

How special diets might aid
cancer treatments p. 20

What happened to Mars'
liquid water? pp. 27 & 56

Translate schooling to learning:
Education reform in India p. 36

Science

\$15
2 APRIL 2021
sciencemag.org

AAAS

FORM & FUNCTION

Designed proteins assemble antibodies
into precise nanostructures p. 47



CUSPEA PRIZE



Call for Nominations for the CUSPEA Prize:

Young distinguished scientists who have made outstanding contributions in scientific exchanges and collaborations

In 1979, **Dr. Tsung-Dao Lee**, Nobel Prize Laureate in Physics (1957), initiated the "China-U.S. Physics Examination and Application" (CUSPEA) program. Through Dr. Lee's efforts, together with strong support from the American and Chinese physics communities, the CUSPEA program ran for a period of ten years and 915 talented young Chinese students had obtained doctoral degrees from the top graduate schools in North America. Today, there are approximately four hundred CUSPEA scientists who are actively working in various academic institutions worldwide. Among them are numerous fellows of scientific societies and twelve scholars who have been inducted into national academies of sciences. Other CUSPEA scholars are working in various private industries around the world.

Dr. Lee and the CUSPEA program have inspired the spirit of "Science belongs to all humankind", creating an invaluable legacy for international scientific exchanges and collaborations. As the beneficiaries and successors of such spirit and goodwill, the CUSPEA scholars have established the CUSPEA Foundation for the Advancement of Science and Technology with the mission of promoting the legacies, and the CUSPEA Prize is created to award individuals who are extraordinary in such endeavors.

In 2021, the CUSPEA Foundation will launch the first award of the CUSPEA Prize to young distinguished scientists who have made outstanding contributions to promote scientific exchanges and collaborations in any scientific fields. The prize will be awarded to selected individuals regardless of nationality after an impartial review by the CUSPEA Prize Award Committee formed by distinguished scientists and chaired by **Sir Anthony J. Leggett**, Nobel Prize Laureate in Physics (2003).

The prize amount each year: **150,000 USD**

Scientists worldwide are welcome to nominate outstanding candidates for the CUSPEA Prize.

For more details and to submit your nominations, please visit the website at <http://www.cuspeaprizet.org.cn/en/>

Or contact us via email: admin@cuspeaprizet.org.cn

The deadline for the 2021 award nominations is **May 31, 2021**



The CUSPEA Foundation for the Advancement of Science and Technology

Science belongs to all humankind



The fight against cancer starts in the lab

Pioneering cancer early-screening technologies developed by Genetron Health can detect cancerous cells and mitigate their threat.

In just over 5 years, Genetron Health has grown into one of China's leading companies in early cancer screening, diagnosis, and monitoring, as well as biopharmaceutical services. Established in 2015, the firm has expanded to nearly 900 staff, opening four laboratories across the country and one in the United States. Each year, the company processes tens of thousands of patient samples to look for DNA alterations that help guide selection of targeted therapies in cancer patients. The company is also working on expanding its technology for early-detection use.

Chief Medical Officer Yunfu Hu, who joined Genetron Health in 2020, credits the company's success to the passion of its management team and their clear vision for developing innovative technologies that can be used to tackle some of China's most pressing health problems, such as liver cancer. Through active participation in public health projects and collaborations with hospitals and medical centers, these technologies are changing people's lives.

"Currently, Genetron Health is the only company in China participating in national key research and development projects for early screening of liver, lung, and digestive cancers," says Hu, a former U.S. Food and Drug Administration (FDA) officer who won over 10 FDA awards for outstanding contributions to FDA authorization of many novel devices.

Improving lab techniques

Genetron Health offers a range of precision oncology products—from liver-cancer early-screening tests to tumor profiling with targeted next-generation sequencing (NGS) panels and whole-genome sequencing. These products are offered through its central laboratory and in-hospital channels.

Currently, the company is focused on developing NGS products based on two key proprietary technologies. The first is its patented One-Step Seq Method, which significantly reduces the time and effort required to build the library of DNA fragments needed for NGS sequencing. This shortens the labor-intensive activity of creating a DNA library from 24 hours to 1.5 hours and also reduces the amount of laboratory spaces required, making it particularly suitable for hospitals wishing to conduct onsite clinical testing. Hu's team improved the system's ability to detect rare, tumor-specific DNA molecules using blood samples, a critical technological advance for accurate early-disease diagnosis, as the circulating tumor DNA yield of a liquid sample has traditionally been a rate-limiting step.

