averaged $\rho(r,t)$ (orange) and $n(k,t)$ (blue) taken at interaction strength $\ln(k_F a_{2D}) \approx 1$. The in situ density oscillates at twice the trap frequency, as expected. The momentum distribution exhibits sharp revivals at twice the rate of the in situ oscillation. The frequency doubling arises from the sinusoidal oscillation of the hydrodynamic velocity field, which vanishes at the inner and outer turning points of the breathing cycle, denoted by the vertical dashed lines.

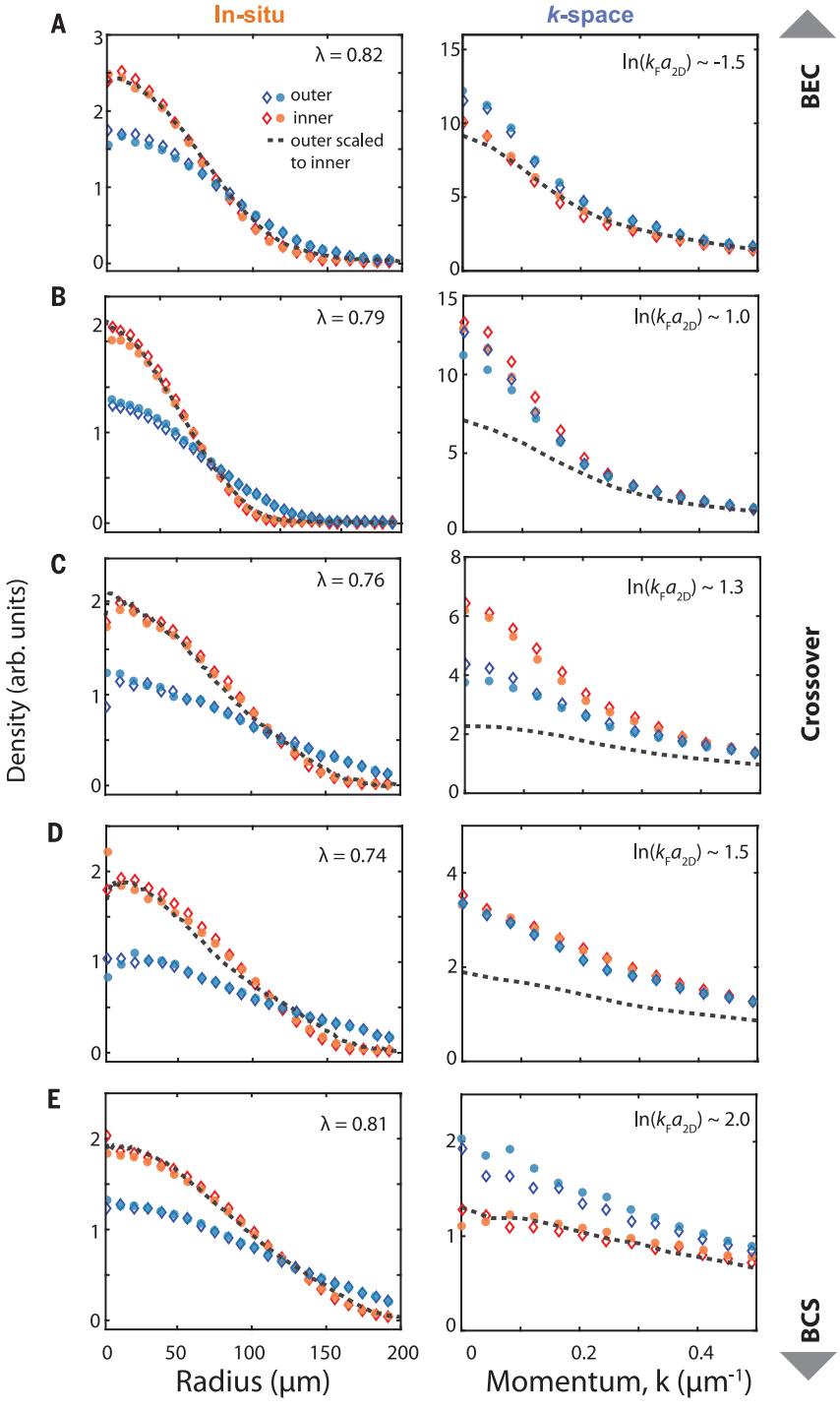



**Fig. 2. Scale-invariance breaking in momentum space.** The in situ (left column) and pair momentum distributions (right column) at the inner and outer turning points for interaction strengths $\ln(k_F a_{2D}) \approx -1.5$, 1.0, 1.3, 1.5, and 2.0 (**A** to **E**, respectively). The diamonds and filled circles represent the distributions at adjacent inner and outer turning points. For a scale-invariant system, the in situ density profiles at $t_o$ (red diamonds) and $t_i$ (blue circles) should be scalable with a single scaling factor λ, as well as the momentum distributions $[n(k,t_o) \rightarrow n(k,t_i)]$ with the inverse factor $\lambda^{-1}$. Such scaling behavior is observed both in the weakly interacting BEC and BCS regimes. However, in the strongly interacting crossover regime, we find a clear departure from scale invariance. Although the evolution of the $\rho(r)$ is still self-similar, the momentum distribution shows a notable discrepancy from the expected result obtained with the inverse scaling factor from the in situ scaling (dashed black line). This scaling violation at strong interactions is attributed to the quantum anomaly.

than $n(k,t_o)$ (see Fig. 2A), but the measured momentum distribution at $\ln(k_F a_{2D}) \sim 1.3$ shows the opposite effect. Here, the occupation of the low-$k$ region of $n(k)$ is strongly enhanced, not only with respect to the expected distribution but also compared with $n(k,t_o)$. This discrepancy is evidence that scale invariance is violated owing to strong interactions, with an unmistakable signature in momentum space.

As shown in Fig. 2, the fermionic interactions have a substantial influence on the low-$k$ modes, which correspond to long-wavelength phase fluctuations in the superfluid. The correlations in the phase are characterized by the first-order correlation function

$$g_1(r) = \int \rho_1(\mathbf{R} - \mathbf{r}/2, \mathbf{R} + \mathbf{r}/2)\, d^2R \qquad (4)$$

where $\rho_1$ is the one-body density matrix. Experimentally, $g_1(r)$ is directly obtained from the $n(k)$ through a Fourier transform. In (*7*), a transition from exponential to algebraic decay in the trap-averaged $g_1(r)$ was observed, in agreement with BKT theory and quantum Monte Carlo computations (*35*). Here, we use the same procedure described in (*7*) to extract $g_1(r)$ at the inner and outer turning points. These are shown in Fig. 3A for $\ln(k_F a_{2D}) = -6$ and 1.3. To account for the change in cloud size when comparing the two correlation functions, we plot $g_1(\lambda r,t_o)$ in rescaled coordinates, where λ is the scaling factor obtained from the procedure described above (Fig. 2 and Eq. 2). In addition, we extract the exponent η by fitting a power law $[f(r) \sim r^{-\eta(t)}]$ to $g_1(r,t)$. Although the exponents in the trap-averaged $g_1(r)$ are substantially larger than the homogeneous BKT predictions, they have the same qualitative behavior (*35*), in particular, a smaller exponent corresponds to a larger superfluid phase space density $D_s = \rho_s \lambda_T^2$, where $\rho_s$ is the superfluid density and $\lambda_T$ is the thermal de Broglie wavelength. The power-law exponents obtained at the two turning points are tabulated in table S1 (*32*).

In the BEC regime, the two curves $[g_1(r,t_i)$ and $g_1(\lambda r,t_o)]$ collapse onto each other (Fig. 3A), whereas in the crossover regime, the correlation functions are substantially different, with the inner $g_1(r,t_i)$ decaying slower than expected. In Fig. 3B, we show the ratio $\eta_i/\eta_o$ for different interaction strengths across the BEC-BCS crossover. For scale-invariant systems, $\eta_i = \eta_o$, i.e., the spectrum of phase fluctuations is unaffected by a change in the density. Indeed, we find $\eta_i/\eta_o \approx 1$ in the BEC regime, but the ratio dips dramatically in the crossover regime to a value of ~0.8 before rising again in the weakly interacting BCS regime. This quantitative deviation proves that the quantum scale anomaly that originates in the short-distance fermionic correlations influences the algebraic decay of correlations in the 2D superfluid. An equivalent signature is obtained by comparing the outer zero-momentum occupation with the rescaled one at the inner turning point; the resulting curve, shown in the inset of Fig. 3B, also deviates from the scale-invariant expectation in the crossover region.

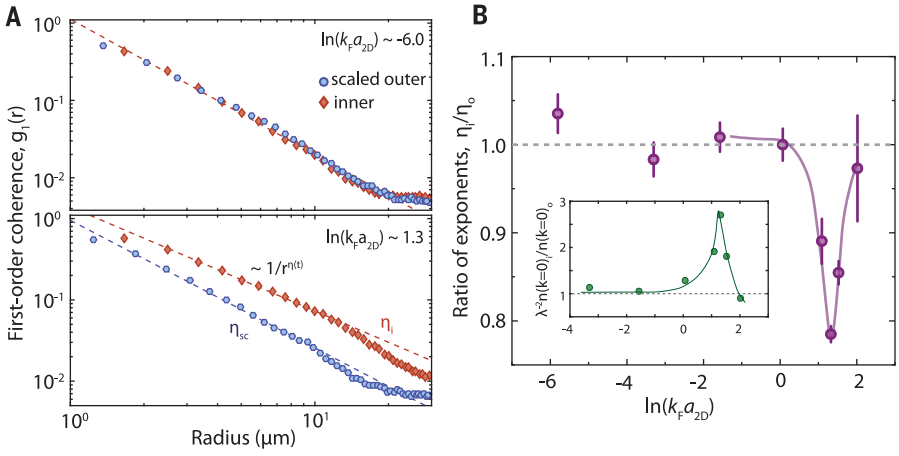



**Fig. 3. The quantum anomaly and spatial coherence.** (**A**) The first-order correlation function $g_1(r,t_i)$ at inner point (red) and rescaled correlation function $g_1(\lambda r,t_o)$ at the outer points (blue), for $\ln(k_F a_{2D}) \sim -6$ (upper panel, BEC) and $\ln(k_F a_{2D}) \sim 1.3$ (lower panel, crossover), where λ is the real space scaling factor obtained in Fig. 2 . In the BEC regime, $g_1(r,t_i)$ and $g_1(\lambda r,t_o)$ coincide, whereas in the crossover regime, the two curves are conspicuously different. From the power-law decay of $g_1(r) \sim r^{-\eta}$, we extract the exponent η. (**B**) The ratio $\eta_i/\eta_o$ across the BEC-BCS crossover. The scale-invariant expectation $\eta_i/\eta_o = 1$ is reproduced in the BEC regime. In the crossover regime, we observe a sharp dip in the ratio signaling the scaling violation in the long-range phase correlations. The minimum ratio is at $\ln(k_F a_{2D}) \sim 1.3$, which coincides with the regime of many-body pairing observed in (*29*). (Inset) The ratio between the zero pair momentum occupation at the inner and outer turning points, divided by $1/\lambda^2$; as above, the largest anomaly is observed in the crossover region. The purple and green curves are guides to the eye.

What is the origin of these effects? The interaction region [$\ln(k_F a_{2D}) \sim 1$] where we see the largest scaling violation in the phase correlations coincides with the regions of (i) maximum critical temperature (*14*), (ii) largest density-dependent pairing (pseudogap) (*29*), and (iii) the maximum breathing mode frequency shift (*26*, *27*). This suggests that all these effects may have a common mechanism. However, the exact dependence of these effects on $\ln(k_F a_{2D})$ is slightly different because local properties such as fermion pairing and long-range properties such as coherence respond differently to temperature. Also, because the breathing motion in the system is much slower than the microscopic scattering rate between fermions, one can apply the traditional hydrodynamic picture where the gas can be considered locally in equilibrium at all times. This allows the dynamical behavior of the gas to be connected with its equilibrium properties.

In this framework, one possible mechanism arises from the density-dependent pairing effect observed in (*29*). In the crossover region, a change in density during the breathing cycle corresponds to a change in the total pairing energy. However, in 2D BCS theory, the coherence length remains fixed to the vacuum expectation $a_{2D}$ irrespective of the density. Accordingly, as the particle spacing is the smallest at the inner turning point, this implies enhanced phase coherence extending over more particle spacings and a smaller decay exponent η. At the same time, enhanced occupation of low-momentum modes requires, at a fixed total number, a reduced occupation at high momenta and hence a depletion in the pair kinetic energy. We have analyzed the kinetic energy extracted from the momentum distribution and found a scaling violation consistent with this argument (fig. S2) (*32*).

The observations in Fig. 3 may also provide hints toward explaining the enhanced critical temperatures in this region. We recall that the power-law exponents are an indicator of superfluid stiffness and phase-space density: a smaller η corresponds to more coherence and greater stiffness $D_s$. For scale-invariant systems, $D_s$ necessarily remains constant throughout the breathing cycle leading to $\eta_i/\eta_o = 1$. However, in the crossover regime, the observation of $\eta_i/\eta_o < 1$ implies that the density-dependent pair correlations enhance the superfluid phase space density for the same effective temperature. In other words, the critical $D_s$ required for the superfluid transition can be attained at higher $T_c/T_F$, as seen in (*14*).

Finally, we note the differences between the manifestations of the anomaly in the breathing mode frequency shifts and coherence measurements. The density profiles at the turning points do not exhibit conspicuous effects of the quantum anomaly and satisfy the prediction of the dynamical SO(2,1) symmetry (*24*). This is consistent with the small shifts in the breathing mode frequency reported in (*26*, *27*). It further shows that the breathing mode dynamics are not fully explained by the equation of state (*15*–*17*), which is scale dependent and would imply a large shift in the breathing frequency accompanied by an observable change in the in situ density profile. On the other hand, the coherence of the system probes the long-wavelength phase fluctuations and thus displays a much larger effect of the quantum anomaly. In addition, the coherence properties in the superfluid phase are more sensitive to temperature, which leads to a slightly different dependence on the $\ln(k_F a_{2D})$ with respect to local measurement of many-body pairing (*29*). An important goal for future investigations is to find a theoretical description that connects these different effects—many-body pairing, enhanced critical temperature, breathing mode frequencies, and spatial coherence—in the crossover region. Although the relation between many-body pairing and breathing mode properties has been recently demonstrated theoretically (*36*), the connection between the quantum anomaly and coherence remains an open question.

## REFERENCES AND NOTES


1. S. Weinberg, *The Quantum Theory of Fields*, vol. 2 (Cambridge Univ. Press, 1995).
2. B. R. Holstein, *Am. J. Phys.* **61**, 142–147 (1993).
3. L. P. Pitaevskii, A. Rosch, *Phys. Rev. A* **55**, R853–R856 (1997).
4. M. Olshanii, H. Perrin, V. Lorent, *Phys. Rev. Lett.* **105**, 095302 (2010).
5. D. S. Petrov, M. Holzmann, G. V. Shlyapnikov, *Phys. Rev. Lett.* **84**, 2551–2555 (2000).
6. D. S. Petrov, G. V. Shlyapnikov, *Phys. Rev. A* **64**, 012706 (2001).
7. P. A. Murthy *et al.*, *Phys. Rev. Lett.* **115**, 010401 (2015).
8. Z. Hadzibabic, P. Krüger, M. Cheneau, B. Battelier, J. Dalibard, *Nature* **441**, 1118–1121 (2006).
9. R. Balili, V. Hartwell, D. Snoke, L. Pfeiffer, K. West, *Science* **316**, 1007–1010 (2007).
10. O. Ovdat, J. Mao, Y. Jiang, E. Y. Andrei, E. Akkermans, *Nat. Commun.* **8**, 507 (2017).
11. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
12. C.-L. Hung, X. Zhang, N. Gemelke, C. Chin, *Nature* **470**, 236–239 (2011).
13. R. Desbuquois *et al.*, *Phys. Rev. Lett.* **113**, 020404 (2014).
14. M. G. Ries *et al.*, *Phys. Rev. Lett.* **114**, 230401 (2015).
15. V. Makhalov, K. Martiyanov, A. Turlapov, *Phys. Rev. Lett.* **112**, 045301 (2014).
16. I. Boettcher *et al.*, *Phys. Rev. Lett.* **116**, 045303 (2016).
17. K. Fenech *et al.*, *Phys. Rev. Lett.* **116**, 045302 (2016).
18. G. Bertaina, S. Giorgini, *Phys. Rev. Lett.* **106**, 110403 (2011).
19. H. Shi, S. Chiesa, S. Zhang, *Phys. Rev. A* **92**, 033603 (2015).
20. J. Hofmann, *Phys. Rev. Lett.* **108**, 185303 (2012).
21. S. Moroz, *Phys. Rev. A* **86**, 011601 (2012).
22. C. Gao, Z. Yu, *Phys. Rev. A* **86**, 043609 (2012).
23. W. Daza, J. E. Drut, C. Lin, C. Ordóñez, *Phys. Rev. A* **97**, 033630 (2018).
24. E. Taylor, M. Randeria, *Phys. Rev. Lett.* **109**, 135301 (2012).
25. E. Vogt *et al.*, *Phys. Rev. Lett.* **108**, 070404 (2012).
26. M. Holten *et al.*, *Phys. Rev. Lett.* **121**, 120401 (2018).
27. T. Peppler *et al.*, *Phys. Rev. Lett.* **121**, 120402 (2018).
28. P. Dyke *et al.*, *Phys. Rev. A* **93**, 011603 (2016).
29. P. A. Murthy *et al.*, *Science* **359**, 452–455 (2018).
30. S. Tung, G. Lamporesi, D. Lobser, L. Xia, E. A. Cornell, *Phys. Rev. Lett.* **105**, 230408 (2010).
31. P. A. Murthy *et al.*, *Phys. Rev. A* **90**, 043611 (2014).
32. See supplementary materials.
33. Y. Y. Atas, I. Bouchoule, D. M. Gangardt, K. V. Kheruntsyan, *Phys. Rev. A* **96**, 041605 (2017).
34. B. Fang, G. Carleo, A. Johnson, I. Bouchoule, *Phys. Rev. Lett.* **113**, 035301 (2014).
35. I. Boettcher, M. Holzmann, *Phys. Rev. A* **94**, 011602 (2016).
36. T. Enss, Dynamical viscosity, contact correlations and pairing in attractive Fermi gases. arXiv:1904.12772 [cond-mat. quant-gas] (29 April 2019).

37. P. A. Murthy, N. Defenu, L. Bayha, M. Holten, P. M. Preiss, T. Enss, S. Jochim, Quantum anomaly and spatial coherence in a 2D Fermi superfluid, Version 1, Zenodo (2019); https://doi.org/10.5281/zenodo.2624085V.

**ACKNOWLEDGMENTS**

We thank I. Boettcher, T. Gasenzer, J. Hofmann, K. V. Kheruntsyan, T. Lompe, and S. Moroz for insightful discussions. **Funding:** This work has been supported by the ERC consolidator grant 725636, the Heidelberg Center for Quantum Dynamics, and is part of the DFG Collaborative Research Center SFB 1225 (ISOQUANT). P.M.P. acknowledges funding from the European Union's Horizon 2020 program under Marie Skłodowska-Curie grant agreement 706487. **Author contributions:** P.A.M. performed the measurements and data analysis. N.D. and T.E. provided the theoretical description of the observations. L.B., M.H., and P.M.P. assisted with experiments and interpretation of data. T.E. and S.J. supervised the project. P.A.M. and N.D. conceptualized and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data shown in this paper are deposited in (*37*).

**SUPPLEMENTARY MATERIALS**

science.sciencemag.org/content/365/6450/268/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Table S1
References (*38*, *39*)

11 June 2018; accepted 18 June 2019
10.1126/science.aau4402

MOLECULAR KNOTS

# Topological molecular nanocarbons: All-benzene catenane and trefoil knot

**Yasutomo Segawa[1,2]*, Motonobu Kuwayama[1], Yuh Hijikata[2,3,4], Masako Fushimi[1,2], Taishi Nishihara[1,2,5], Jenny Pirillo[3,4], Junya Shirasaki[1], Natsumi Kubota[2], Kenichiro Itami[1,2,3]***

**The generation of topologically complex nanocarbons can spur developments in science and technology. However, conventional synthetic routes to interlocked molecules require heteroatoms. We report the synthesis of catenanes and a molecular trefoil knot consisting solely of *para*-connected benzene rings. Characteristic fluorescence of a heterocatenane associated with fast energy transfer between two rings was observed, and the topological chirality of the all-benzene knot was confirmed by enantiomer separation and circular dichroism spectroscopy. The seemingly rigid all-benzene knot has rapid vortex-like motion in solution even at −95°C, resulting in averaged nuclear magnetic resonance signals for all hydrogen atoms. This interesting dynamic behavior of the knot was theoretically predicted and could stimulate deeper understanding and applications of these previously untapped classes of topological molecular nanocarbons.**

Carbon nanostructures such as fullerenes (*1*), carbon nanotubes (CNTs) (*2*), and graphene (*3*) and their partial molecular substructures (molecular nanocarbons) (*4*–*9*) have revolutionized the research fields of chemistry, physics, materials science, and nanoscience (Fig. 1A). Historically, these studies have shown that the emergence of distinct geometries and morphologies of carbon leads to the discovery of functions and applications that are not initially predicted nor expected (*10*–*12*). Currently, the known variations of nanocarbon structure are all topologically simple. There are numerous theoretically proposed nanocarbons with unexplored topologies, such as carbon nanotori (Fig. 1A) (*13*–*15*), nanocoils (*13*–*15*), and Mackay crystals (*16*). However, the lack of methods for creating such carbon topologies hampers the advancement of this field. In particular, mechanically interlocked molecules (MIMs) such as catenanes and molecular knots (*17*–*20*) are interesting motifs that provide heretofore unrealized topologies in the context of nanocarbon structures.