The second technology is Genetron's proprietary Mutation Capsule system, which enables the detection of both tumor-specific methylation patterns and mutations in cell-free DNA. Most current products detect only one or the other.

This breakthrough is important, says Hu, as the parallel profiling of genetic and epigenetic alterations in a single reaction enables screening for multiple tumor types with enhanced sensitivity, as compared to profiling them separately. The samples can also be retested up to 10 times for other biomarkers.

Powered by innovative and proprietary technology, Genetron Health's liquid biopsy early-screening test for hepatocellular carcinoma, called HCCscreen, received breakthrough device designation from the FDA in September 2020. It is the first device from China's molecular diagnostic industry to receive such recognition. This status should smooth its regulatory progress towards market approval in the United States.

The latest data from a collaborative prospective study between the company and the National Cancer Center of China shows that among a total of 1,615 hepatitis B surface antigen-positive individuals, HCCscreen achieved 88% sensitivity and 93% specificity. In early-stage hepatocellular carcinoma patients, the sensitivity of HCCscreen was 85% and 96% for tumor sizes of <w3 cm and 3–5 cm, respectively.

Hu says the company's strategic approach to early cancer-screening technology differs from its competitors. Instead of developing technologies to identify multiple cancer biomarkers, his team focuses on developing NGS panels that screen for individual cancer types, and then combines these panels for a multicancer test product.

Clearly, it's only a matter of time before patients across China benefit from the work of this pioneering company.

• WEBINAR NOTICE

For more info on liquid biopsies, please view our on-demand webinar, "Early cancer screening today and tomorrow: Exploring liquid biopsy profiling for diagnosis and treatment," available at webinar.sciencemag.org

Sponsored by

GENETRON 泛生子
ANSWERS FOR CANCER

MICHELSON PRIZES: *NEXT GENERATION GRANTS*

The Michelson Prizes: Next Generation Grants support young investigators applying disruptive research concepts to significantly advance the development of vaccines and immunotherapies for major global diseases.

The 2021 Michelson Prizes will be awarded for research proposals in two focus areas:

- Human Immunology and Vaccine Research
- Climate Change and Human Immunology

Researchers from a wide array of disciplines are encouraged to apply.

Are you 35 or younger?
Apply now for the 2021 Michelson Prizes!

www.humanvaccinesproject.org/michelsonprizes



Dr. Danika Hill,
Michelson Prize Winner 2020
Monash University

"The Michelson Prize comes at a critical point in my career, enabling me to follow-up on exciting discoveries that I've made."