We report the synthesis of all-benzene catenanes **1a** and **1b** and trefoil knot **2** (Fig. 1B). These molecules are cycloparaphenylenes (CPPs) (*7*, *8*)—a class of molecular nanocarbon that comprises a sidewall segment structure of CNTs—with the topology of catenanes and trefoil knots. They retain the high symmetry and radial π-conjugation modes characteristic of CPPs, but also demonstrate distinctive intramolecular electronic interactions and dynamic motion. Geometrically, trefoil knots are in the class of torus knots (*21*),

[1]JST-ERATO, Itami Molecular Nanocarbon Project, Chikusa, Nagoya 464-8602, Japan. [2]Graduate School of Science, Nagoya University, Chikusa, Nagoya 464-8602, Japan. [3]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Chikusa, Nagoya 464-8602, Japan. [4]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo, Hokkaido 001-0021, Japan. [5]Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan.
***Corresponding author. Email: itami@chem.nagoya-u.ac.jp (K.I.); ysegawa@nagoya-u.jp (Y.S.)**

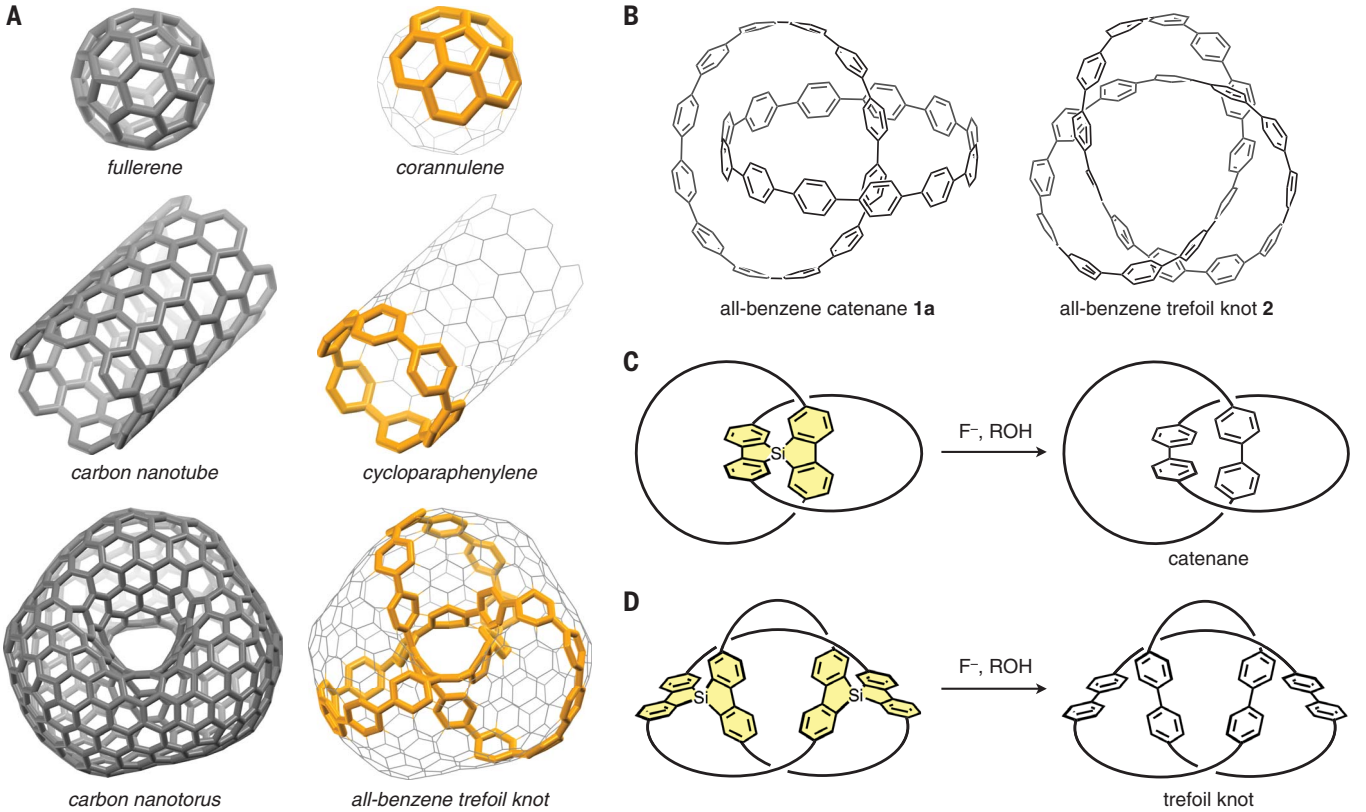



**Fig. 1. Topological molecular nanocarbons.** (**A**) Structures of fullerene, CNT, carbon nanotorus, and their segmental molecules (corannulene, CPP, and all-benzene trefoil knot). (**B**) Structures of all-benzene catenane **1a** and all-benzene trefoil knot **2**. (**C** and **D**) Strategy for the synthesis of all-benzene catenane (C) and the trefoil knot (D). $F^-$, fluoride anion; ROH, alcohol.

which can be placed on the surface of a torus. As shown in Fig. 1A, the all-benzene trefoil knot **2** is a partial segment of a carbon nanotorus.

Conventionally, MIMs have been synthesized by using several types of reversible interactions such as metal–ligand coordination (*22*), electrostatic and π–π stacking interactions (*23*, *24*), as well as hydrogen bonding (*25*). All-hydrocarbon structures have been left behind in the history of MIMs since the report on the synthesis of a cycloalkane catenane in 1983 (*26*), presumably because of the lack of an efficient synthetic method. In 2016, the existence of all-benzene catenanes and possibly trefoil knots was suggested from a detailed mass spectral analysis of a mixture of CPPs. However, results of further spectral or structural investigations were not reported, as individual components were difficult to isolate (*27*). Two recent reports of MIMs with nitrogen-containing CPP derivatives rely on the metal–ligand coordination strategy (*28*, *29*). Thus, a distinct synthetic method to access all-benzene catenanes and knots is necessary to explore this exciting field. Our strategy (Fig. 1, C and D) drew inspiration from the known quantitative conversion of a spirobi(dibenzosilole) to two biphenyls upon treatment with fluoride in

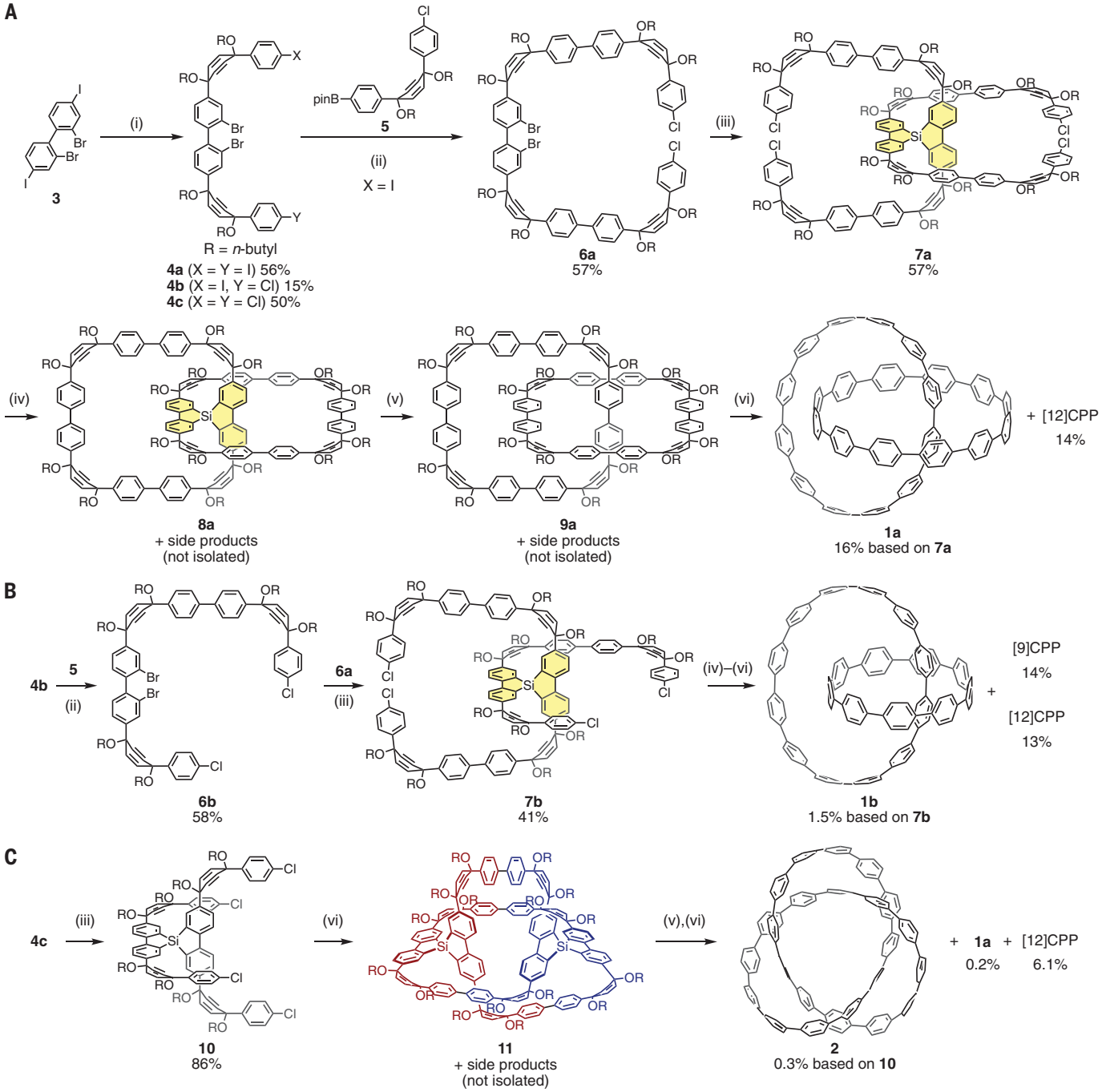



**Fig. 2. Synthetic routes to 1a (A), 1b (B), and 2 (C).** Reaction conditions: (i) Four steps (see the supplementary materials for details). (ii) Catalytic $Pd(PPh_3)_4$, $K_2CO_3$, toluene/EtOH/water or THF/DMF/water, reflux. (iii) *n*-BuLi; then TMEDA, $SiHCl_3$, THF/$Et_2O$. (iv) $Ni(cod)_2$, 2,2′-bipyridyl, DMF. (v) *n*-$Bu_4NF$, THF/EtOH. (vi) Sodium naphthalenide, THF, and then $I_2$. THF, tetrahydrofuran; cod, 1,5-cyclooctadiene; DMF, *N*,*N*-dimethylformamide; TMEDA, *N*,*N*,*N*′,*N*′-tetramethylethylenediamine; pinB, 4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl.

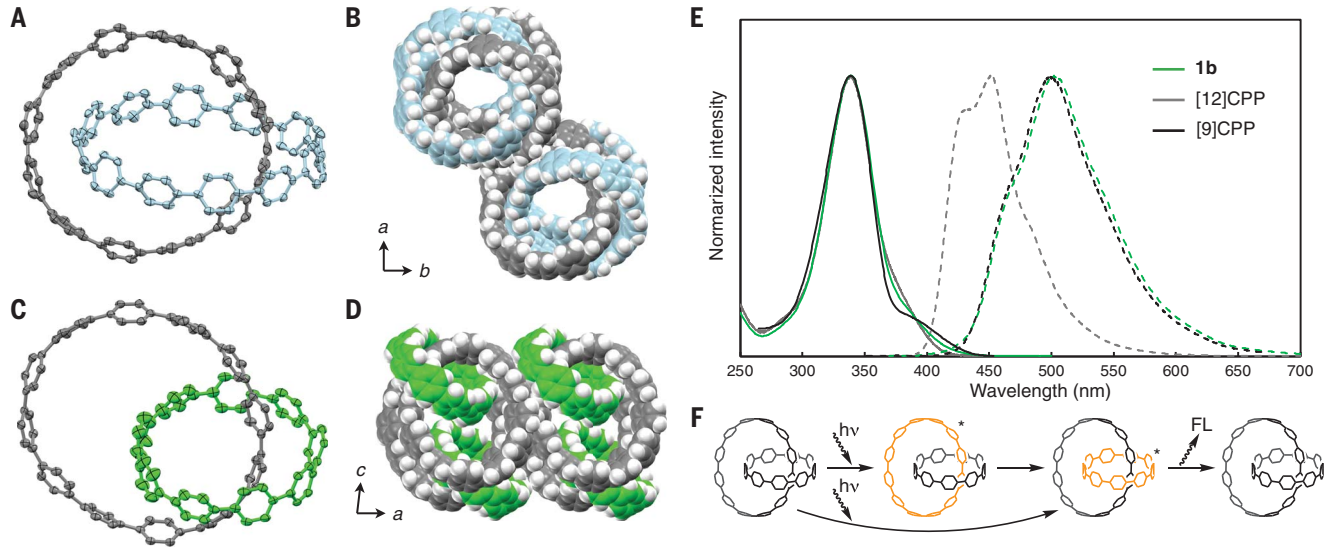



**Fig. 3. Structures and properties of all-benzene catenanes 1a and 1b.** (**A** and **C**) Oak Ridge thermal ellipsoid plot (ORTEP) drawings of **1a** (A) and **1b** (C) with thermal ellipsoids set to 50% probability. Hydrogen atoms and solvent molecules are omitted for clarity. (**B** and **D**) Packing structures of **1a** (B) and **1b** (D); carbon: gray, blue, or green; hydrogen: white. Solvent molecules are omitted for clarity. (**E**) Ultraviolet-visible absorption (solid lines) and fluorescence (dashed lines) spectra of the dichloromethane solutions of **1b**, [9]CPP, and [12]CPP. The fluorescence spectra were measured upon excitation at 340 nm for [12]CPP and [9]CPP or 360 nm for **1b**. (**F**) Hypothetical illustration of the fluorescence mechanism of **1b**. Orange color with asterisk (*) represents excited moieties. hν, light irradiation; FL, fluorescence.

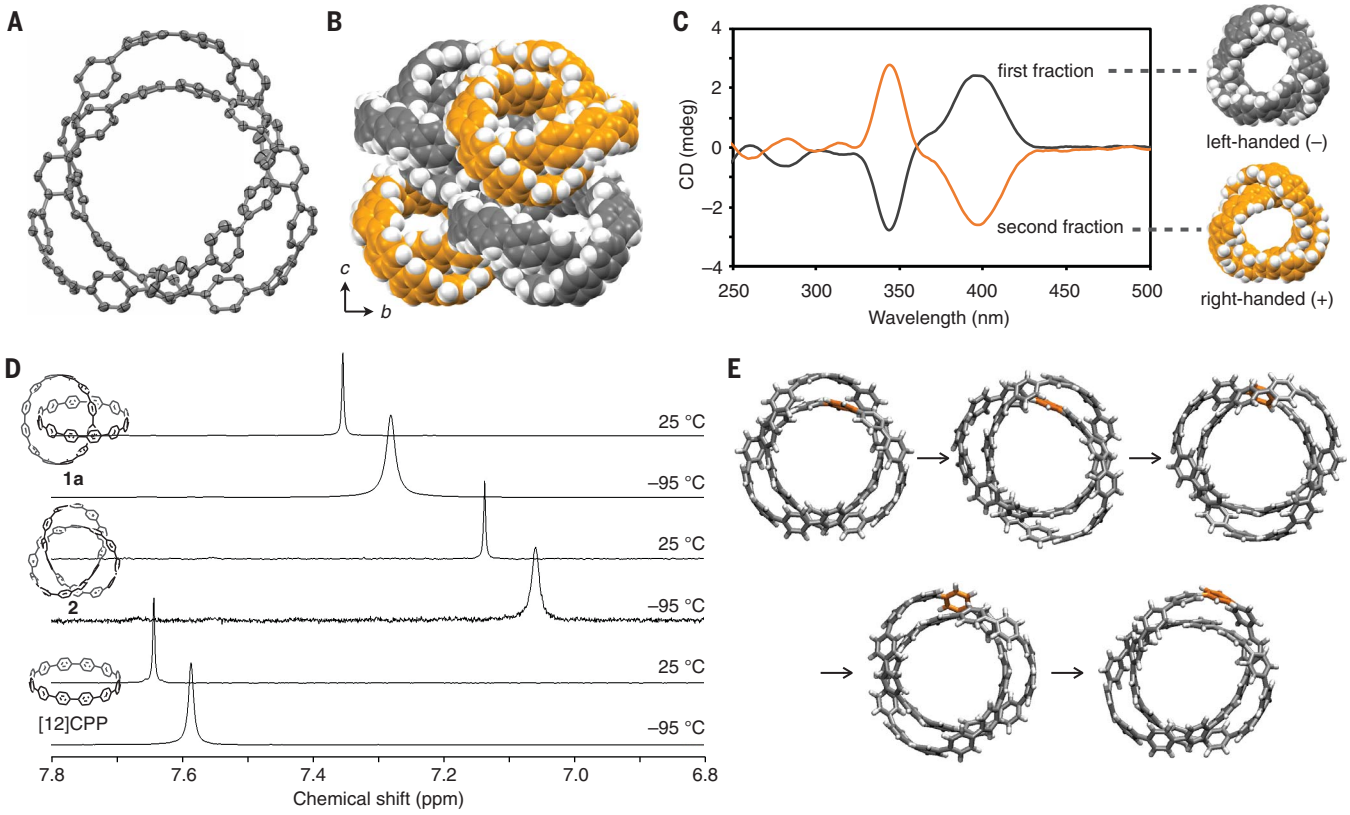



**Fig. 4. Structure and properties of all-benzene knot 2.** (**A**) ORTEP drawing of **2** with thermal ellipsoids set to 50% probability. Hydrogen atoms and solvent molecules are omitted for clarity. (**B**) Packing structures of **2**; carbon: gray or orange; hydrogen: white. Solvent molecules are omitted for clarity. (**C**) CD spectra of the enantiomers of **2** and their assignment. (**D**) $^{1}$H-NMR spectra of **1a**, **2**, and [12]CPP in $CD_2Cl_2$ at 25°C and −95°C. (**E**) Snapshots of the DFTB-MD simulation of **2** (carbon: gray or orange; hydrogen: white). See the supplementary materials and methods and movie S1 for details.

ethanol (*30*). We envisioned that the spirosilane moiety could similarly be used as a traceless template for the synthesis of all-benzene catenanes and trefoil knots, which would contain no remnants of it.

The synthetic route to [12]CPP-based [2]catenane (**1a**) is shown in Fig. 2A. According to the synthetic protocol reported by Jasti and colleagues (*31*), dialkoxycyclohexadiene moieties were selected as bent paraphenylene precursors, whereas *n*-butoxy groups were used instead of methoxy groups to increase the solubility of the intermediates. The starting material, 2,2′-dibromo-4,4′-diiodobiphenyl (**3**), was converted into a dibrominated U-shaped unit **4a** in four steps that included an iodo-selective lithiation, a nucleophilic addition to a *p*-quinone derivative, a nucleophilic addition of *p*-halolithiobenzenes, and an *n*-butylation of the resulting hydroxy groups. A Suzuki–Miyaura coupling of **4a** and L-shaped unit **5**, which bears chloro and boryl groups (*31*), took place selectively at the iodo moieties of **4a** to form C-shaped unit **6a** in moderate yield. For the formation of spirosilane **7a**, conventional reaction conditions (dilithiation with *n*-BuLi, followed by spirosilylation with $SiCl_4$) were not applicable because of the high Lewis acidity of $SiCl_4$, which reacts quickly with the *n*-butoxy moieties. After screening the reaction conditions using a model reaction (compare table S1), we discovered that the complex of $SiHCl_3$ and TMEDA (*N,N,N′,N′*-tetramethylethylenediamine) (*32*) was suitable for this spirosilylation step. Sequential lithiation of **6a** and addition of $SiHCl_3$ and TMEDA afforded spirosilane **7a** in 57% yield.

Catenane **1a** was synthesized from **7a** in three steps including a Ni(0)-mediated intramolecular aryl–aryl coupling reaction at the C–Cl moieties, a fluoride-mediated desilylation, and a reductive aromatization of the dialkoxycyclohexadiene units (Fig. 2A). As a result, the desired catenane **1a** was isolated in 16% yield (9.0 mg) from **7a** by column chromatography and preparative thin-layer chromatography (PTLC), along with the generation of pristine [12]CPP (14% yield, 8.0 mg). Although intermediates were neither isolated nor characterized, precursors **8a** and **9a** were expected to be generated from the aryl–aryl coupling step and the desilylation step, respectively (see fig. S1 for the plausible intermediate of [12]CPP). The $^1$H–nuclear magnetic resonance (NMR) spectrum of **1a** in $CD_2Cl_2$ showed a singlet peak at 7.35 parts per million (ppm), which indicates fast mutual rotation of the two [12]CPP components of **1a** in solution.

By applying this strategy, the all-benzene catenanes consisting of two different-sized rings could also be synthesized (Fig. 2B). Starting from C-shaped unit **6b** in addition to **6a**, hetero-spirosilylation successfully occurred to furnish **7b** in 41% yield together with relatively small amounts of the homo-spirosilylation products **7a** and **7c**, which is a possible precursor for [9]CPP catenane. The thus-obtained hetero-spirosilane **7b** was converted into [9]CPP–[12]CPP heterocatenane **1b** in 1.5% yield (2.0 mg) by sequential homocoupling reaction, desilylation, and reductive aromatization reactions. The $^1$H-NMR signals of **1b** in $CD_2Cl_2$ were observed at 7.29 and 7.38 ppm, with the integral ratio of 9:12 indicating 1:1 catenated structure of [9]CPP and [12]CPP in **1b**.

The catenated structures of **1a** and **1b** were confirmed by x-ray crystallography (Fig. 3, A and C). In the solid state, catenanes **1a** and **1b** stack in one dimension to form void channels (Fig. 3, B and D), wherein solvent molecules used for recrystallization (hexane and chloroform for **1a** and 1,4-dioxane and chloroform for **1b**) were incorporated. Intramolecular π–π interactions are evident, with the C–C distance of the nonbonding benzene rings ~3.4 to 3.6 Å (see fig. S2 for details). Absorption and fluorescence measurements of **1b** (Fig. 3E) revealed unusual features of this heterocatenane compared with [9]CPP and [12]CPP as reference molecules (*33*). The fluorescence spectrum of **1b** is almost identical to that of [9]CPP without any trace of the fluorescence peaks that originated from [12]CPP, whereas the absorption spectrum of **1b** is the simple combination of those of [9]CPP and [12]CPP. The fluorescence lifetime of **1b** (10.9 ns) is also similar to that of [9]CPP (10.6 ns) (*33*). Judging from the fact that the solution of the mixture of [9]CPP and [12]CPP shows the fluorescence peaks of both [9]CPP and [12]CPP (fig. S4), the complete fluorescence quenching of the [12]CPP moiety in **1b** seems to be a consequence of the catenated structure. As depicted in Fig. 3F, fast energy transfer from the excited [12]CPP moiety to the [9]CPP moiety in **1b** would account for these observations (fig. S5). This result clearly demonstrates the effect of the catenated structure on the photophysical properties of fully conjugated rings, in that catenation is the only way to connect the all-benzene rings without breaking their high symmetry.

We hypothesized that our traceless synthetic strategy could be extended to the generation of an even more challenging, all-benzene trefoil knot **2**, as shown in Fig. 1C. Inspired by the report of Dietrich-Buchecker and Sauvage, who constructed the trefoil-knot–type topology using two phenanthroline-Cu moieties as template units (*34*), we designed **11** (Fig. 2C) as the key intermediate, in which a loop consisting of 16 paraphenylenes and 8 cyclohexadiene-diyl units is knotted through two spirosilane joints. As the key intermediate **11** was a dimerized product of spirosilane **10**, we started with the synthesis of **10**. The U-shaped unit **4c** was lithiated and silylated under the aforementioned conditions ($SiHCl_3$ and TMEDA) to afford **10** in 86% yield. In a manner similar to the synthesis of **1a** and **1b**, **10** was subjected to a sequence of Ni(0)-mediated aryl-aryl coupling, desilylation, and reductive aromatization without isolating the intermediates. Fortunately, after extensive chromatographic separation of the crude mixture, the all-benzene trefoil knot **2** was isolated by PTLC and preparative high-performance liquid chromatography (HPLC) in 0.3% yield (0.8 mg), along with [12]CPP (17.3 mg, 6.1%) and **1a** (0.5 mg, 0.2%; see fig. S1 for the plausible intermediate of **1a** from **10**).

Single crystals of **2** were obtained from a hexane and 1,2-dichloroethane solution of **2**, and x-ray crystallography unambiguously corroborated the trefoil-knot structure (Fig. 4A). Similar to **1a** and **1b**, intramolecular π-π stacking was observed (see the supplementary materials for details). Trefoil knot **2** is chiral, and both enantiomers (gray and orange) were incorporated in a 1:1 ratio in the crystal (Fig. 4B). Separation of the enantiomers of **2** was achieved by chiral HPLC, and the circular dichroism (CD) spectra of both enantiomers were successfully recorded. Judging from the simulated CD spectra by density functional theory (DFT) calculations, the first and second HPLC fractions could be assigned to the left-handed (–) and right-handed (+) (*35*) trefoil-knot structures, respectively (Fig. 4C).

The seemingly rigid knot **2** was demonstrated to have rapid dynamic motion in solution. The $^1$H-NMR spectrum for **2** in $CD_2Cl_2$ exhibited a sharp singlet at room temperature (7.14 ppm), and even at –95°C, a slightly broadened peak was present, similar to those of **1a** and [12]CPP (Fig. 4D). This result clearly indicates that all benzene rings of **2** were equivalent on the NMR time scale even at –95°C because of the rapid dynamic motion of the molecule (see fig. S7 for simulated NMR spectra). The dynamic behavior of **2** was simulated by density-functional tight-binding (DFTB) with molecular dynamics (MD) methods. In this simulation, we observed the intrinsic dynamics whereby the paraphenylene chains of **2** coil around a thin torus like the typical motion of trefoil-knot vortices (*36*). Because this dynamic motion seamlessly shifts the benzene rings (orange in Fig. 4E) from the undercrossing to the overcrossing regions of the knot (see movie S1), it plausibly accounts for the averaged NMR signals of **2**.

These catenane and trefoil-knot molecules represent the cornerstone objects for topological molecular nanocarbons. The properties of these π-conjugated molecules raise further scientific questions such as how the structural features (size, components, substituents) affect their dynamic motions, physical properties, π-conjugation, and optoelectronic properties. We anticipate that this traceless synthetic method will generate a broader variety of topological molecular nanocarbons and open the door to expanded research areas in nanocarbon science.

## REFERENCES AND NOTES


1. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, R. E. Smalley, *Nature* **318**, 162–163 (1985).
2. S. Iijima, *Nature* **354**, 56–58 (1991).
3. K. S. Novoselov *et al*., *Science* **306**, 666–669 (2004).
4. W. E. Barth, R. G. Lawton, *J. Am. Chem. Soc.* **88**, 380–381 (1966).
5. L. T. Scott *et al*., *Science* **295**, 1500–1503 (2002).
6. H. Sakurai, T. Daiko, T. Hirao, *Science* **301**, 1878 (2003).
7. K. Tahara, Y. Tobe, *Chem. Rev.* **106**, 5274–5290 (2006).
8. Y. Segawa, A. Yagi, K. Matsui, K. Itami, *Angew. Chem. Int. Ed.* **55**, 5136–5158 (2016).
9. G. Povie, Y. Segawa, T. Nishihara, Y. Miyauchi, K. Itami, *Science* **356**, 172–175 (2017).
10. M. F. L. De Volder, S. H. Tawfick, R. H. Baughman, A. J. Hart, *Science* **339**, 535–539 (2013).
11. Y.-T. Wu, J. S. Siegel, in *Polyarenes i*, J. S. Siegel, Y.-T. Wu, Eds. (Springer, 2014), pp. 63–120.

12. Y. Segawa, H. Ito, K. Itami, . *Nat. Rev. Mater.* **1**, 15002 (2016).
13. B. I. Dunlap, *Phys. Rev. B* **46**, 1933–1936 (1992).
14. E. C. Kirby, R. B. Mallion, P. Pollak, *J. Chem. Soc., Faraday Trans.* **89**, 1945–1953 (1993).
15. L. Liu, F. Liu, J. Zhao, *Nano Res.* **7**, 626–657 (2014).
16. A. L. Mackay, H. Terrones, *Nature* **352**, 762 (1991).
17. J.-P. Sauvage, C. O. Dietrich-Buchecker, *Molecular Catenanes, Rotaxanes and Knots: A Journey Through the World of Molecular Topology* (Wiley, 2008).
18. R. S. Forgan, J.-P. Sauvage, J. F. Stoddart, *Chem. Rev.* **111**, 5434–5464 (2011).
19. C. J. Bruns, J. F. Stoddart, *The Nature of the Mechanical Bond: From Molecules to Machines* (Wiley, 2016).
20. S. D. P. Fielden, D. A. Leigh, S. L. Woltering, *Angew. Chem. Int. Ed.* **56**, 11166–11194 (2017).
21. W. P. Thurston, *Bull. Am. Math. Soc.* **6**, 357–382 (1982).
22. C. O. Dietrich-Buchecker, J.-P. Sauvage, J. P. Kintzinger, *Tetrahedron Lett.* **24**, 5095–5098 (1983).
23. P. R. Ashton *et al*., *Angew. Chem. Int. Ed. Engl.* **28**, 1396–1399 (1989).
24. M. Fujita, F. Ibukuro, H. Hagihara, K. Ogura, *Nature* **367**, 720–723 (1994).
25. C. A. Hunter, *J. Am. Chem. Soc.* **114**, 5303–5311 (1992).
26. G. Schill, N. Schweickert, H. Fritz, W. Vetter, *Angew. Chem. Int. Ed. Engl.* **22**, 889–891 (1983).
27. W. Zhang, A. Abdulkarim, F. E. Golling, H. J. Räder, K. Müllen, *Angew. Chem. Int. Ed.* **56**, 2645–2648 (2017).
28. Y.-Y. Fan *et al*., *Nat. Commun.* **9**, 3037 (2018).
29. J. M. Van Raden, B. M. White, L. N. Zakharov, R. Jasti, *Angew. Chem. Int. Ed.* **58**, 7341–7345 (2019).
30. H. Lenormand, J.-P. Goddard, L. Fensterbank, *Org. Lett.* **15**, 748–751 (2013).
31. E. R. Darzi, T. J. Sisto, R. Jasti, *J. Org. Chem.* **77**, 6624–6628 (2012).
32. G. W. Fester *et al*., *Inorg. Chem.* **49**, 2667–2673 (2010).
33. Y. Segawa *et al*., *Org. Biomol. Chem.* **10**, 5979–5984 (2012).
34. C. O. Dietrich-Buchecker, J.-P. Sauvage, *Angew. Chem. Int. Ed. Engl.* **28**, 189–192 (1989).
35. S. J. Tauber, *J. Res. Natl. Bur. Stand. Sect. A* **67A**, 591–599 (1963).
36. O. Velasco Fuentes, "Quasi-steady endless vortices with chaotic streamlines," in *Experimental and Computational Fluid Mechanics*, J. Klapp, A. Medina, Eds. (Springer, 2014), pp. 111–128.

## ACKNOWLEDGMENTS

We thank Rigaku Co. for x-ray analysis of **1a**, **1b**, and **2** and K. Kato, A. Miyazaki, D. R. Levine, and S. Ogi (Nagoya University) for assistance with experiments and fruitful advice. **Funding:** This work was supported by the ERATO program from JST (JPMJER1302 to K.I.), the Funding Program for KAKENHI from MEXT (JP19H05463 to K.I., JP16K05771 and JP19H02701 to Y.S., and JP17K14461 to Y.H.), a grant-in-aid for Scientific Research on Innovative Areas "π-Figuration" (JP17H05149 to Y.S.) and "Coordination Asymmetry" (JP17H05364 to Y.H.), and the Noguchi Institute (to Y.S). Computations were partially performed at the Research Center for Computational Science, Okazaki, Japan. ITbM is supported by the World Premier International Research Center Initiative (WPI), Japan. **Author contributions:** Y.S. and K.I. conceived the concept and directed the project. M.F. and N.K. attempted the synthesis of **1a** and modified the synthetic protocol. M.K. synthesized **1a**, **1b**, and **2**, and J.S. improved the synthesis. Y.S. performed x-ray crystallography and DFT calculations. Y.S. and T.N. conducted the photophysical measurements. Y.H. and J.P. performed the DFTB-MD calculations. Y.S. and K.I. prepared the manuscript with feedback from the other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Crystallographic data for compounds **1a**, **1b**, **2**, and **S4a** are available free of charge from the Cambridge Crystallographic Data Centre under CCDC identifiers 1860595–1860597 and 1908693 (www.ccdc.cam.ac.uk/structures/).

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/272/suppl/DC1
Materials and Methods
Figs. S1 to S29
Tables S1 to S3
References (*37*–*55*)
Movie S1

21 September 2018; resubmitted 14 May 2019
Accepted 11 June 2019
10.1126/science.aav5021

PALEONTOLOGY

# New Jurassic mammaliaform sheds light on early evolution of mammal-like hyoid bones

**Chang-Fu Zhou[1,2]*, Bhart-Anjan S. Bhullar[3]*, April I. Neander[4], Thomas Martin[5]†, Zhe-Xi Luo[4]†**

**We report a new Jurassic docodontan mammaliaform found in China that is preserved with the hyoid bones. Its basihyal, ceratohyal, epihyal, and thyrohyal bones have mobile joints and are arranged in a saddle-shaped configuration, as in the mobile linkage of the hyoid apparatus of extant mammals. These are fundamentally different from the simple hyoid rods of nonmammaliaform cynodonts, which were likely associated with a wide, nonmuscularized throat, as seen in extant reptiles. The hyoid apparatus provides a framework for the larynx and for the constricted, muscularized esophagus, crucial for transport and powered swallowing of the masticated food and liquid in extant mammals. These derived structural components of hyoids evolved among early diverging mammaliaforms, before the disconnection of the middle ear from the mandible in crown mammals.**

Early diverging mammaliaforms of the Mesozoic are the nearest predecessors to modern Mammalia (*1*) and provide direct fossil evidence of how the mammalian structures have evolved (*2*–*4*). One early mammaliaform group is the docodontans, with a wide distribution on Laurasian continents in the Jurassic and Cretaceous (*5*–*13*). We report here a newly discovered, exquisitely preserved docodontan skeleton that offers fresh insight on the transformation of hyoid bones and the earliest evolution of hyolingual function in mammals (*14*–*18*).

*Microdocodon gracilis*, gen. et sp. nov. (*16*) is from the Middle Jurassic and was found in the Daohugou localities; it is preserved in part and counterpart (holotype PMOL-AM00025A and B, respectively: Fig. 1, fig. S1, and movie S1). The fully erupted teeth (I?/i4?-C1/c1-P6/p6-M4/m4) indicate that it was an adult. Among docodontans, *Microdocodon* is phylogenetically nested in the Tegotheriidae (*16*), but differs from other tegotheriids by several dental characteristics (fig. S2) (*9*–*12*). Geologically, it is the oldest of the tegotheriid taxa (*16*).

This animal has a diminutive size with a body mass ranging from 5 g (as estimated from limb bone lengths) to 9 g (as estimated from skull length) (tables S1 and S2) (*16*). It is much smaller than other docodontans from the paleoecological community of Daohugou (*7*, *11*, *12*, *16*). Its skeleton shows a "T"-shaped interclavicle, curved clavicles, and strap-shaped scapulae, which are similar to those of other mammaliaforms (*16*). The girdle and limb bones are very slender, and the radius and ulna are elongate relative to the humerus. It has an exceptionally long tail: the postpelvic length of the caudal vertebrae is ~120% of the head–body length. On the basis of these characteristics, we interpret this to mean that it was a scansorial animal (figs. S1 and S10 to S12) (for details, see the supplementary materials) (*16*).

The middle ear and hyoids are preserved with the skull (Figs. 2 and 3, figs. S2 to S5, and movie S2). The middle ear is preserved in the postdentary trough of the mandible, as in other mammaliaforms (Figs. 2 and 3 and figs. S1 to S4) (*5*–*7*, *11*, *16*). The basihyal, thyrohyal, ceratohyal, and epihyal bones (distal part incomplete and possibly cartilaginous) are preserved near the base of the skull (figs. S3 and S4). The basihyal is a rod-like bone with two slightly enlarged ends that bear articulating facets for the paired thyrohyals and ceratohyals. These facets are similar to the articular facets of the basihyal–thyrohyal and basihyal–ceratohyal joints in the monotreme *Tachyglossus* (Fig. 3 and figs. S4 and S5). As in monotremes, the thyrohyals in *Microdocodon* have proximal ends wider than the shaft, whereas the distal ends are incomplete or unossified in the fossil. The right thyrohyal (best seen on the main part) overlaps the well-preserved manubrium of the right malleus (best seen on the counterpart). The ceratohyals are slender rods. The ceratohyal–epihyal joint was mobile and forms a sharp angle. We interpret this to mean that the basihyal, thyrohyals, and ceratohyals formed a "⟩—⟨" configuration, in which the basihyal was the transverse strut (Figs. 2 and 3 and movie S2). This configuration, herein called the saddle-shaped configuration, is typical of the basal parts (basihyal, ceratohyals, thyrohyals) of the hyoid apparatus in monotremes and placentals, and in a modified condition in marsupials (for details, see the supplementary materials) (*16*–*18*).

Based on the distinctly mammal-like morphology of the hyoid elements of *Microdocodon*, we can now identify hyoid elements of several other mammaliaforms (*16*) (figs. S4 to S9). The Jurassic haramiyidan *Vilevolodon* has preserved basihyal and ceratohyal bones (fig. S6 and movie S2) (*19*). The Cretaceous eutriconodontan *Yanoconodon* also has preserved hyoid elements (fig. S6) (*20*). Computed tomography (CT) scanning has revealed the basihyal, thyrohyals, ceratohyals, and epihyals of the trechnotherian *Maotherium* and the multituberculate *Sinobaatar* of the Cretaceous (figs. S7 and S8 and movie S2) (*16*). Additionally, *Sinobaatar* has preserved stylohyal bones (fig. S7), which is similar to the Cretaceous multituberculate *Kryptobaatar* (*21*). We interpret this to mean that multituberculates have an integrocornuate anterior hyoid cornu, characterized by the complete series of ceratohyal–epihyal–stylohyal connected to the basicranium (movie S2) (*16*). Moreover, we now can identify the hyoid elements in the early eutherian *Eomaia* (fig. S9) (*16*), which are similar to those of the recently discovered eutherian *Ambolestes* (*22*).

The newly identified hyoids in the early mammaliaform *Microdocodon* (fig. S4 and movie S2) show the rodlike basihyal, thyrohyals, ceratohyals, and epihyals, which are similar to those of *Yanoconodon*, *Sinobaatar*, *Maotherium*, and *Eomaia* (figs. S6 to S9), in an extant mammal-like configuration. Therefore, the hyoids of *Microdocodon* represent the ancestral characters of hyoid apparatus for the clade of *Microdocodon* and crown mammals. Phylogenetically, *Microdocodon* and *Vilevolodon* are the earliest-known mammaliaform fossils with mammal-like hyoids (Fig. 3, fig. S5, and movie S2).

Although the hyoids are mammal like, the middle ear in docodontans is fully attached in the postdentary trough of the mandible (Fig. 3B), which is also present in haramiyidans, albeit in a modified condition (Fig. 3C) (*19*). The hyoids are also mammal like in eutriconodontans and spalacotherioids, whereas the middle ears in these animals are still connected by an ossified Meckel's element to the mandible (Fig. 3, fig. S6, and movie S2). The mammal-like, jointed anterior hyoid cornu and the saddle-like basal hyoid structure (Fig. 3) evolved before the separation of the middle ear from the mandible by resorption of Meckel's cartilage, which occurred convergently in separate clades of crown Mammalia. The development of a mobile linkage of the mammal-like hyoids is decoupled from transformation of the ear bones in articulation with the mandible into the separated middle ear during mammaliaform evolution (Fig. 3).

Mammaliaform hyoids show neomorphic characters not developed in nonmammaliaform therapsids (Fig. 2). (i) The basihyal is ossified and forms a transverse strut in a saddle configuration with the ceratohyals, and with the thyrohyals that connect the hyoids to the thyroid cartilage

[1]Paleontological Museum of Liaoning, Shenyang Normal University, Shenyang Liaoning, 110034, China. [2]College of Earth Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China. [3]Department of Geology and Geophysics and Peabody Museum of Natural History, Yale University, New Haven, CT 06511, USA. [4]Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, IL 60637, USA. [5]Section Paleontology, Institute of Geosciences, Rheinische Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany.
*These authors contributed equally to this work.
**†Corresponding author. Email: tmartin@uni-bonn.de (T.M.); zxluo@uchicago.edu (Z.-X.L.)**

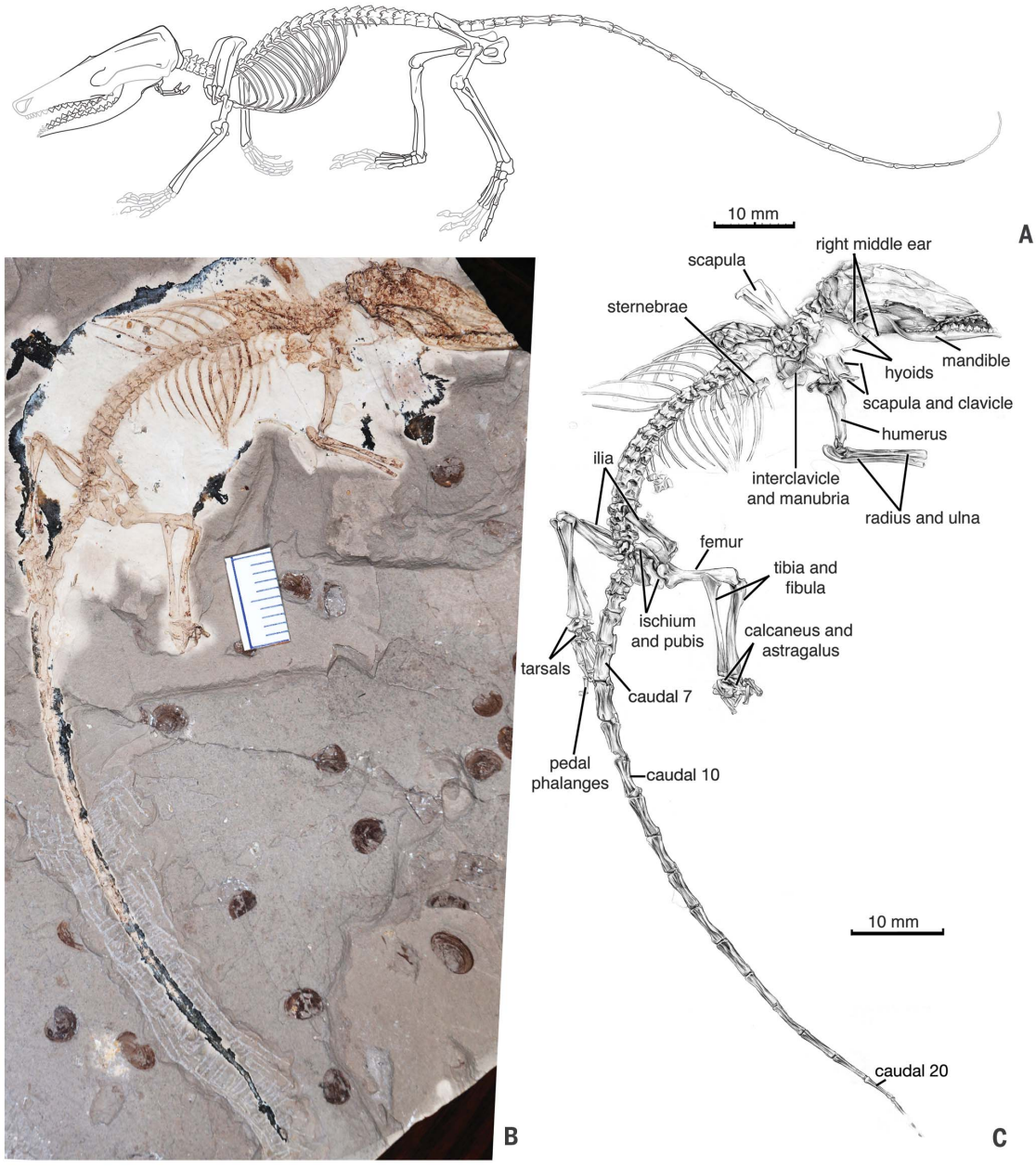



**Fig. 1. Mammaliaform *Microdocodon gracilis*.** (**A**) Skeletal reconstruction, (**B**) photo, and (**C**) illustration of holotype main part PMOL-AM00025A. PMOL-AM00025B is shown in fig. S1 and movie S1, mandible and dentition in fig. S2, and middle ear and hyoids in the details in figs. S3 and S4 and movie S2.

(*14*, *16*). By contrast, the basihyal and thyrohyals are not ossified (unknown as fossils) in nonmammaliaform therapsids (*2*, *24*–*26*) (for details, see the supplementary materials (*16*). (ii) The basihyal, ceratohyal, and epihyal in the anterior hyoid cornu (which hypothetically also includes the stylohyal) form a jointed linkage to the crista parotica of the petrosal of mammaliaforms (Fig. 3 and fig. S5) (*16*, *23*). However, in nonmammaliaform therapsids, the hyoids are a pair of long and unsegmented rods that lack the flexibility of the anterior hyoid cornu of mammals (*24*–*26*). (iii) The jointed anterior hyoid cornu swings into a posterior posture in mammaliaforms, enabled by the angled joints of the basihyal–ceratohyal and the ceratohyal–epihyal (Figs. 2 and 3 and fig. S5). This posterior posture is necessary for the basihyal–thyrohyal to cradle the thyroid cartilage and the larynx for a narrow and muscularized pharynx, creating an oropharyngeal space in which to swallow the masticated food, as seen in modern mammals (*3*, *14*, *27*). By contrast, in nonmammaliaform therapsids, the long hyoid rods are situated anteriorly in an "A-shaped" configuration, as seen in early synapsids (*28*) and in extant nonmammalian amniotes (*15*, *27*).

In extant mammals, under the control of the suprahyoid and infrahyoid musculature, the mobile linkage of the jointed anterior hyoid cornu plays a crucial role in cyclic movement of the

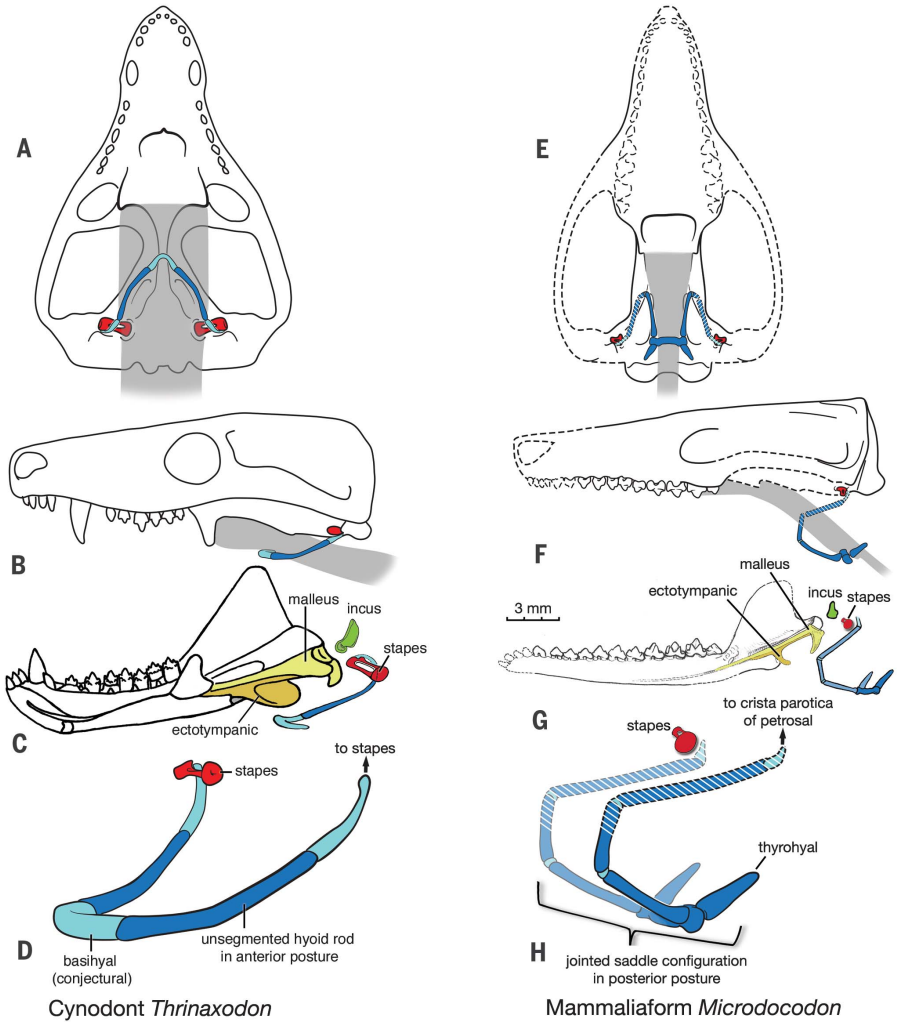



**Fig. 2. Apomorphic hyoid characteristics of mammaliaforms compared with nonmammaliaform cynodonts.** (**A** to **D**) Cynodont plesiomorphies of *Thrinaxodon*: "A"-shaped configuration of unsegmented hyoid rods in anterior hyoid posture; basihyal and thyrohyals unossified, as in therapsids. The hyoid–stapes connection in *Thrinaxodon* is based on comparison with the hyoid–stapes connection in other synapsids (*2*, *24*). [(A) and (B)] Skull with hyoid–stapes in ventral and lateral views (for details, see movie S2). Gray shading indicates a wide pharyngeal passage conforming to hyoid rods, as in extant reptiles (*27*). (C) Reconstructed hyoids and middle ear bones of cynodonts; the incus and hyoids are separated to show their relationships. (D) Reconstructed hyoid and stapes: hypothetical cartilaginous connection (light blue) to stapes (red) [following (*2*)]; the cartilaginous basihyal is conjectural (*16*). (**E** to **H**) Mammaliaform apomorphies of *Microdocodon*: the hyoids form a rectangular and saddle-shaped configuration with a transverse basihyal. The hyoid anterior cornu forms a mobile linkage of the jointed basihyal, ceratohyal, and epihyal (incomplete). The hyoid apparatus in a posterior hyoid posture, all like the crown Mammalia (Fig. 3, figs. S3 and S4, and movie S2). The lack of full ossification of epi-stylohyals (dashed lines) in *Microdocodon* is consistent with the condition of their homologs in monotremes, all marsupials, eutriconodontans, and spalacotherioids, but not multituberculates and the early eutherians (*16*). [(E) and (F)] Skull in ventral and lateral views. Gray shading indicates a muscularized pharyngeal passage, constrained by rectangle-shaped hyoids, as in extant mammals (*14*, *27*). (G) Mandible and hyoids in medial view. (H) Reconstructed hyoid and stapes: the soft tissues in hyoid joints are conjectural. Missing parts of the epihyals, stylohyals, and tympanohyals are conjectural (hatched pattern). Separation of the stapes from the hyoid contact to the crista parotica of petrosal is based on *Haldanodon* (*23*).

tongue and the larynx simultaneously during mastication. After mastication, this apparatus is also responsible for transport and swallowing of the masticated food (*3*, *14*, *29*, *30*) and for stabilizing the hyolingual structure by maintaining a dynamic hyoid posture (*31*). The mobile linkage and the saddle-shaped basal hyoid configuration of the hyoid apparatus can now be traced to *Microdocodon* (also *Vilevolodon*) in synapsid phylogeny (Figs. 2 and 3 and figs. S5 and S6). Docodontans and haramiyidans have more complex teeth for mastication than does *Morganucodon* and most of nonmammaliaform therapsids; they also have the hyoid apparatus and related hyolingual function of transport and swallowing of the masticated food and liquid. It is likely that the mammal-like hyoid structure evolved with the first appearance of complex mastication in early divergent mammaliaforms. We hypothesize that this system evolved no later than the common

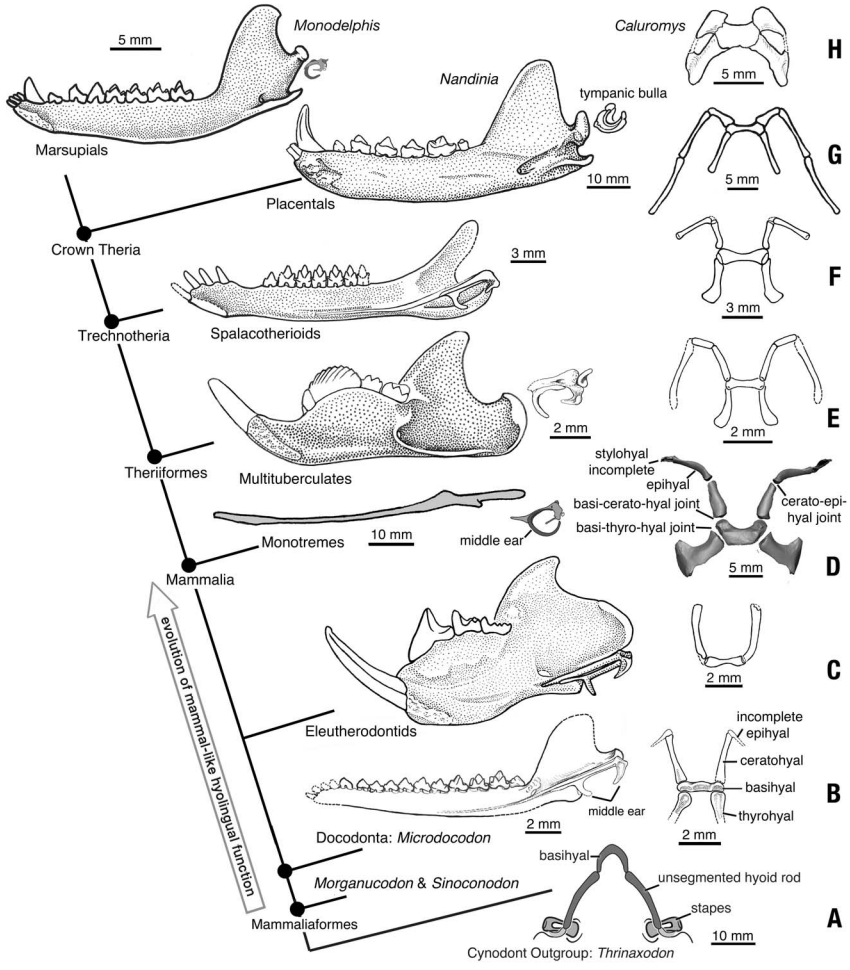



**Fig. 3. Mosaic evolution of the hyoid structure and the middle ear in mammaliaform phylogeny.** The jointed anterior hyoid cornu from embryonic pharyngeal arch 2 had evolved into mammal-like structures before the separation of the middle ear bones from the mandible, both derived from (or in association with) pharyngeal arch 1. The mammaliform *Microdocodon* has mammal-like hyoids with a saddle-shaped formation in posterior posture, with apomorphies of a transverse basihyal, and mobile joints between basi-thyrohyals, basi-ceratohyals, and cerato-epihyals. This indicates that the transformation to mammal-like hyoids and hyolingual function occurred no later than the common ancestor of crown mammals and docodontans. The hyoids, as exemplified by *Microdocodon*, are also partly preserved in other early mammaliaforms. The first appearance of mammal-like hyoids preceded, phylogenetically, the separation of middle ear bones from the dentary, which occurred later and convergently in monotremes, multituberculates, and crown therians. (**A**) Cynodont *Thrinaxodon* (based on UCMP40446, see also movie S2); basihyal is conjectural. (**B**) Mandible, middle ear, and preserved hyoid parts of *Microdocodon* (details in figs. S1 to S4 and movie S2). (**C**) Mammaliaform *Vilevolodon* (an eleutherodontid from the Late Jurassic) (fig. S6) (*20*). (**D**) Monotreme *Tachyglossus*, hyoids digitally disarticulated, with thyroid arches omitted to facilitate comparison (details in figs. S3 and S4). (**E**) Multituberculate *Sinobaatar* with reconstruction of hyoids (fig. S7 and movie S2) [for more details, see (*16*)]. (**F**) Cretaceous trechnotherian *Maotherium*, also shown in a CT rendering of hyoids and middle ear (JZT-B0064; fig. S8 and movie S2). (**G**) Extant carnivoran *Nandinia* [redrawn from (*18*)]. (**H**) Marsupials *Monodelphis* (mandible and middle ear) and *Caluromys* (hyoids, FMNH60697). Mandibles (left on cladogram) and middle ears are shown in medial view. Hyoids (right column) are shown in splayed ventral view or digitally disarticulated in ventral view.

ancestor of docodontans and crown Mammalia (Fig. 3).

## REFERENCES AND NOTES


1. T. B. Rowe, *J. Vertebr. Paleontol.* **8**, 241–264 (1988).
2. E. F. Allin, J. A. Hopson, in *The Evolutionary Biology of Hearing*, D. B. Webster, R. R. Fay, A. N. Popper, Eds. (Springer-Verlag, 1992), pp. 587–614.
3. A. W. Crompton, in *Functional Morphology in Vertebrate Paleontology*, J. J. Thomason, Ed. (Cambridge Univ. Press, 1995), pp. 55–75.
4. Z.-X. Luo, *Nature* **450**, 1011–1019 (2007).
5. J. A. Lillegraven, G. Krusat, *Contrib. Geol. Univ. Wyoming* **28**, 39–138 (1991).
6. Z. Kielan-Jaworowska, R. L. Cifelli, Z.-X. Luo, *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure*. (Columbia Univ. Press, 2004).
7. Q. Ji, Z.-X. Luo, C.-X. Yuan, A. R. Tabrum, *Science* **311**, 1123–1127 (2006).
8. Z.-X. Luo, T. Martin, *Bulletin of the Carnegie Museum of Natural History* **39**, 27–47 (2007).
9. A. O. Averianov, A. V. Lopatin, S. A. Krasnolutskii, S. V. Ivantsov, *Proceedings of the Zoological Institute of the Russian Academy of Sciences* **314**, 121–148 (2010).
10. T. Martin, A. O. Averianov, H.-U. Pfretzschner, *Palaeobiodivers. Palaeoenviron.* **90**, 295–319 (2010).
11. Q.-J. Meng *et al*., *Science* **347**, 764–768 (2015).
12. Z.-X. Luo *et al*., *Science* **347**, 760–764 (2015).
13. J. A. Schultz, B.-A. S. Bhullar, Z.-X. Luo, *J. Mamm. Evol.* **26**, 9–38 (2019).
14. K. M. Hiiemae, A. W. Crompton, in *Functional Vertebrate Morphology*, M. Hildebrand, D. M. Bramble, K. F. Liem, D. B. Wake, Eds. (Harvard Univ. Press, 1985), pp. 262–290.
15. W. Maier, *Zool. Anz.* **338**, 55–74 (1999).
16. See the supplementary materials.
17. J.-P. Gasc, Squelette hyobranchial, in *Traité de Zoologie Anatomie, Systématique, Biologie*, Tome XVI, Fascicle I, P.-P. Grassé, Ed. (Masson, 1967), pp. 550–583; 1103–1106.
18. J. R. Wible, M. Spaulding, *Ann. Carnegie Mus.* **82**, 1–114 (2013).
19. Z.-X. Luo *et al*., *Nature* **548**, 326–329 (2017).
20. Z.-X. Luo, P. Chen, G. Li, M. Chen, *Nature* **446**, 288–293 (2007).
21. J. R. Wible, G. W. Rougier, *Bull. Am. Mus. Nat. Hist.* **247**, 1–124 (2000).
22. S. Bi *et al*., *Nature* **558**, 390–395 (2018).
23. I. Ruf, Z.-X. Luo, T. Martin, *J. Vertebr. Paleontol.* **33**, 382–400 (2013).
24. T. H. Barry, *Annals of the South African Museum* **50**, 275–281 (1968).
25. S. C. Jasinoski, F. Abdala, V. Fernandez, *Anat. Rec.* **298**, 1440–1464 (2015).
26. A. K. Huttenlocker, C. A. Sidor, *J. Vertebr. Paleontol.* **36**, e1111897 (2016).
27. V. E. Negus, *The Comparative Anatomy and Physiology of the Larynx* (Heinemann, 1949).
28. E. C. Olson, *Fieldiana Geol.* **17**, 225–349 (1968).
29. S. W. Herring, R. P. Scapino, *J. Morphol.* **141**, 427–460 (1973).
30. R. Z. German, A. W. Crompton, *Brain Behav. Evol.* **48**, 157–164 (1996).
31. R. Z. German *et al*., *Dysphagia* **26**, 97–98 (2011).


## ACKNOWLEDGMENTS


For full acknowledgments, please see the supplementary materials. **Author contributions**: Acquisition of fossils: C.-F.Z.; laboratory study of fossils: T.M., C.-F.Z., and Z.-X.L.; comparative data acquisition: B.-A.S.B., C.-F.Z., A.I.N., and Z.-X.L.; CT analysis and graphic presentation: A.I.N.; development of comparative analysis: B.-A.S.B., Z.-X.L.; taxonomy and phylogenetics: Z.-X.L., T.M.; logistic coordination: T.M. and Z.-X.L.; project design: Z.-X.L., T.M., C.-F.Z., and B.-A.S.B. **Competing interests**: The authors declare no competing interests. **Data and materials availability**: The holotype is accessioned in the Paleontological Museum of Liaoning (PMOL), China; comparative data and phylogenetic datasets are provided in the supplementary materials.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/276/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S3
References (*32–151*)
Movies S1 and S2


29 July 2018; accepted 12 June 2019
10.1126/science.aau9345

NATURAL PRODUCTS

# Use of a scaffold peptide in the biosynthesis of amino acid–derived natural products

Chi P. Ting[1], Michael A. Funk[2]*, Steve L. Halaby[3,4], Zhengan Zhang[2], Tamir Gonen[3,4]†, Wilfred A. van der Donk[1,2,5]†

**Genome sequencing of environmental bacteria allows identification of biosynthetic gene clusters encoding unusual combinations of enzymes that produce unknown natural products. We identified a pathway in which a ribosomally synthesized small peptide serves as a scaffold for nonribosomal peptide extension and chemical modification. Amino acids are transferred to the carboxyl terminus of the peptide through adenosine triphosphate and amino acyl-tRNA–dependent chemistry that is independent of the ribosome. Oxidative rearrangement, carboxymethylation, and proteolysis of a terminal cysteine yields an amino acid–derived small molecule. Microcrystal electron diffraction demonstrates that the resulting product is isosteric to glutamate. We show that a similar peptide extension is used during the biosynthesis of the ammosamides, which are cytotoxic pyrroloquinoline alkaloids. These results suggest an alternative paradigm for biosynthesis of amino acid–derived natural products.**

Bacteria produce many small-molecule natural products that play important roles in communication, symbiosis, and competition (*1*). Historically, these compounds have been discovered by activity-based screens. An alternative avenue for their discovery starts with identification of their biosynthetic gene clusters, now that bacterial genomes have revealed the tremendous diversity of natural products that remain to be discovered (*2*). In this study we focus on a group of biosynthetic gene clusters for which the final products were not known and could not be predicted.

Ribosomally synthesized and posttranslationally modified peptides (RiPPs) (*3*) include lantibiotics and thiopeptides that are used in food and agriculture (*4*). They are biosynthesized from a precursor peptide consisting of a leader peptide that serves as a recognition motif for the biosynthetic enzymes and a core peptide that is converted to the final product. During their maturation, Ser and Thr residues are glutamylated by LanB enzymes through a glutamyl-tRNA–dependent mechanism (*5*, *6*). Subsequently, the glutamate is eliminated to generate dehydroamino acids (Fig. 1A). A survey of >100,000 publicly available bacterial genomes revealed more than 600 genes that encode LanB-like proteins in which the elimination domain is not present within the cluster or genome.

In the plant pathogen *Pseudomonas syringae* pv. maculicola ES4326, such a protein (TglB) is encoded near an open reading frame for a 50-amino-acid peptide (TglA) (Fig. 1B). Coexpression of $His_6$-TglA and -TglB in *Escherichia coli* and subsequent purification of the peptide demonstrated an increase in mass by 103 Da (Fig. 1C). This increase is inconsistent with glutamylation but could be the result of condensation with a cysteine residue. High-resolution tandem mass spectrometry (MS/MS) analysis of the peptide suggested that the adduct was attached to the C-terminal alanine instead of the anticipated ester linkage to a Ser in the peptide (Fig. 1D). We expressed TglA and TglB individually as $His_6$-tagged proteins and purified them. In vitro incubation with Cys, ATP, $tRNA^{Cys}$, and Cys tRNA synthetase (CysRS) resulted in the same product (TglA-Cys) (Fig. 2A) as that isolated from coexpression in *E. coli*, confirming that TglB adds a Cys to the C terminus of TglA in a tRNA-dependent manner (fig. S1A). This C-terminal peptide extension not only constitutes a previously unknown posttranslational modification but also seems counterintuitive, because a more logical route to the product would entail a Cys encoded by *tglA*. We next purified Cys-$tRNA^{Cys}$ and showed that TglB does not transfer the Cys to the C terminus of TglA unless ATP is present, which is converted to ADP and phosphate (fig. S1B). Performing the reaction in buffer made with $H_2{}^{18}O$ and subsequent MS analysis demonstrated that the product contains one $^{18}O$ atom (fig. S1C), and addition of hydroxyl amine to the assay mixture allowed trapping of C-terminally activated TglA as the hydroxamate (fig. S1D). These findings are consistent with activation of the C terminus of TglA by phosphorylation, subsequent amide bond formation with the amino group of Cys-tRNA, and release of the tRNA by hydrolysis (fig. S1E, mechanism 1). The observations rule out the use of the activated ester of Cys-tRNA for the nonribosomal peptide extension (fig. S1E, mechanism 2). TglB accepted a 12-mer peptide corresponding to the C terminus of TglA as a minimal substrate (fig. S1F), and kinetic experiments showed that TglB has a turnover of 28 $min^{-1}$ in the presence of full-length TglA (fig. S1G).

We next interrogated the other proteins encoded in the biosynthetic gene cluster. TglH has low homology to a structurally characterized dinuclear nonheme iron–dependent protein for which no activity has been reported (*7*). The C-terminal domain of TglI has homology with known leader peptide-binding domains in RiPP biosynthetic enzymes (Fig. 1B) (*5*, *8*). We coexpressed TglA with TglB, TglH, and TglI in *E. coli* and isolated a product that was decreased in mass by 14 Da from TglA-Cys (fig. S2A). We treated the peptide with trypsin to generate a C-terminal tetrapeptide. Chemical assays with thiol- and carboxylate-reactive electrophiles indicated that the product still contained these functional groups (fig. S3), suggesting structure **1** as the product of TglHI (Fig. 2A). We next repeated this experiment but used an *E. coli* strain that is auxotrophic for Cys and that was grown in minimal medium supplemented with $^{13}C$-labeled Cys. Isolation of the peptide and analysis by MS showed that it is the cysteine β carbon that is removed (fig. S4).

The biosynthetic cluster also contains a pair of genes (*tglEF*) encoding proteins similar to a recently characterized carboxy-*S*-adenosylmethionine (Cx-SAM) synthase and a Cx-SAM–dependent methyltransferase, respectively (*9*, *10*). We added compound **1** to Cx-SAM and TglF in vitro and isolated product **2**, with a mass increase of 58 Da (Fig. 2C), consistent with carboxymethylation of a thiol. This hypothesis was confirmed by treating the TglHI product with iodoacetic acid, which resulted in the same outcome, as did coexpression of TglABEFHI in *E. coli* (fig. S2B). The in vitro–prepared peptide was treated with trypsin and the C-terminal tetrapeptide **3** was characterized by $^1H$ nuclear magnetic resonance (NMR) spectroscopy and tandem MS, which supported structure **2** for the TglF product (Fig. 2A and fig. S5). Given the unusual architecture, we also chemically synthesized peptide **3** as two diastereomers (supplementary materials and methods) and demonstrated that the $^1H$ NMR spectrum of one isomer was identical to that of the enzymatic product (fig. S5). We tried to obtain crystals to assign the stereochemistry of either isomer and made several chemical derivatives but were unable to obtain crystals for X-ray diffraction.

We next turned to the cryo–electron microscopy (cryo-EM) method microcrystal electron diffraction (MicroED) (*11*–*13*). A small amount of powder of the diastereomer that eluted first during high-performance liquid chromatography purification was placed onto an EM grid, plunged into liquid nitrogen, and investigated

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles CA 90095, USA. [4]Departments of Biological Chemistry and Physiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles CA 90095, USA. [5]Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA.
*Present address: American Association for the Advancement of Science, Washington, DC, USA.
†Corresponding author. Email: tgonen@ucla.edu (T.G.); vddonk@illinois.edu (W.A.v.d.D.)

under cryogenic conditions in an electron microscope. The seemingly amorphous powder contained numerous nanocrystals on the grid that were suitable for MicroED analysis, each consisting roughly of femtograms of material that diffracted to ~1-Å resolution. MicroED data were collected from each nanocrystal, but the sample was highly susceptible to beam damage such that no useful diffraction was observed after the first few frames of the MicroED movie. Despite collection of >150 datasets on a complementary metal oxide semiconductor-based CetaD camera, nanocrystals succumbed to radiation damage too fast, preventing structure determination. It is possible that the peptide was particularly susceptible to damage because of the 3-thiaglutamate, consistent with an earlier study that showed that radiation damage is particularly prevalent at Cys residues (*14*). We then turned to the Falcon III direct electron detector, one of the most sensitive cameras for cryo-EM that was recently demonstrated to be suitable for MicroED data collection and structure determination and that minimizes radiation damage because of its high sensitivity and high frame rate (*15*). Atomic resolution data from seven nanocrystals were collected, each covering an angular range of ~50° before damage was observed. Data from five nanocrystals were merged to yield a 96% complete dataset to 1.0-Å resolution, and the structure was determined by direct methods (Fig. 3; PDB 6PO6; EMD-20411 crystallographic data in table S3 and supplementary materials and methods). The atomic resolution MicroED structure revealed the D configuration of the 3-thiaGlu in this peptide (D-***3***), which in turn provided the stereochemical assignment for L-***3***, which coelutes with and has the same spectral data as the enzymatic product. These results demonstrate that the TglHI-catalyzed reaction occurred with retention of configuration at the α carbon (Fig. 3, B to D). These findings highlight the utility of MicroED to determine the structure and stereochemistry of a previously unknown natural product. Thus, collectively, TglBEFHI convert TglA into a peptide containing L-3-thiaglutamate at its C terminus (TglA-thiaGlu; **2** in Fig. 2A).

We next investigated the TglHI-catalyzed reaction with purified proteins. Neither protein could be expressed in soluble form individually, but coexpression resulted in copurification and metal analysis indicated TglHI contained 2.5 Fe. In vitro TglHI converted TglA-Cys to **1** under aerobic conditions with a turnover of 1.1 $min^{-1}$ (Fig. 2B), whereas under low oxygen concentrations product formation was negligible, confirming oxygen dependency of the reaction (fig. S6A). To investigate if TglHI can functionalize internal cysteine residues, the extension mutant TglA-CysAla was prepared. This peptide was not modified by TglHI (fig. S6B). TglHI also did not modify other unrelated peptides that end in Cys (fig. S6C), and N-terminal truncation of TglA-Cys led to diminished or abolished TglHI activity (fig. S6D). Thus, the enzyme has high specificity for TglA-Cys. To identify the fate of the lost carbon atom, $^{13}C$-labeled TglA-Cys was reacted with TglHI, and formate was observed by $^{13}C$ NMR spectroscopy (Fig. 3E). Moreover, when [2,3,3,$^{2}H^{3}$]Cys was used, the product contained one deuterium, illustrating that the α hydrogen is likely not removed during the transformation (fig. S4D). Thus, TglHI catalyzes a net four-electron oxidation of TglA-Cys, modifying the redox states of both the α and β carbons of the C-terminal cysteine installed by TglB. Based on the in vitro studies, we propose a mechanism for the formation of **1** and formate from TglA-Cys (fig. S7). The chemistry catalyzed by TglHI expands the range of posttranslational modifications in natural product biosynthesis (*16*) to include a remarkable excision of a methylene group from cysteine. Additional TglHI-like enzymes are encoded in the genomes (fig. S8), including in the biosynthetic gene cluster for methanobactin (*17*–*19*).

The last four genes in the biosynthetic cluster encode a putative membrane-bound protease (TglG), a putative pyridoxal-phosphate–dependent

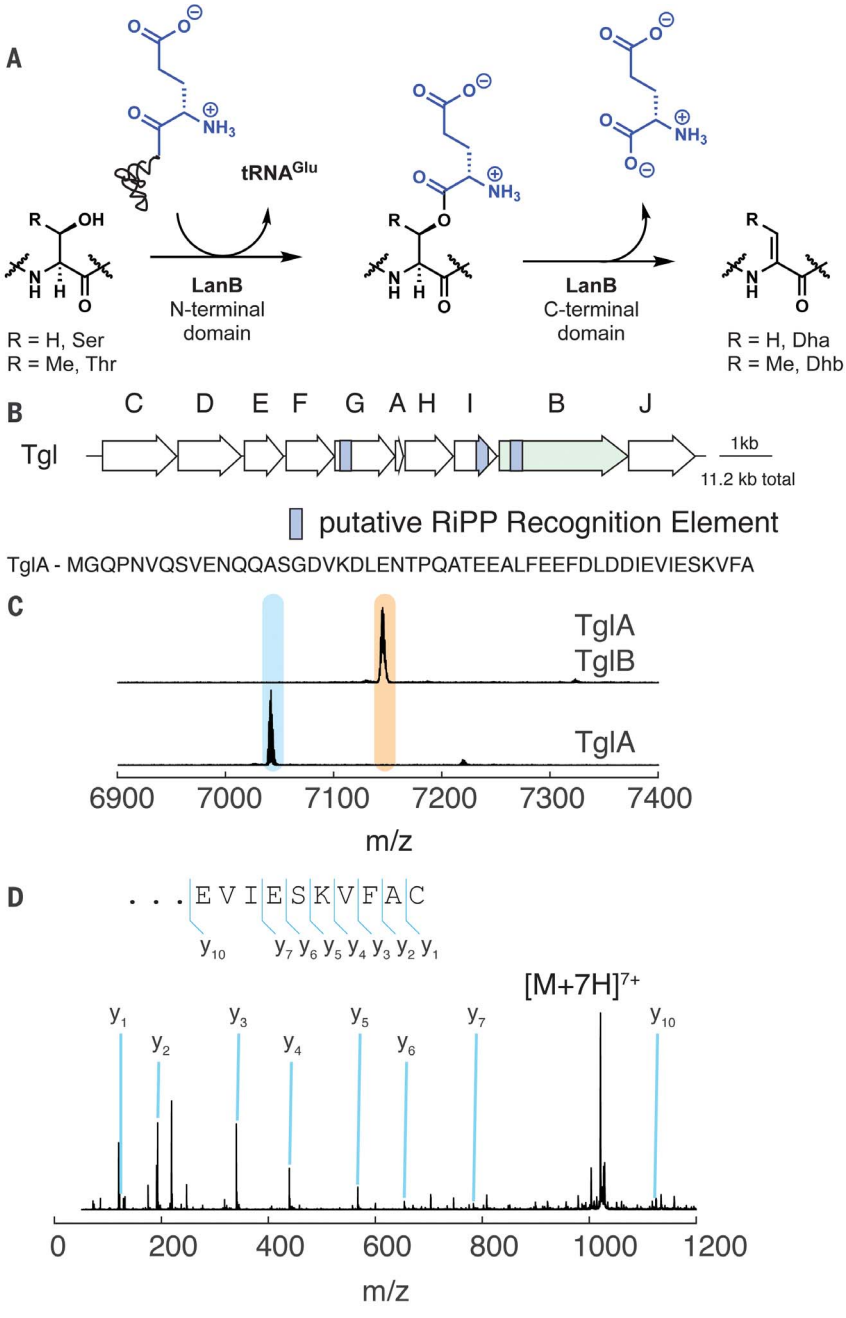



**Fig. 1. Function of a small LanB enzyme, TglB, found in *P. syringae*.** (**A**) LanB enzymes glutamylate Ser/Thr residues and subsequently eliminate the glutamate to form dehydroamino acids. Small LanB proteins lack the elimination domain. Dha, dehydroalanine; Dhb, dehydrobutyrine. (**B**) Biosynthetic gene cluster in *P. syringae* that encodes a small LanB (TglB). (**C**) Matrix-assisted laser desorption ionization with time-of-flight (MALDI-TOF) mass spectra of TglA coexpressed with TglB. (**D**) Analysis of the TglB product by electrospray ionization–tandem MS (ESI-MS/MS).

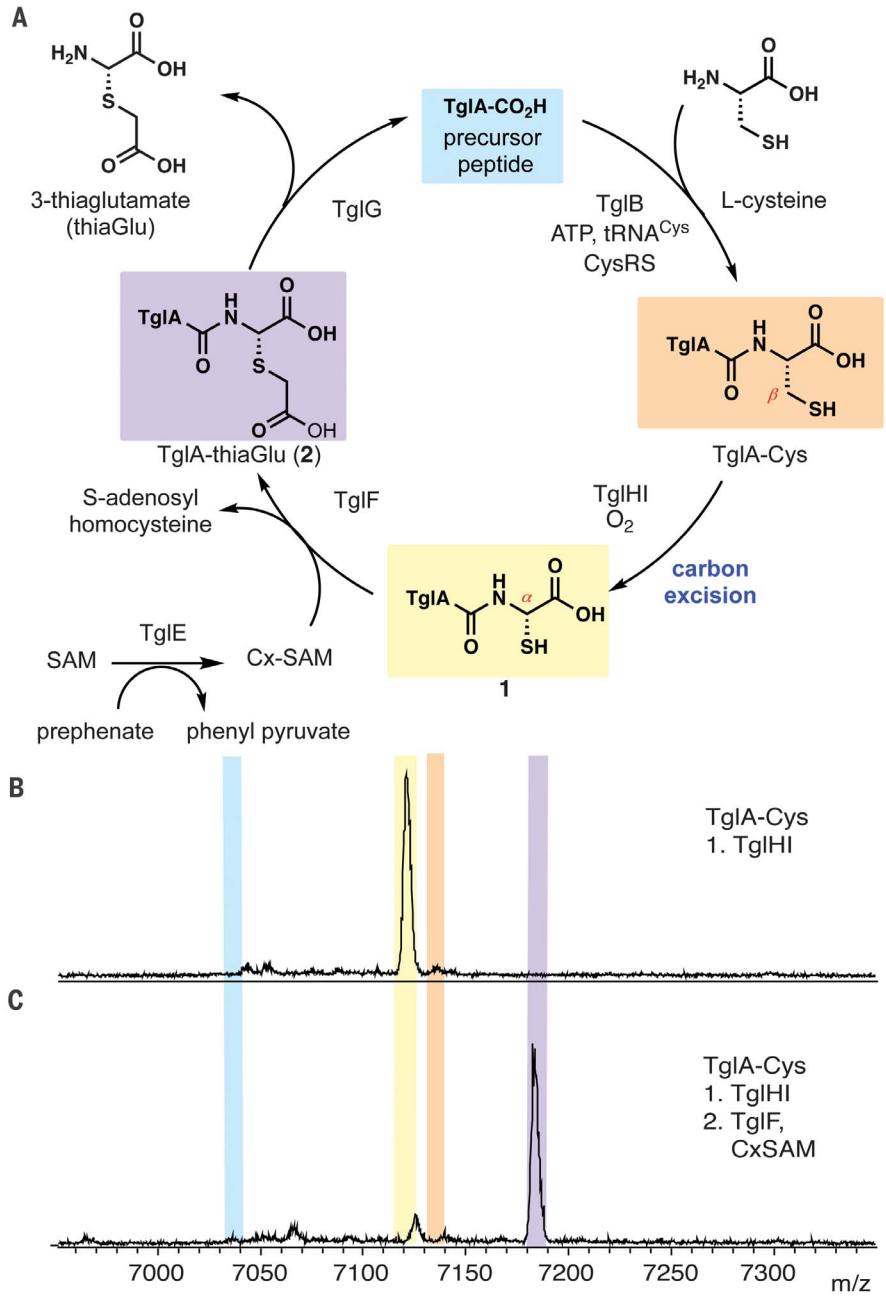



**Fig. 2. The cysteine added by TglB is modified by other enzymes encoded by the *tgl* cluster.** (**A**) Inferred biosynthetic pathway toward 3-thiaglutamate. (**B**) MALDI-TOF mass spectrum of in vitro reaction of TglHI with TglA-Cys. (**C**) MALDI-TOF mass spectrum of in vitro reaction of TglF with compound **1**. Color coding of shaded peaks in (B) and (C) is shown in (A).

enzyme (TglC) that is sometimes missing in homologous clusters, and two putative transporters (TglD and TglJ). Like TglB and TglI, TglG contains a RiPP leader peptide recognition motif, suggesting that it will act on a TglA-derived peptide (Fig. 1B), and homologous enzymes have cytoplasmic active sites (*20*). When TglA-thiaGlu was exposed to the membrane fraction of cell lysate of *E. coli* expressing green fluorescent protein–tagged TglG, the peptide was cleanly converted into TglA (fig. S9). TglA-Glu was also a substrate, but TglA-GluAla was not, illustrating that the protease cannot distinguish Glu and 3-thiaGlu but does not tolerate extension of the peptide. Thus, TglA appears to be a scaffold on which 3-thiaGlu is assembled and final proteolytic release regenerates TglA for another round of biosynthesis (Fig. 2A). Were cysteine merely encoded by *tglA*, then each ribosomally produced peptide could make only a single 3-thiaglutamate. Instead, the use of TglA as a scaffold peptide is conceptually more efficient than the stoichiometric use of leader peptide in other RiPP pathways (*4*). At present we do not know the function of 3-thiaGlu, nor whether this unstable compound is further chemically modified. Plants were recently shown to use Glu for a systemic signaling response to pathogens (*21*), and it is possible that 3-thiaGlu or a product derived from it interferes with Glu signaling similarly to other antimetabolite toxins made by *P. syringae* that block jasmonate and ethylene signaling pathways (*22*).

We note that 3-thiaGlu is not a RiPP, because it is not ribosomally synthesized, but it is made by posttranslational modification reactions. Perhaps this unusual pathway evolved because of the significant relative burden of leader peptide production for a single amino acid product. Bioinformatic prediction of TglA transcriptional regulation (*23*) suggests that precursor production is not driven by a separate promoter, which is consistent with putative catalytic use of the peptide (fig. S10). This contrasts with most RiPP pathways in which expression of the substrate peptide is controlled by its own promoter followed by a readthrough transcriptional terminator to allow the precursor peptide to be present in excess over the biosynthetic machinery (*24*, *25*).

It is the Cys-tRNA–dependent enzyme TglB that allows the proposed catalytic use of TglA. Similar small LanB-encoding genes are found in several bacterial phyla, with some clusters encoding multiple such proteins and a range of additional putative modification enzymes (fig. S11). To assess the generality of the function of small LanB proteins and provide further support for a catalytic role of the scaffold peptide, we investigated ammosamide biosynthesis. A previous study of these Trp-derived pyrroloquinoline natural products (Fig. 4A) hinted that the compounds could be derived from a small peptide, AmmA ending in Trp, encoded in the gene cluster (Fig. 4B) (*26*). However, when this Trp was mutated to Ser or deleted altogether, ammosamide was still produced (*26*). The ammosamide gene cluster encodes four small LanB proteins. We tested all four for activity in vitro and in *E. coli* with AmmA (previously annotated Amm6) and AmmA lacking the C-terminal Trp, but we observed no activity. We noted that AmmA has homology with other peptides encoded in clusters with small LanB proteins (Fig. 4B) but that AmmA appears to have a C-terminal extension. When we removed this extension, AmmB2 (previously annotated Amm9), but not the other three AmmB proteins, added a Trp in a Trp-tRNA–dependent fashion to the C terminus of the peptide in vitro and in *E. coli* (Fig. 4C). This finding explains the observation that mutation or deletion of the C-terminal Trp still resulted in ammosamide production and supports catalytic use of the peptide. Such use provides an attractive explanation for the 134 mg/L of ammosamide C generated by the producing bacterium (*26*), because stoichiometric use would require production of 3.0 g of AmmA. Given this second example of tRNA-dependent activity, we suggest the name peptide-amino acyl tRNA ligase

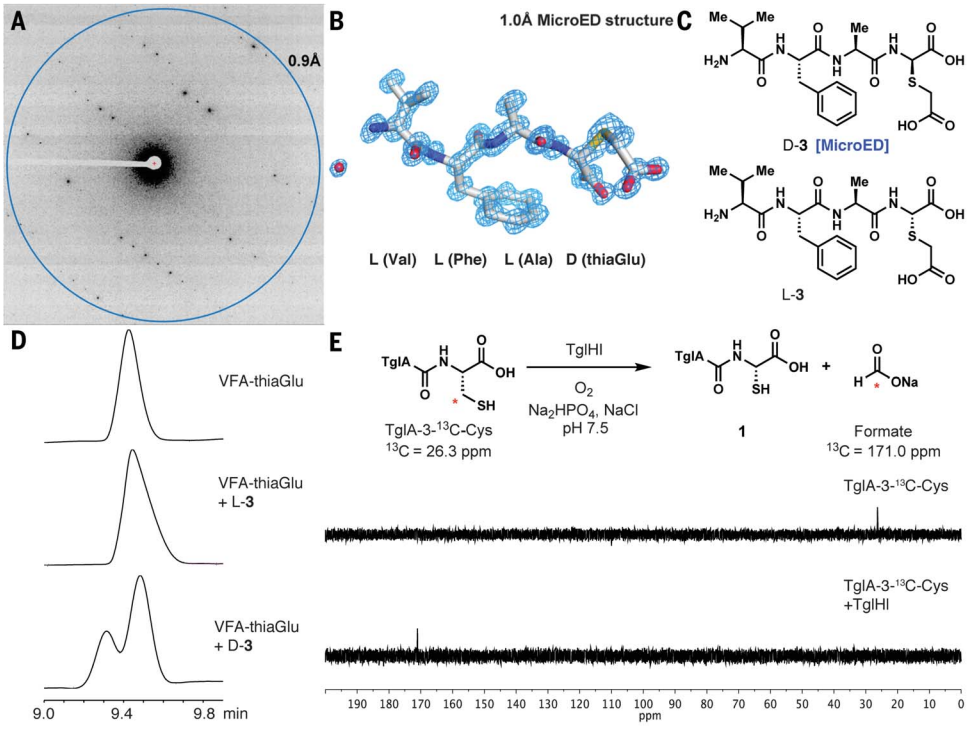



**Fig. 3. In vitro TglHI reacts with $^{13}$C-labeled TglA-Cys to produce $^{13}$C-formate and compound 1 with retention of configuration.** (**A**) Diffraction pattern of D-**3** with resolution ring at 0.9 Å. (**B**) Atomic MicroED structure of D-**3** determined at 1.0-Å resolution. (**C**) Structure of chemically synthesized tetrapeptides (VFA-thiaGlu) containing D-thiaGlu (D-**3**) and L-thiaGlu (L-**3**). (**D**) Determination of stereochemical configuration of thiaGlu by comparison with synthetic standards. High-performance liquid chromatograms are shown. VFA-thiaGlu was obtained by TglHI modification of TglA-Cys and then 2-iodoacetic acid alkylation and trypsin digest. (**E**) $^{13}$C NMR spectra showing the β carbon of the C-terminal cysteine in $^{13}$C-labeled TglA-Cys (26.3 ppm; top), and a new signal at 171.0 ppm that corresponds to $^{13}$C-formate after reaction with TglHI (bottom).

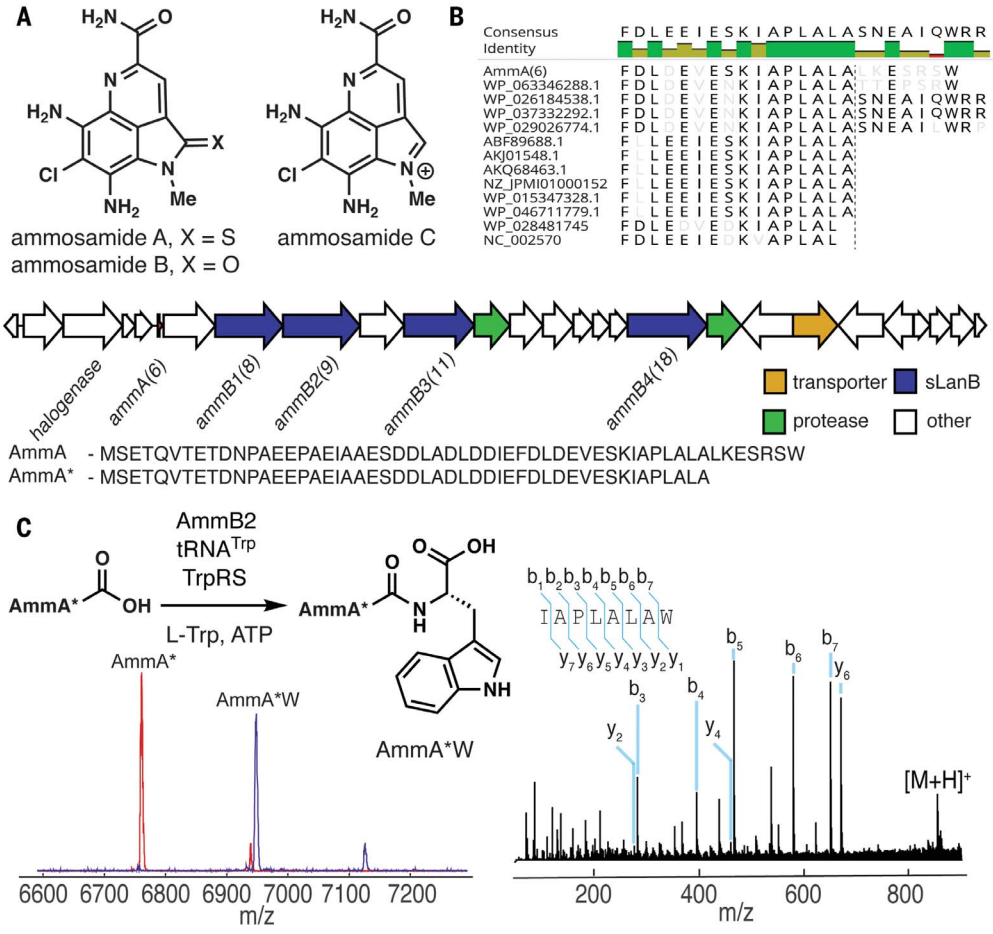



**Fig. 4. Ammosamide biosynthesis involves addition of L-Trp to the C terminus of a ribosomally synthesized peptide.** (**A**) Pyrolloquinoline alkaloid ammosamides A to C. (**B**) Sequence alignment of the C terminus of the AmmA precursor peptide and its homologs, showing a C-terminal extension for AmmA relative to most homologs. The gene cluster for ammosamide biosynthesis in *Streptomyces* sp. CNR698 comprises 27 open reading frames. The encoded proteins include four small LanBs, two proteases, one halogenase, and a transporter. (**C**) AmmB2 adds L-Trp to AmmA*, a truncated peptide of AmmA, to afford AmmA*W in vitro in an ATP–, tRNA$^{Trp}$–, and Trp-RS–dependent reaction. The red MALDI-TOF mass spectrum is AmmA*, and the blue spectrum shows the product of the reaction. High-resolution ESI-MS/MS confirmed addition of L-Trp to the C terminus.

(PEARL) for the small LanB proteins. The biosynthesis of a metabolite on a small peptide scaffold is uncommon, with the closest similarity found in the biosynthesis of amino acids linked by isopeptide bonds to a glutamate residue on amino-carrier proteins in some bacteria (*27*, *28*).

## REFERENCES AND NOTES


1. N. Adnani, S. R. Rajski, T. S. Bugni, *Nat. Prod. Rep.* **34**, 784–814 (2017).
2. P. Cimermancic *et al.*, *Cell* **158**, 412–421 (2014).
3. M. A. Ortega, W. A. van der Donk, *Cell Chem. Biol.* **23**, 31–44 (2016).
4. P. G. Arnison *et al.*, *Nat. Prod. Rep.* **30**, 108–160 (2013).
5. M. A. Ortega *et al.*, *Nature* **517**, 509–512 (2015).
6. G. A. Hudson, Z. Zhang, J. I. Tietz, D. A. Mitchell, W. A. van der Donk, *J. Am. Chem. Soc.* **137**, 16012–16015 (2015).
7. Joint Center for Structural Genomics, 3BWW, Crystal structure of a duf692 family protein (hs_1138) from Haemophilus somnus 129pt at 2.20 Å resolution. Protein Data Bank (2008).
8. B. J. Burkhart, G. A. Hudson, K. L. Dunbar, D. A. Mitchell, *Nat. Chem. Biol.* **11**, 564–570 (2015).
9. J. Kim *et al.*, *Nature* **498**, 123–126 (2013).
10. J. Kim *et al.*, *Nucleic Acids Res.* **43**, 4602–4613 (2015).
11. C. G. Jones *et al.*, *ACS Cent. Sci.* **4**, 1587–1592 (2018).
12. B. L. Nannenga, D. Shi, A. G. W. Leslie, T. Gonen, *Nat. Methods* **11**, 927–930 (2014).
13. D. Shi, B. L. Nannenga, M. G. Iadanza, T. Gonen, *eLife* **2**, e01345 (2013).
14. J. Hattne *et al.*, *Structure* **26**, 759–766.e4 (2018).
15. J. Hattne, M. Martynowycz, T. Gonen, *bioRxiv* (2019).
16. J. A. McIntosh, M. S. Donia, E. W. Schmidt, *Nat. Prod. Rep.* **26**, 537–559 (2009).
17. G. E. Kenney, A. C. Rosenzweig, *ACS Chem. Biol.* **7**, 260–268 (2012).
18. H. J. Kim *et al.*, *Science* **305**, 1612–1615 (2004).
19. G. E. Kenney *et al.*, *Science* **359**, 1411–1416 (2018).
20. L. Feng *et al.*, *Science* **318**, 1608–1612 (2007).
21. M. Toyota *et al.*, *Science* **361**, 1112–1115 (2018).
22. E. Arrebola, F. M. Cazorla, A. Perez-García, A. de Vicente, *Toxins* **3**, 1089–1110 (2011).
23. M. N. Price, E. J. Alm, A. P. Arkin, *Nucleic Acids Res.* **33**, 3224–3234 (2005).
24. C. Chatterjee, M. Paul, L. Xie, W. A. van der Donk, *Chem. Rev.* **105**, 633–684 (2005).
25. A. Dufour, T. Hindré, D. Haras, J. P. Le Pennec, *FEMS Microbiol. Rev.* **31**, 134–167 (2007).
26. P. A. Jordan, B. S. Moore, *Cell Chem. Biol.* **23**, 1504–1514 (2016).
27. F. Hasebe *et al.*, *Nat. Chem. Biol.* **12**, 967–972 (2016).
28. T. Ouchi *et al.*, *Nat. Chem. Biol.* **9**, 277–283 (2013).


## ACKNOWLEDGMENTS

*P. syringae* pv. maculicola ES4326 was provided by D. Desveaux (University of Toronto). We thank R. Splain (van der Donk lab) and M. W. Martynowycz and J. Hattne (Gonen lab, UCLA) for advice and useful discussions. **Funding:** This work was supported by the National Institutes of Health (R37 GM058822 to W.A.v.d.D.; F32 GM129944 to C.P.T.; F32 GM120868 to M.A.F.). T.G. and W.A.v.d.D. are investigators of the Howard Hughes Medical Institute. **Author contributions:** C.P.T., M.A.F., and Z.Z. performed biochemical assays. M.A.F. performed bioinformatics analysis. C.P.T., M.A.F., and W.A.v.d.D. designed experiments, analyzed data, and wrote the manuscript. S.L.H and T.G. designed the MicroED experiment; performed MicroED data collection, processing, and refinement of the structure; and contributed to writing of the manuscript and figure preparation. **Competing interests:** The authors declare no competing financial interests. M.A.F. is employed by the American Association for the Advancement of Science, and his editorial access to the paper was blocked. **Data and materials availability:** All data supporting the findings in this study are provided in the main text and supplementary materials.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/280/suppl/DC1
Materials and Methods
Scheme 1
Figs. S1 to S11
Tables S1 to S3
References (*29*–*49*)

29 June 2018; resubmitted 8 May 2019
Accepted 20 June 2019
10.1126/science.aau6232

MARINE MANAGEMENT

# Seasonal to multiannual marine ecosystem prediction with a global Earth system model

Jong-Yeon Park[1,2,3]*†, Charles A. Stock[2]†, John P. Dunne[2], Xiaosong Yang[2], Anthony Rosati[2]

**Climate variations have a profound impact on marine ecosystems and the communities that depend upon them. Anticipating ecosystem shifts using global Earth system models (ESMs) could enable communities to adapt to climate fluctuations and contribute to long-term ecosystem resilience. We show that newly developed ESM-based marine biogeochemical predictions can skillfully predict satellite-derived seasonal to multiannual chlorophyll fluctuations in many regions. Prediction skill arises primarily from successfully simulating the chlorophyll response to the El Niño–Southern Oscillation and capturing the winter reemergence of subsurface nutrient anomalies in the extratropics, which subsequently affect spring and summer chlorophyll concentrations. Further investigations suggest that interannual fish-catch variations in selected large marine ecosystems can be anticipated from predicted chlorophyll and sea surface temperature anomalies. This result, together with high predictability for other marine-resource–relevant biogeochemical properties (e.g., oxygen, primary production), suggests a role for ESM-based marine biogeochemical predictions in dynamic marine resource management efforts.**

The incorporation of biogeochemical processes into global climate models has transformed them into Earth system models (ESMs) that aspire to holistically represent the interacting physical, chemical, and biological processes shaping global carbon and nutrient cycles (*1*). Unlike physical climate models, ESMs can explore feedbacks between global change and carbon fluxes within and between terrestrial, ocean, and atmospheric reservoirs (*2*–*4*). For oceans, ESMs have further provided outlooks for marine-resource–relevant changes beyond warming, including ocean acidification, deoxygenation, and changing ocean productivity (*5*–*7*).

Although knowledge of potential multidecadal marine resource shifts associated with climate change is strategically vital, these trends underlie often irregular seasonal to interannual climate and marine resource variations. Failure to anticipate such fluctuations has been a major contributor to past fisheries collapses (*8*). The desire to sustain marine resources subject to climate-driven fluctuations and change has prompted efforts toward more dynamic, environmentally informed marine resource decisions (*9*), including integration of seasonal to multiannual physical climate forecasts into management frameworks (*10*–*12*). Whereas the reliability of physical predictors alone for anticipating marine ecosystem responses is often limited (*13*), recent observations and idealized modeling studies suggest that biogeochemical drivers (e.g., acidity, oxygen, primary production) may be more predictable than their physical counterparts (*14*).

The development of seasonal to interannual marine biogeochemical predictions has been impeded by diverse challenges. These include difficulties associated with the integration of biogeochemical models with ocean data-assimilation systems used for forecast initialization (*15*–*17*), uncertainty in both physical and biogeochemical model structure (*18*), limited availability of and difficulties associated with assimilating biogeochemical data (*19*, *20*), and the large computational cost of retrospective forecast experiments required to rigorously assess biogeochemical prediction skill. As a result, studies of global biogeochemical prediction have relied upon limited reforecast experiments and idealized configurations distinct from those used for operational seasonal to multiannual physical climate predictions (*14*, *21*).

In this study, we present results from 2-year global biogeochemical forecasts initialized on the first of each month between 1991 and 2017. Each prediction has 12 ensemble members, creating a database of nearly 4000 forecasts and 8000 simulation years. The prediction system was constructed by integrating the Carbon, Ocean Biogeochemistry and Lower Trophics (COBALT) marine biogeochemical model (*22*) with seasonal to multiannual climate predictions from the Geophysical Fluid Dynamics Laboratory's (GFDL) CM2.1 climate model (*23*). CM2.1 has been shown to skillfully recreate primary modes of natural climate variability (*24*) and has been applied extensively to study seasonal and multiannual climate prediction (*25*, *26*). The physical initialization for CM2.1 forecasts was based on GFDL's ensemble-coupled data assimilation (ECDA) system (*27*). For the biogeochemical initialization, COBALT was integrated with the data-assimilative ocean physics following a strategy that carefully avoids spurious vertical motions that can degrade biogeochemical simulations (*15*). This integration resulted in substantial reductions of biogeochemical biases relative to nonassimilative simulations (*15*, *22*). The resulting ocean-state estimate captures large-scale sea surface temperature (SST) and chlorophyll variations (figs. S1 to S4).

Predictions were assessed against 20 years of satellite-derived chlorophyll estimates (*28*), which offer a near-global, continuous, multidecade time series of ocean ecosystem anomalies. Chlorophyll has also been found to be a robust indicator of cross-ecosystem (*29*) and, in some places, interannual fish-catch variations (*30*). Ensemble mean predictions are drift corrected with a lead-dependent monthly forecast climatology from the 27-year ensemble mean forecasts (see the materials and methods in the supplementary materials for further details).

The global marine biogeochemical prediction system produces skillful chlorophyll predictions one season in advance in many ocean regions (Fig. 1A). Significant chlorophyll prediction skill above that of a persistence forecast extends beyond 1 year in some regions (Fig. 1, B to F, and fig. S5). Although skill varies by region and initialization month, characteristic patterns emerged for tropical and extratropical regions. In the tropical Pacific (Fig. 1B), prediction skill is limited to maximum leads of 12 months, with peak skill for fall/winter forecasts and reduced skill for boreal spring predictions. This pattern closely resembles the El Niño–Southern Oscillation (ENSO) SST prediction skill (fig. S6) and is consistent with prediction of ENSO-associated nutricline variations, which peak in boreal winter but have a weak boreal spring signal (*31*, *32*). Strong negative and positive winter chlorophyll anomalies are tightly linked to El Niño and La Niña events, respectively (Fig. 2A). This relationship also holds for the tropical Indian Ocean (Fig. 2B), which is subject to a lagged ENSO signal carried into the Indian Ocean through atmospheric teleconnections (*33*).

Chlorophyll predictions in extratropical systems are characterized by alternating predictable and unpredictable forecast windows (Fig. 1, D to F; note diagonal bands of alternating high- and low-anomaly correlation coefficients). In the subtropical to temperate North Atlantic (Fig. 1D), chlorophyll anomalies are not predictable in winter but are predictable during the productive spring, summer, and fall (i.e., the growing season). Furthermore, prediction skill remains evident through two growing seasons with leads up to 24 months. This skill results from successfully simulating the persistence of initial subsurface nutrient anomalies across seasons and successfully simulating the subsequent impact of these anomalies on surface chlorophyll. Winter nitrate ($NO_3$) anomalies linked to North Atlantic Oscillation (NAO)–driven wind anomalies (fig. S7)

[1]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ 08540, USA. [2]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540, USA. [3]Department of Earth and Environmental Sciences, Chonbuk National University, Jeonju-si, Jeollabuk-do 54896, Republic of Korea.
*Corresponding author. Email: jongyeon.park@jbnu.ac.kr
†These authors contributed equally to this work.

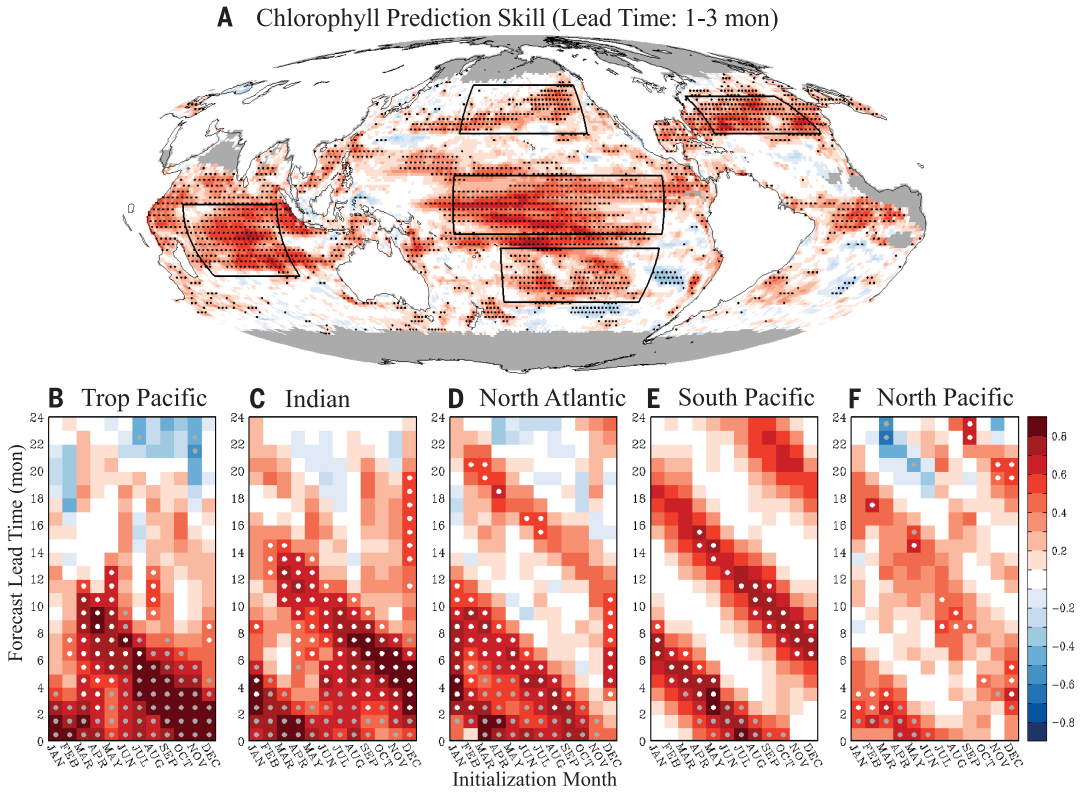



**Fig. 1. Prediction skill in reproducing observed variations of monthly chlorophyll anomaly.** (**A**) Chlorophyll prediction skill measured by the mean monthly anomaly correlation coefficient (ACC) between predicted and observed (satellite) chlorophyll at a 1- to 3-month lead time during the period 1997–2017. Stippled areas indicate that the correlation is significantly greater than 0 with 95% confidence. Areas with less than 80% satellite chlorophyll coverage are masked in gray. (**B** to **F**) Chlorophyll prediction skill as a function of forecast initialization month (*x* axis) and lead time (*y* axis) in the Tropical Pacific (170°E–100°W, 10°S–10°N), Indian (55°E–95°E, 25°S–0°S), North Atlantic (70°W–20°W, 25°N–40°N), North Pacific (170°E–130°W, 25°N–45°N), and South Pacific (170°W–100°W, 35°S–15°S) oceans. Circles indicate significant ($P < 0.05$) prediction skill: White circles indicate that the chlorophyll forecast skill from the biogeochemical prediction system exceeds the persistent forecast skill; gray circles indicate that the skill of the biogeochemical forecast is significant, but it is not significantly better than a persistence forecast. Three-month running mean anomalies are used for the calculation of ACCs.

remain evident beneath the mixed layer during summer and reemerge when the mixed layer deepens during the subsequent fall and winter (Fig. 3A). High-$NO_3$ anomalies then lead to elevated chlorophyll during the following growing season (Fig. 3B). A composite of high- versus low-chlorophyll years (fig. S8) confirms that high-$NO_3$ anomalies are associated with enhanced $NO_3$-based spring phytoplankton production, followed by enhanced ammonium-based (i.e., recycled) production during the summer and fall. This chlorophyll reemergence pattern resembles the mechanism for predicting midlatitude SST anomalies (*34*) but, unlike the predictable SST signal, which occurs after the breakdown of summer stratification, the predictable chlorophyll signal occurs during the stratified period.

Other extratropical areas exhibit variations on the basic reemergence mechanism illustrated for the North Atlantic. The South Pacific reemergence signal remains exceptionally strong through 2 years (Fig. 1E; Fig. 3, C and D; and fig. S9). In the North Pacific, prediction skill is weaker and limited in spatial extent (Fig. 1F; Fig. 3, E and F; and fig. S10). This may reflect a greater role of atmospheric iron deposition in the North Pacific masked by our current use of a constant deposition climatology (*22*) or stronger stochastic atmospheric forcing causing irregular and spatially less homogeneous chlorophyll fluctuations relative to other regions (*35*).

Successful prediction of chlorophyll anomalies in some regions across multiple years gives cause for optimism concerning the utility of biogeochemical predictions for marine resource application. As a proof of concept, we considered the capacity to anticipate interannual fluctuations in aggregate fish catch using predicted chlorophyll and SST, two known "bottom-up" drivers of fish catch (*12*, *29*, *30*). We assessed predictions in coastal large marine ecosystems (LMEs, fig. S11) accounting for over 95% of global fish catch (*36*). Annual mean fish-catch data were obtained from the Sea Around Us project (*37*). Despite coarse ocean grids that limit resolution of coastal circulation and ecosystem processes, global climate prediction systems have significant SST-forecasting skill for many LMEs (*38*, *39*), and our results show that this also holds for interannual chlorophyll anomalies (fig. S12).

We assessed the potential for biogeochemical predictions to inform interannual fish-catch fluctuations in a subset of LMEs based on three conditions. First, we identified LMEs in which past interannual catch fluctuations are significantly correlated with observed SST or chlorophyll anomalies over the retrospective forecast period. We considered both concurrent and 1-year–lagged relationships. The concurrent relationship tests for rapid catch responses such as immigration during favorable conditions. A 1-year lag allows for propagating environmental effects such as recruitment of short-lived species. Whereas longer lag responses between environmental drivers and ecosystem responses are possible (*40*, *41*), contemporaneous or short lag signals have proven to be the most tractable for management-driven forecasts (*42*) and are of primary interest for assessing the utility of the interannual biogeochemical predictions herein. Twenty-five out of the 54 heavily fished LMEs considered satisfied this first condition (Fig. 4 and figs. S13 and S14). The absence of a significant relationship in 29 LMEs does not imply that there are no "bottom-up" constraints on these systems, only that a relationship between interannual aggregate catch and SST or chlorophyll anomalies could not be discerned over the retrospective forecast period.

Second, we subselected LMEs for which the global biogeochemical prediction system could predict observed annual mean SST or chlorophyll anomalies with significant skill. Such cases were ubiquitous, with 38 of 54 LMEs satisfying this condition despite the model's coarse ocean resolution (fig. S12). Fifteen of these LMEs also satisfied our first condition (Fig. 4).

Third, we subselected LMEs in which the bottom-up relationship was strong enough and environmental predictions were skillful enough to significantly explain the reported aggregate interannual fish-catch anomalies. Consistent with our first selection condition, predicted

fish catch is based on a simple linear regression of catch anomalies against predicted environmental (SST, chlorophyll) anomalies with contemporaneous or 1-year–lagged relationships (see materials and methods). We focused on LMEs for which significant fish-catch relationships remained after detrending to reduce the potential of an erroneous attribution of fishing effort trends onto environmental factors. Six systems satisfied all three conditions (Fig. 4, table S1, and figs. S15 and S16): Gulf of Alaska, California Current, Humboldt Current, Canadian Eastern Arctic, Agulhas Current, and Somali Coastal Current systems (Fig. 4, lower panels). Four out of six LMEs exhibited significant fish-catch prediction skill for an additional year (i.e., at a 1- to 2-year lead time; Fig. 4, green lines).

In the Gulf of Alaska LME, catch fluctuations covaried with predictable coastal SST variations associated with the Pacific Decadal Oscillation, a basin-scale mode of climate variability with diverse fisheries links (*38*). In the California Current LME, the model predicts a recent observed chlorophyll increase off the Baja Peninsula that covaried with increased reported catch from this region (Fig. 4C). Weak ENSO imprints are also apparent, particularly a catch reduction after the 1997–1998 El Niño. This signal is more prominent in the Humboldt Current (Fig. 4D). In the Canadian Eastern Arctic LME, a predictable warming in the 1990s covaried with increasing northern prawn catches, both of which leveled off as temperatures stabilized in the 2000s (Fig. 4E). This LME is toward the cold-water end of the northern prawn range (*43*), suggesting a favorable response to warming. More complex interactions with changing plankton dynamics, however, cannot be ruled out (*43*).

In the Indian Ocean, increasing Agulhas LME catch between 1996 and 2004 corresponds to the increased prominence of the sardine fishery

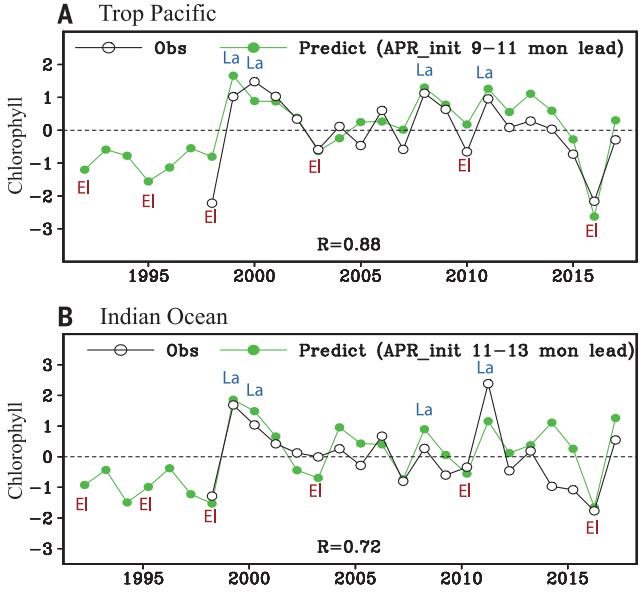



**Fig. 2. Observed and predicted chlorophyll anomalies in the tropical oceans.** (**A**) Observed (black) and predicted (green) wintertime (December-January-February) normalized chlorophyll anomalies in the tropical Pacific (170°E–100°W, 10°S–10°N). The predicted anomalies are 1 April–initialized chlorophyll predictions (i.e., forecast lead time is 9 to 11 months). (**B**) Similar to (A) but for 1 April–initialized springtime (February-March-April) chlorophyll anomalies in the Indian Ocean (55°E–95°E, 25°S–0°S; i.e., forecast lead time is 11 to 13 months). El Niño and La Niña years are marked with "El" and "La," respectively.

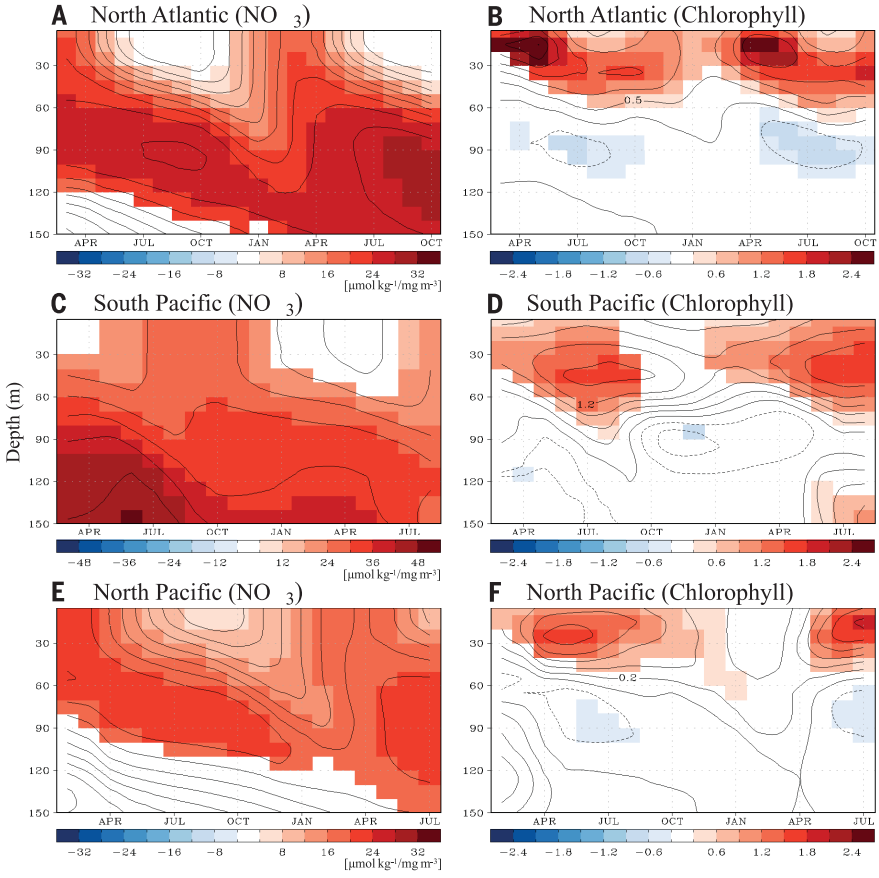



**Fig. 3. Reemergence of subsurface biogeochemical anomalies linked to chlorophyll prediction skill in the extratropical oceans.** (**A**) Temporal evolution of 1 March–initialized $NO_3$ anomaly prediction in the North Atlantic Ocean (70°W–20°W, 25°N–40°N). $NO_3$ anomalies are regressed onto predicted SON (September-October-November) surface chlorophyll concentrations during the following year. That is, positive values indicate that elevated $NO_3$ at the specified time and depth are associated with elevated SON chlorophyll 18 to 21 months after the initialization. The anomalies are 3-month running means. (**B**) Similar to (A) but for 1 March–initialized chlorophyll anomaly prediction. (**C**) and (**D**) are similar to (A) and (B), respectively, but for 1 March–initialized prediction in the south Pacific (170°W–100°W, 35°S–15°S) regressed onto 1 March–initialized July-August-September chlorophyll prediction of the following year. (**E**) and (**F**) are similar to (**A**) and (**B**), respectively, but for 1 February–initialized prediction in the North Pacific (170°E–130°W, 25°N–45°N) regressed onto 1 February–initialized June-July-August chlorophyll prediction of the following year. Shaded areas represent the 95% confidence region.

(*44*). Although debate over the cause of this increase remains, its consistency with a lagged chlorophyll relationship is suggestive of a recruitment link. Catch variations in the neighboring Somali Coastal Current are similar to those in the Agulhas, but the underlying characteristics of chlorophyll predictability are different. Agulhas anomalies follow a midlatitude reemergence pattern similar to that of the extratropical areas shown in Fig. 1, whereas the Somali system exhibits relatively limited chlorophyll prediction skill because of equatorial waves repeatedly triggered by ENSO and the Indian Ocean dipole (fig. S17) (*45*). In both the Agulhas and Somali systems, annual fish catch was predictable up to 2 to 3 years in advance using the 1-year–lagged relationship between catch and chlorophyll.

Although only six LMEs met the most stringent criteria for skillful prediction of interannual catch anomalies from SST or chlorophyll, many notable relationships for individual climate-sensitive fish stocks may underlie aggregate catch relationships (fig. S18). More detailed accounting for fish-stock dynamics (*42*) and variations in fishing effort could also better isolate predictable bottom-up signals (*30*). The prediction skill threshold at which forecasts become useful is fishery dependent, but recent management strategy evaluations suggest an elevated likelihood of utility for species with short prerecruit survival windows or strong environmental bottlenecks (*12*, *42*). In addition, biogeochemical prediction systems can extend beyond SST and chlorophyll to include other potential drivers, including oxygen, acidity, net primary production (NPP), and zooplankton. Assessment of NPP predictions

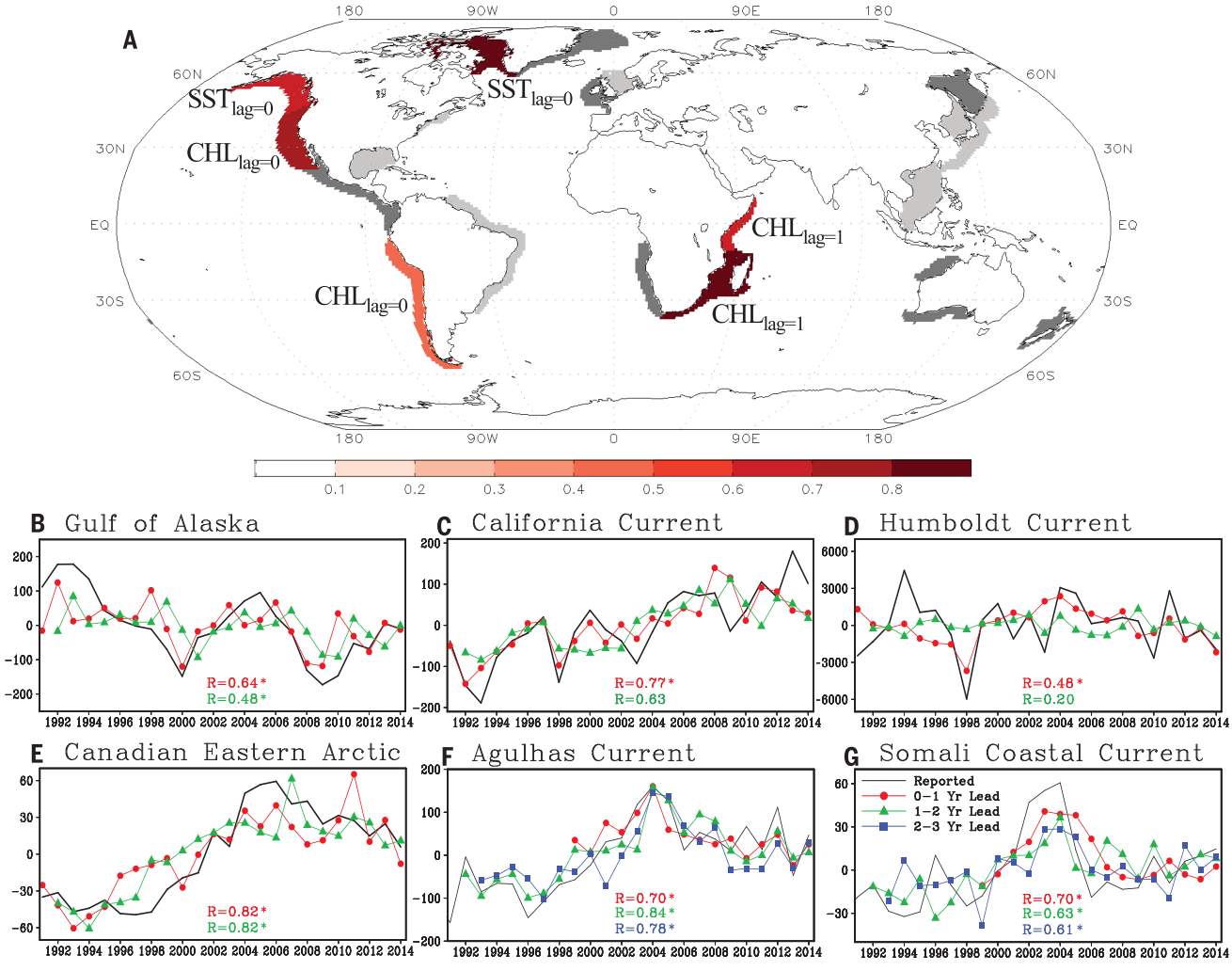



**Fig. 4. Potential utility of marine biogeochemistry prediction for annual fish-catch prediction.** (**A**) All shaded LMEs represent regions where past annual fish-catch fluctuations are significantly correlated with observed bottom-up factors [i.e., SST or chlorophyll (CHL)]. Dark gray represents regions where the ESM-based prediction system can predict bottom-up forcing changes, and color shadings represent regions where the ESM-based prediction system can predict both bottom-up forcing changes and reported fish catch. Bottom-up factors and the time lag used for fish-catch predictions are shown near each predictable LME. (**B** to **G**) Reported (black lines) and predicted (colored lines) annual mean fish catches in (B) Gulf of Alaska, (C) California Current, (D) Humboldt Current, (E) Canadian Eastern Arctic, (F) Agulhas Current, and (G) Somali Coastal Current LMEs (unit: $10^3$ tonnage). Predicted annual fish catches are based on the 1 January–initialized SST (for the Gulf of Alaska and Canadian Eastern Arctic systems) or chlorophyll (for other systems) predictions in the coming year (i.e., forecast lead time is 0 to 1 year; red lines) and the following year (i.e., forecast lead time is 1 to 2 years; green lines). In each LME, the time lag at which the maximum fish-catch prediction skill occurs is used for the fish-catch prediction. For example, the fish-catch predictor for Agulhas Current is the annual mean chlorophyll in the previous year, thus the observed annual mean chlorophyll in the previous year is used for the 0- to 1-year lead-time forecast. Similarly, 1 January–initialized chlorophyll predictions for the coming and following years are used for the 1- to 2-year and 2- to 3-year lead-time forecasts in the Agulhas Current system. Asterisks indicate the significant ($P < 0.05$) correlation between reported and predicted annual fish catches.

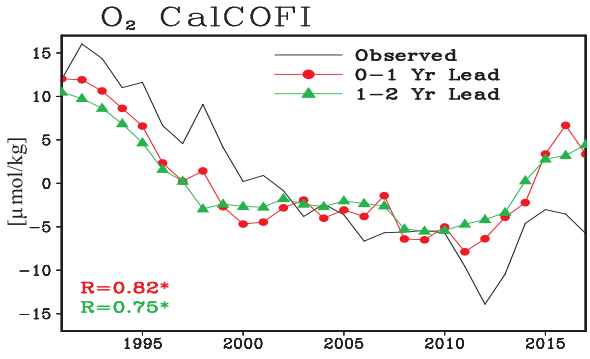



**Fig. 5. Oxygen prediction in the California Current LME as an example of extension to other biogeochemical stressor properties.** Observed oxygen from the California Cooperative Fisheries Investigations (CalCOFI) program (black lines) and 1 January–initialized oxygen prediction for the coming year (0- to 1-year lead time; red line) and for the following year (1- to 2-year lead time; green line). Dissolved oxygen shown here is the averaged value between the 200- and 500-m depth in the CalCOFI region. Asterisks indicate a significant ($P < 0.05$) correlation between observed and predicted oxygen.

against satellite-based NPP algorithms suggests patterns of predictability similar to those of chlorophyll (fig. S19), whereas model-based assessments of the potential predictability of other drivers suggest that they may be more predictable than chlorophyll or SST (fig. S20). For example, subsurface oxygen predictions that accompany the skillful California Current LME chlorophyll predictions highlighted in Fig. 4 were robust throughout our 2-year prediction horizon (Fig. 5). Such lasting and predictable "biogeochemical memory" gives cause for further optimism concerning the benefit of extending physical climate predictions to marine biogeochemical predictions for marine resource management in a dynamic environment.

## REFERENCES AND NOTES


1. G. B. Bonan, S. C. Doney, *Science* **359**, eaam8328 (2018).
2. P. Friedlingstein *et al.*, *J. Clim.* **27**, 511–526 (2014).
3. A. Anav *et al.*, *J. Clim.* **26**, 6801–6843 (2013).
4. J. L. Sarmiento, T. M. C. Hughes, R. J. Stouffer, S. Manabe, *Nature* **393**, 245–249 (1998).
5. W. W. L. Cheung *et al.*, *ICES J. Mar. Sci.* **73**, 1283–1296 (2016).
6. C. A. Stock *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1441–E1449 (2017).
7. L. Bopp *et al.*, *Biogeosciences* **10**, 6225–6245 (2013).
8. F. P. Chavez, J. Ryan, S. E. Lluch-Cota, C. M. Niquen, *Science* **299**, 217–221 (2003).
9. K. N. Marshall, L. E. Koehn, P. S. Levin, T. E. Essington, O. P. Jensen, *ICES J. Mar. Sci.* **76**, 1–9 (2019).
10. A. J. Hobday, C. M. Spillman, J. P. Eveson, J. R. Hartog, *Fish. Oceanogr.* **25**, 45–56 (2016).
11. D. Tommasi *et al.*, *Prog. Oceanogr.* **152**, 15–49 (2017).
12. D. Tommasi *et al.*, *Ecol. Appl.* **27**, 378–388 (2017).
13. R. A. Myers, *Rev. Fish Biol. Fish.* **8**, 285–305 (1998).
14. R. Séférian *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11646–11651 (2014).
15. J. Y. Park *et al.*, *J. Adv. Model. Earth Syst.* **10**, 891–906 (2018).
16. K. Raghukumar *et al.*, *Prog. Oceanogr.* **138**, 546–558 (2015).
17. J. Waters, M. J. Bell, M. J. Martin, D. J. Lea, *QJRMS* **143**, 195–208 (2017).
18. E. Hawkins, R. S. Smith, J. M. Gregory, D. A. Stainforth, *Clim. Dyn.* **46**, 3807–3819 (2016).
19. D. A. Ford *et al.*, *Ocean Sci.* **8**, 751–771 (2012).
20. H. Song, C. A. Edwards, A. M. Moore, J. Fiechter, *Ocean Model.* **106**, 131–145 (2016).
21. C. S. Rousseaux, W. W. Gregg, *Front. Mar. Sci.* **4**, 236 (2017).
22. C. A. Stock, J. P. Dunne, J. G. John, *Prog. Oceanogr.* **120**, 1–28 (2014).
23. T. L. Delworth *et al.*, *J. Clim.* **19**, 643–674 (2006).
24. A. T. Wittenberg, A. Rosati, N. C. Lau, J. J. Ploshay, *J. Clim.* **19**, 698–722 (2006).
25. B. P. Kirtman *et al.*, *Bull. Am. Meteorol. Soc.* **95**, 585–601 (2014).
26. G. A. Meehl *et al.*, *Bull. Am. Meteorol. Soc.* **95**, 243–267 (2014).
27. S. Zhang, M. J. Harrison, A. Rosati, A. Wittenberg, *Mon. Weather Rev.* **135**, 3541–3564 (2007).
28. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, Moderate-resolution Imaging Spectroradiometer (MODIS) Terra Aqua Chlorophyll Data (NASA OB.DAAC, 2014).
29. K. D. Friedland *et al.*, *PLOS ONE* **7**, e28945 (2012).
30. C. J. McOwen, W. W. L. Cheung, R. R. Rykaczewski, R. A. Watson, L. J. Wood, *Fish Fish.* **16**, 623–632 (2015).
31. M. J. Behrenfeld *et al.*, *Science* **291**, 2594–2597 (2001).
32. F. P. Chavez *et al.*, *Science* **286**, 2126–2131 (1999).
33. S. Venzke, M. Latif, A. Villwock, *J. Clim.* **13**, 1371–1383 (2000).
34. M. A. Alexander, C. Deser, M. S. Timlin, *J. Clim.* **12**, 2419–2433 (1999).
35. A. Morel, H. Claustre, B. Gentili, *Biogeosciences* **7**, 3139–3151 (2010).
36. K. Sherman, in *Sustaining Large Marine Ecosystems: The Human Dimension*, T. M. Hennessey, J. G. Sutinen, Eds. (2005), vol. 13, pp. 3–16.
37. D. Pauly, D. Zeller, *Nat. Commun.* **7**, 10244 (2016).
38. C. A. Stock *et al.*, *Prog. Oceanogr.* **137**, 219–236 (2015).
39. D. Tommasi *et al.*, *Front. Mar. Sci.* **4**, 201 (2017).
40. E. Di Lorenzo, M. D. Ohman, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2496–2499 (2013).
41. P. Le Mezo *et al.*, *J. Mar. Syst.* **153**, 55–66 (2016).
42. M. Haltuch *et al.*, *Fish. Res.* **217**, 198–216 (2019).
43. P. Koeller *et al.*, *Science* **324**, 791–793 (2009).
44. L. Hutchings *et al.*, *Prog. Oceanogr.* **83**, 15–32 (2009).
45. J. Currie *et al.*, *Biogeosciences* **10**, 6677–6698 (2013).


## ACKNOWLEDGMENTS

We thank the anonymous reviewers for helpful and constructive comments on the manuscript. We also thank T. Delworth and F. Gonzalez-Taboada for internal reviews of the manuscript and J. Link, R. Watson, and V. Lam for discussions and input during the revision process. **Funding:** This work was supported by NOAA's marine ecosystem tipping points initiative. J.-Y.P. was also supported by the National Research Foundation of Korea (grant NRF-2018R1C1B5086584). **Author contributions**: J.-Y.P., C.A.S., J.P.D, X.Y., and A.R. conceived the experiments. J.-Y.P. and C.A.S. wrote the manuscript. J.-Y.P. performed experiments with X.Y. and analyzed the data. All authors discussed the results and commented on the manuscript at all stages. **Competing interests**: The authors declare no competing interests. **Data and materials availability**: The model simulation data used here are archived at NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) server. Complete datasets may be requested from J.-Y.P. Key data used in the main manuscript and supplementary materials are archived at GFDL's public server (ftp://nomads.gfdl.noaa.gov/users/Jong-Yeon.Park/ESMprediction/).

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6450/284/suppl/DC1
Materials and Methods
Table S1
Figs. S1 to S20
References (*46*–*54*)

8 October 2018; accepted 25 June 2019
10.1126/science.aav6634


## Present your work to the world.

Are you a representative of the upcoming generation of thought leaders in your field? Together we look forward to your application for the new Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy.

## Apply now!

The Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy

www.sartorius.com/sartorius-and-science-prize

Awarded by sartorius Science


### Nanoparticle Tracking Analyzer

Analytik has launched ZetaView QUATT, with a multiwavelength video-tracking microscope for rapid analysis of size, concentration, fluorescence, electrophoretic mobility, and subpopulations of individual nanoparticles. Using proprietary Nanoparticle Tracking Analysis (NTA) technology, ZetaView instruments can capture the Brownian motion of each particle in a video to determine its hydrodynamic diameter. Furthermore, following the movement of the particles in an applied electric field, the charge state of the particle surface (zeta potential) can be measured. Consequently, pattern parameters, such as intensity fluctuations, surface geometry, and particle and concentration, are recorded and can be used to distinguish subpopulations. ZetaView QUATT houses four different wavelength lasers, so one, two, or four different fluorophores or dyes with different excitation and emission spectra can be analyzed on the same sample, allowing calculation of biomarker ratios. The innovation of multiwavelength fluorescence-NTA also reduces total measurement times, minimizes the amount of sample required, and improves experimental reproducibility.

**Analytik**
For info: +44-(0)-1954-232776
www.analytik.co.uk

### High-Performance *Drosophila* ChIP Kits

Chromatrap offers the first commercial ChIP kit range dedicated to *Drosophila* (fruit fly). Benefiting from proprietary bead-free chromatin immunoprecipitation (ChIP) technology, these new kits deliver high-quality, ChIP-ready chromatin from as little as five *Drosophila* specimens and from a range of *Drosophila* tissues. The kit range now offers a single optimized Chromatrap kit with unique buffers and components for *Drosophila* chromatin extraction and ChIP. Incorporating correctly oriented capture proteins throughout the inert, filter-based ChIP technology, the new kits offer outstanding sensitivity that is especially important for detection of binding of low-abundance targets. Coupling these advantages with the well-documented benefits of bead-free ChIP assay technology provides researchers with an easy, sensitive workflow that can be completed in less than 5 h. The new Chromatrap range includes *Drosophila* UniqSeq, ChIP-Seq, Antibodies, and Primer Sets.

**Chromatrap**
For info: +44-(0)-1978-661144
www.chromatrap.com/drosophila-chip-seq-range1

### T-Cell Functional Analytics Platform

Berkeley Lights announces the launch of its Lightning optofluidic platform, which directly visualizes phenotype and function of thousands of T cells in days, to help scientists address challenges in developing cancer immunotherapies. Over the last decade, scientists have advanced their understanding of immune-cell function in many contexts, including cancer. However, progress has been slow, because identification of immune-cell functional signatures requires time-consuming processes that average individual cell behaviors and interactions. Using Berkeley Lights' proprietary optofluidic technology, individual T cells can be loaded into NanoPen chambers with other cell types to assess cell–cell interactions, cell-surface phenotype, cytotoxicity, and cytokine secretion, yielding a more detailed picture of individual cell function within a population. The entire process is digitally recorded for visual confirmation of results and on-instrument data analysis. The platform is also designed to overcome one of the major hurdles to most existing cellular assays—destruction of the cell during analysis—by enabling scientists to recover live cells after completion of functional assays.

**Berkeley Lights**
For info: 510-858-2855
www.berkeleylights.com

### CRISPR-Based Mitochondrial DNA Depletion Kits

Arbor Biosciences announces the launch of myNGS Guides MitoDeplete Kits as an introduction to their growing catalog of guide RNA pools for targeted sequencing applications. This new depletion system combines Arbor's massively parallel, high-fidelity oligo synthesis platform with best-in-class Cas enzymes for efficiently targeting unwanted molecules from NGS libraries. In-stock MitoDeplete catalog kits have been specifically designed and validated for human and mouse applications, while kits for other custom species can be designed and synthesized on demand. The kits are designed for use on any NGS library containing a significant percentage of mitochondrial DNA (mtDNA) molecules, such as ATAC-Seq (assay for transposase-accessible chromatin sequencing) libraries, which typically contain an abundance of mtDNA-derived molecules. CRISPR-based depletion with MitoDeplete delivers up to a 95% reduction in mtDNA-derived molecules in a simple, single-step process, which translates into a 30%–70% savings in sequencing costs.

**Arbor Biosciences**
For info: 734-998-0751
arborbiosci.com

### Biomarker Validation and Assay Development

Immuno-Parallel Reaction Monitoring (iPRM) is a multiplexed LC/MS assay service that combines the specificity of mass spectrometry, the throughput of an immunoassay, and the power of multiplexing to accelerate the time-to-market of your drug and/or diagnostic test. Our iPRM services let you leverage the proteomic expertise of Cell Signaling Technology (CST) scientists and the established specificity and sensitivity of CST antibodies for streamlined assay development and reliable validation of your therapeutic target(s). iPRM services enable you to seamlessly transition from biomarker discovery to biomarker validation; profile the impact of your drug across your choice of targets to determine on- and off-target effects; and quantify dozens of analytes in a single, complex biological sample with a high degree of confidence, sensitivity, and specificity. Learn more about iPRM services or consult with a CST in-house expert.

**Cell Signaling Technology**
For info: 877-678-8324
www.cellsignal.com

### Mass Spectrometry Systems

Waters offers two new high-resolution mass spectrometry systems. The Waters SELECT SERIES Cyclic IMS seamlessly integrates cyclic ion mobility (cIM) technology into a high-performance, research-grade time-of-flight mass spectrometer providing limitless experimental potential. Combining novel ion mobility spectrometry (IMS) capability with significantly improved MS performance and enabling software, this platform allows researchers to unlock the potential in scientific discovery. The SYNAPT XS is a new, highly flexible, high-resolution mass spectrometer for R&D labs focused on discovery applications that need the greatest variety of analytical strategies to tackle challenging questions. By providing increased flexibility through inlets and acquisition modes, the SYNAPT XS delivers greater freedom of analytical choice to support scientific creativity and technical success for any application.

**Waters Corporation**
For info: 800-252-4752
www.waters.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

BIOLOGY ■ EARTH ■ ENVIRONMENT ■ LIFE ■ INTERDISCIPLINARY ■ PHYSICAL ■ MATERIAL ■ SOCIAL SCIENCES

# Science Advances | AAAS

**OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS**


**OPEN ACCESS**

## Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**


AAAS | ANNUAL MEETING
Seattle, WA | February 13–16, 2020

## Mark Your Calendar!

The Annual Meeting event lasts 4 days, but AAAS staff, the program committee, disciplinary sections, a local committee, and others work hard all year to put on the best event we can! Make sure you don't miss a deadline.

**aaas.org/meetings**


# Astronomy at Nanjing University

The School of Astronomy and Space Science of Nanjing University was established in March 2011, and its predecessor, the Department of Astronomy, was founded in 1952. It earns the longest history and high reputation among the astronomy departments in China. The School owns the national first-level key discipline of astronomy, consisting of two national second-level key disciplines: Astrophysics and Astrometry & Celestial Mechanics. It also consists of several laboratories, such as Key Laboratory of Modern Astronomy and Astrophysics of the Ministry of Education, Central Laboratory for Teaching, Solar Tower Laboratory, etc. During the previous National Primary Discipline Evaluations, astronomy discipline of Nanjing University always ranked 1st.

The school has established a strong faculty consisting of both distinguished senior scholars and a large number of young scholars. There are 50 faculty members, including 23 professors, with 4 Academicians of the Chinese Academy of Science, 4 Chang Jiang Scholars, 1 Chief Scientist for the National Basic Research Program (973 Program), 9 obtainers of National Outstanding Youth Funds and 9 youth high-level talents. The percentage of young scholars under the age of forty-five is around 71%, which guarantees an exuberant momentum of development.

Known as the "Whampoa Military Academy" in astronomical society, the school focused upon cultivation of excellent talents with international perspective and innovation ability. It has fostered more than 2000 graduated students, most of them have grown into leading experts, including 5 members of Chinese Academy of Sciences and a group of academic leaders in top-ranked Chinese and overseas universities and observatories.

To conform with the frontier of modern Astronomy and Astrophysics, the school's research targets at six influential fields, including (1) high-energy astrophysics, (2) physics of solar activity region, (3) galaxies and cosmology, (4) physics of planetary systems and exoplanet search, (5) theory of astrometry and astronomical reference frame, and (6) deep space exploration and astrodynamics. The high-energy astrophysics research group studies supernovae and its remnants, gamma-ray bursts and their afterglows, compact stellar objects, and particle acceleration, all of which are related to the most violent explosions in our Universe. In particular, the school's gamma-ray burst research is internationally well-recognized. The solar physics research group, aided with its unique spectrographic and imaging telescope, studies the three-dimensional structure, the eruption mechanisms and the magnetohydrodynamic processes in various solar activities. The planet research group explores planet formation and evolution, and recently discovered a sizable sample of exoplanet candidates with the telescope at the South Pole. The astrodynamics research group plays a leading role in the dynamical theory of man-made satellites. In recent years, the school has won a series of national and provincial awards, such as the second class prize of the National Award of Natural Sciences. The works on high-precision measurement of the distance to the Perseus Spiral Arm and the study of X-ray flares in gamma-ray bursts were published in *Science*. The former was thought to open the new era of VLBI measurement of the Galactic structure, and was selected to the Top 10 Science and Technology Development of China's Higher Education and Top 10 News of Chinese Fundamental Research. Recently, the school spared no effort to construct the Time Domain Observatory in Tibet which is helpful to TESS mission, and will give star catalogues around North Ecliptic Pole.

In the past four years, the school received a large number of grants (more than 85 million RMB) from both the NSFC and the industry. More than 120 papers were published per year and about 85 percent of them were in high quality journals, in particular one in *Nature* and 9 in *Nature* Series. In addition, the project of the Optical-NIR Solar Eruption Explore Telescope was awarded first class Ministerial Prize for Advances in Science and Technology as well as Top 10 Science and Technology Achievements in Chinese Astronomy.

In order to establish close collaboration between universities and observatories, in December 2012, Nanjing University, Peking University, the University of Science & Technology of China, National Astronomical Observatories, and Purple Mountain Observatory co-founded the "Collaborative Innovation Center of Modern Astronomy and Space Exploration". Additionally, the school has led to finish a number of the strategic reports that guide the future development of Chinese Astronomy. It also serves as an active member in the scientific researches with national astronomical facilities such as LAMOST, FAST and the South Pole Observatory. Furthermore, the school explores the channels of cooperation with overseas partners to push forward the process of internationalization and has established steady cooperative relationship with over 20 top-level institutes in USA, UK, Australia, Japan, etc. To be more specific, the NJU-IRAM Collaboration has been established in December 2017, and the school signed the MoU with CfA in January 2018. Fruitful academic accomplishments and awards have been obtained owe to those vigorous exchanges with both domestic and outside world.

**Email:** qulian@nju.edu.cn
**Website:** http://enastronomy.nju.edu.cn/
http://astronomy.nju.edu.cn/

UCLA

## Integrative Biology and Physiology

The UCLA Department of Integrative Biology and Physiology invites applications for a tenure-track faculty position at the level of Assistant Professor or higher. We seek applicants working in the area of cardiovascular physiology who are committed to training the next generation of scientists and who embrace the concept of collaboration, which is a strength of our UCLA community. In addition, UCLA has a well-established cardiovascular theme (https://medschool.ucla.edu/cardiovascular). In the cover letter, please state how your research will synergize with existing research programs in the Department and within the broader campus community. The successful candidate will be expected to establish an internationally recognized, externally funded research program, in addition to participating in undergraduate and graduate teaching, training, and mentoring. The undergraduate teaching involves contributing to a team-taught core physiology course for undergraduates. The cardiovascular component of the course covers many topics including: cardiac anatomy, blood flow and vascular regulation, cardiomyocyte cell biology, myofilament organization and function, pacemaker activity, excitation-contraction coupling, electrocardiograms, pressure-volume relationships, parasympathetic and sympathetic control, Frank-Starling relationships, as well as basic principles of heart failure, arrhythmia, and atherosclerosis. The applicant is expected to teach the cardiovascular component within the course and to also adapt a growth mindset to teaching excellence as part of the academic mission. Junior faculty in the Department are supported with multiple, well-qualified mentors during their path to tenure.

Please submit CV and statements of research, teaching, mentoring and contributions to equity, diversity, and inclusion. before **September 15, 2019**, to **https://recruit.apo.ucla.edu/JPF04523**. Competitive candidates will have a PhD or equivalent degree within the fields of biology. The faculty in the Department of Integrative Biology and Physiology embraces excellence in diversity and inclusion through the mentoring of under-represented students. Individuals with a history of mentoring students under-represented in the sciences are encouraged to apply and to describe their experience in a cover letter. Queries can be addressed to: Rachelle H. Crosbie, Chair of Integrative Biology and Physiology (rcrosbie@physci.ucla.edu).

**UCLA** is California's largest university, with an enrollment of nearly 38,000 undergraduate and graduate students. The Biosciences at UCLA has more than 300 faculty members, includes many top ten ranked departments, and is consistently in the top ten in NIH funding. As a campus with a diverse student body, we encourage applications from women, minorities, and individuals with a history of mentoring under-represented groups in the sciences.

*The University of California is an Equal Opportunity/Affirmative Action Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, age or protected veteran status. For the complete University of California nondiscrimination and affirmative action policy, see: UC Nondiscrimination & Affirmative Action Policy. (http://policy.ucop.edu/doc/4000376/NondiscrimAffirmAct)*



## Department of Pharmacology
## Yale University

### Assistant or Associate Professor of Pharmacology

The Department of Pharmacology at Yale University School of Medicine invites applications for Assistant or Associate Professor positions in the tenure track. Applicants must have an M.D. and/or Ph.D. or equivalent degree and have demonstrated excellent qualifications in research and education. Areas of interest include studies of molecular function such as proteomics, chemical biology, structural biology (including electron microscopy), and computational biology in different areas of drug discovery, cell signaling, metabolism, genetic diseases, and epigenetic mechanisms.

Responsibilities include establishing a vigorous and independently funded research program in cancer biology while supervising and mentoring students with diverse backgrounds plus contributing to the graduate and medical school educational missions. We seek individuals with strong records of independent creative accomplishments, who will interact productively with colleagues within the Department and across Yale School of Medicine.

Pharmacology at Yale University has a rich history of pre-eminence in drug discovery, indeed the first cancer chemotherapy clinical trials were conducted here. Today, the research in our Department encompasses a broad range of fields, with researchers aiming to both understand and treat human disease, with particular strengths in drug discovery, integrative cell signaling, structural biology, and neuroscience. For more details, see https://medicine.yale.edu/pharm.

Please submit a letter describing qualifications, along with a CV, a two-page summary of current and proposed research, and three letters of reference to the following Interfolio website: **apply.interfolio.com/65128**. Consideration of applications will begin October 1st, 2019. Informal inquiries may be submitted electronically to Prof. Joseph Schlessinger, Chair of Pharmacology, c/o Nathan Kucera (nathan.kucera@yale.edu).

*Yale University is an Equal Opportunity/Affirmative Action Employer. We seek candidates who embrace and reflect diversity in the broadest sense. Yale values diversity among its students, staff, and faculty and strongly welcomes applications from women, persons with disabilities, protected veterans, and underrepresented minorities.*

UR MEDICINE | Orthopaedics & Rehabilitation

Center for MusculoSkeletal Research

The *Center for Musculoskeletal Research* (CMSR: **www.urmc.rochester.edu/musculoskeletal-research.aspx**) at the University of Rochester Medical Center (URMC: **www.urmc.rochester.edu**) is recruiting candidates with expertise in musculoskeletal biology for tenure-track Assistant Professor positions. The successful candidates will benefit from a multidisciplinary musculoskeletal research community, a vibrant graduate program and state of the art infrastructure and core facilities at the University of Rochester. Competitive start-up packages are available. The CMSR has a strong track record and commitment to career development.

Qualified candidates will hold PhD or MD/PhD degrees with an accomplished publication record who study areas relevant to musculoskeletal biology, including but not limited to genomics, cancer biology, cartilage biology, and aging. Candidates at the Assistant Professor level in the Department of Orthopaedics (**www.urmc.rochester.edu/orthopaedics**) are preferred, although outstanding candidates at the Associate and Full Professor levels will be given full consideration, with commensurate expectations of research and funding accomplishment. Secondary departmental affiliations will be determined according to best fit.

The scientific interests pursued in the CMSR include musculoskeletal development, biomechanics and regenerative biology, skeletal pathology, bone cancer, arthritis, osteoimmunology and infections, musculoskeletal stem cell biology and metabolism, tissue engineering and targeted cell- and drug-delivery, population health, and artificial intelligence/machine learning.

Candidates with a strong record of accomplishments should submit a CV, statement of research interests/plans, PDFs of two or three key publications, and arrange for three reference letters to be sent to: **Jaycee_Bristol@urmc.rochester.edu**. Review of applications will begin upon receipt.

*The University of Rochester is an Equal Opportunity Employer and had a strong commitment to diversity and actively encourages applications from candidates from groups underrepresented in higher education.*

EGL
Charitable Foundation

The EGL Charitable Foundation
invites you to apply to the

## Gruss Lipper Post-Doctoral Fellowship Program

### Eligibility

- Israeli citizenship
- Candidates must have completed PhD and/or MD/PhD degrees in the Biomedical Sciences at an accredited Israeli University/Medical School or be in their final year of study
- Candidates must have been awarded a postdoctoral position in the U.S. host research institution.

**Details regarding the fellowship are available at www.eglcf.org**

***Application Deadline is September 29, 2019***

---

**European Innovation Council (EIC) Programme Managers**
**"Join the EIC and Make Europeans world innovation leaders again!"**

You are passionate by radical new and emerging technologies that can lead to disruptive market-creating innovation? You have a vision for and a leading expertise in potential emerging breakthrough technologies in areas such as human-centric artificial intelligence; implantable autonomous devices; zero-emissions energy generation; future technologies for social experience; nanometrology; digital twins for the life-sciences; environmental intelligence? You have demonstrated creativity and determination to realise your vision within a compressed timeframe? You have multidisciplinary experience across academia and industry (including on finance), between the public and private sectors? You have a vast network and collaboration potential across the EU and beyond? You can steer and support teams, set reasonable milestones and seek results and impact? You have excellent written and oral communicating skills? You are a national of one the EU Members States?

**Then, we have just the position for you! The EC is recruiting its first five EIC Programme Managers** for contracts of 3 to 4 years (an initial period up to the end of 2020, renewable once for an additional period of 2 to 3 years) based in Brussels, to play an active role in the origination and management of projects supported by the Enhanced EIC pilot, an ambitious pilot of the forthcoming **European Innovation Council (EIC)** that will focus on detecting, nurturing, and scaling-up EU market-creating and disruptive innovation, from the idea down to market deployment and scale-up.

**Join us to help and accompany EU breakthrough – market creating inventors and innovators, who will generate jobs, strengthen our global competitiveness and improve our daily lives!**

**Submit your candidature before 31 July 2019, 12:00 noon Brussels time.**

Find out more at the following address: **https://europa.eu/!Mn87Nv**

---

ETH zürich

## Professor or Assistant Professor (Tenure Track) of Physics of Soils and Terrestrial Ecosystems

→ The Department of Environmental Systems Science (www.usys.ethz.ch) at ETH Zurich invites applications for the above-mentioned position focusing on (1) physical processes controlling water, energy, and/or mass fluxes in the subsurface and (2) their implications for biogeochemical processes, organisms and/or ecosystems. Candidates are expected to pursue a system-oriented research program linking physical and biological processes across a range of scales. A focus on forest systems and forested landscapes is particularly welcome.

→ The successful candidate is expected to have a strong background in soil physics, physical hydrology, geophysics, or related fields, to have an international track record in research (commensurate with the career stage), and to be a motivated and effective teacher. Teaching duties will include introductory and advanced-level courses in soil physics and environmental physics. In general, at ETH Zurich under-graduate level courses are taught in German or English and graduate level courses are taught in English.

→ The level of the appointment will depend on the candidate's career track and qualifications. Assistant professorships have been established to promote the careers of younger scientists. ETH Zurich implements a tenure track system equivalent to other top international universities.

→ **Please apply online: www.facultyaffairs.ethz.ch**

→ Applications should include a curriculum vitae, a list of publications, a statement of future research and teaching interests, and a description of the three most important achievements. The letter of application should be addressed to the President of ETH Zurich, Prof. Dr. Joël Mesot. The closing date for applications is 15 September 2019. ETH Zurich is an equal opportunity and family friendly employer and is responsive to the needs of dual career couples. We specifically encourage women to apply.

*By* **S. Hussain Ather**

# *Livestreaming science*

I arrive at the lab where I work as a postbaccalaureate researcher and, like many scientists starting their days, I open my laptop, check my email, and plan my agenda. But then, when I'm ready to get to work, I turn on my laptop camera. I perform computational research studying the neuroscience of zebrafish—and I use an online streaming service to share it with the world in real time. I had started to use the platform, called Twitch, a few years ago to watch people play video games. But until last year, it had never occurred to me that I could use it for research, too.

When I used Twitch for video games, I never streamed myself, but I enjoyed watching other gamers play and share their strategies. I also liked interacting with streamers and engaging with other viewers. For me, it was an open, trusting, and supportive environment that offered a great way to build community and learn. (There are a few similar platforms, but Twitch is the most popular among gamers.)

Then I went to a conference, where I met a computational scientist who was considering livestreaming his research. He thought it would be a good way to share his work more widely with other researchers, and to give the general public an intimate window into how research is conducted. He even proposed real-time programming competitions, which would foster community discussion.

***"Anyone who visits my channel can watch me and my computer screen as I work."***

Initially, I thought it was a crazy idea. Twitch is for gaming, not science. But I was intrigued. An internet search revealed a few scientists who streamed their work, and I started to wonder. Maybe there was something to this.

The idea of opening up my work for public viewing and commenting was a little nerve-wracking. I was worried that the whole world would be judging my abilities. But our lab already shares much of our work publicly on GitHub, a website for posting software and code, and streaming seemed like the next logical step. It would be more efficient to get input from users as I worked than to wait until I had updated GitHub. I decided that the possible benefits outweighed the risks. My boss thought streaming my work was the "weirdest thing" and doubted that anyone would want to watch, but he gave his blessing.

So I started to livestream my research for a few hours a day, sharing work that won't compromise my, or my lab's, ability to publish. Anyone who visits my channel can watch me and my computer screen as I work. Sometimes I stream myself reading scientific materials and writing on my blog. One time, I even picked up my laptop and gave a virtual tour of the lab.

My audience is on the small side—some days I have a handful of viewers; other days, I have just one or two—but regardless, the interactions are valuable. Viewers ask me why I choose to run specific programs and about my career goals. They comment on my code and provide feedback. I love engaging with people while still getting work done—and in some cases doing it more effectively than I would have otherwise. One viewer with a bioinformatics background suggested that I use a different tool for one of my analyses, which helped fix a problem I had been stuck on. Another told me about a machine learning tutorial that I have found tremendously useful. Sharing how I write and edit my code—not just the final product—also helps me hold myself accountable to appropriate standards and conventions.

The benefits have extended beyond the nitty-gritty details of my research. I have chatted with my viewers about the philosophy of neuroscience. I've practiced explaining my research in ways a general audience can understand—a skill that will serve me well in my planned career as a scientist and communicator. My boss has even begun to stream some of his research, inspired by my positive experience.

As I open myself to this new method of engagement, I find novel ideas I can embrace and opportunities to grow and improve. Viewers gain insight and trust in science from watching how I do my research. By sharing my work, I show myself to the world. ■

*S. Hussain Ather recently completed a postbaccalaureate research program at the National Institutes of Health in Bethesda, Maryland, and will be starting a master's degree in science communication at the University of California, Santa Cruz, in the fall. Send your career story to SciCareerEditor@aaas.org.*

ILLUSTRATION: ROBERT NEUBECKER