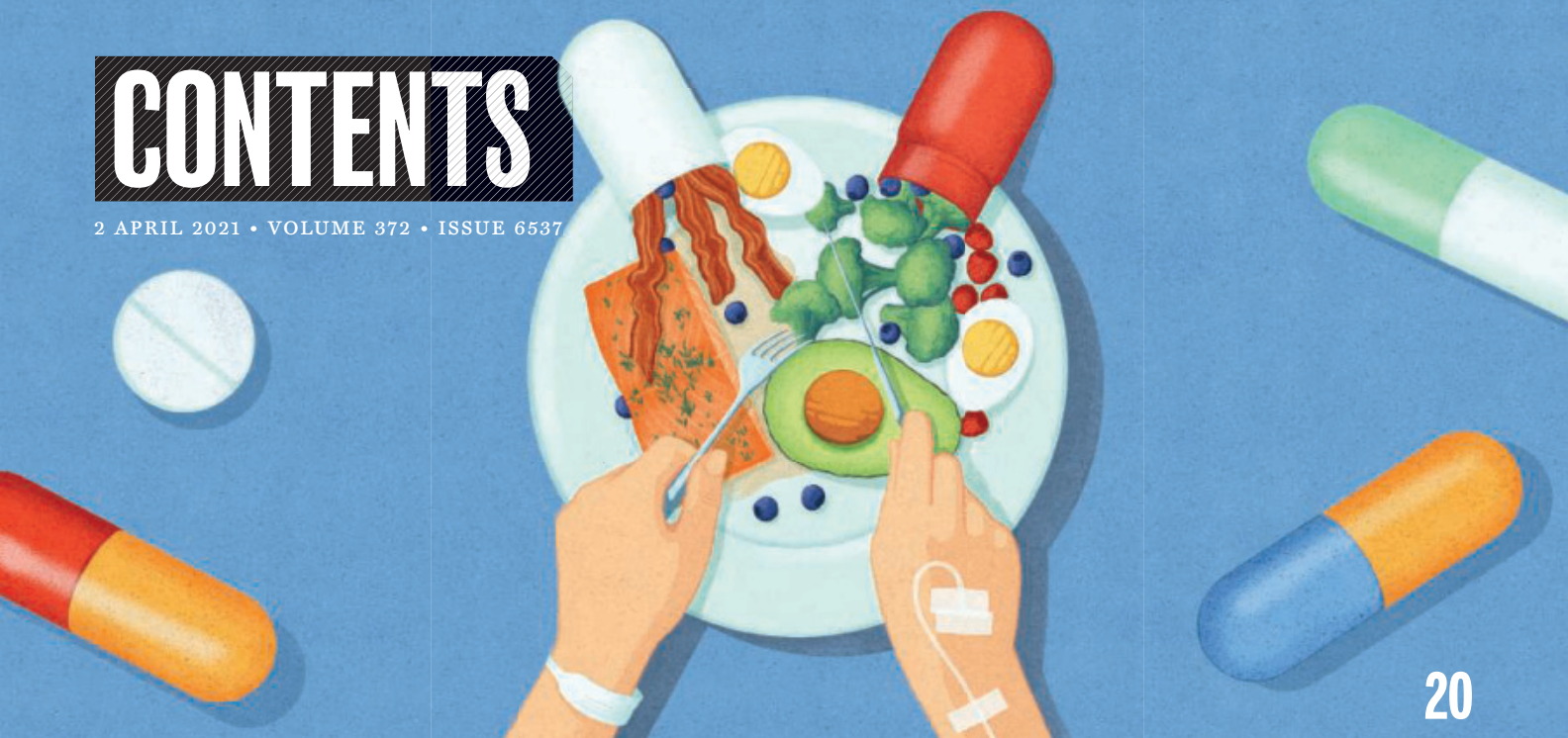
APPLY TODAY

**APPLICATIONS DUE:
JUNE 18, 2021**

**GRANT AWARD:
\$150,000**

CONTENTS

2 APRIL 2021 • VOLUME 372 • ISSUE 6537



20

NEWS

IN BRIEF

10 News at a glance

IN DEPTH

12 India speeds up vaccinations as cases soar again

Relaxed control measures, virus variants, and weather may drive powerful second wave
By V. Chandrashekar

13 Pandemic scientists fight burnout

Running on empty at bedside, bench, and beyond
By M. Wadman

14 Side effect worry grows for AstraZeneca vaccine

Some nations limit shot to older people, as probe of clotting disorders continues
By G. Vogel and K. Kupferschmidt

15 Underwater neutrino traps take shape

Detectors aim to trace high-energy particles to violent cosmic accelerators in deep space
By D. Clery

17 China's scientific treasures tempt foreign collaborators

Outside researchers are offered telescope time and access to Moon rocks, but some are uneasy about the relationship
By D. Normile

18 Researchers coax stripped-down cells to grow normally

Restoring seven removed genes rescues normal division
By M. Leslie

19 U.S. needs solar geoengineering research program, report says

Proposed \$200 million effort could include limited field studies of ways to dim sunlight
By P. Voosen

FEATURES

20 A plateful of medicine

Special diets might boost the power of drugs to vanquish tumors
By J. Kaiser

INSIGHTS

LETTERS

24 NextGen Voices: Forging remote relationships

PERSPECTIVES

27 Hydrated crust stores Mars' missing water

Mars' liquid water might not have been lost to space as thought
By H. Kurokawa
RESEARCH ARTICLE p. 56



28 The impactful origin of neotropical rainforests

A mass extinction event led to vast diversity and structural complexity of neotropical rainforests
By B. F. Jacobs and E. D. Currano
RESEARCH ARTICLE p. 63

30 Achieving fairness in medical devices

Learning from computer science can make medical devices fair for all races and sexes
By A. Kadambi

32 Death and rebirth through nonlinear control

Nonlinearity in complex photonic systems uncovers implications beyond optics
By P. Roztocky and R. Morandotti
REPORT p. 72

33 How dopamine leads to hallucinations

An increase of dopamine in the striatum virtualizes a nonexistent auditory signal in mice
By M. Matamalas
RESEARCH ARTICLE p. 51

34 Reversal of exhaustion in engineered T cells

Antitumor activity of exhausted engineered T cells is improved with pulsatile signaling
By M. Mamonkin and M. K. Brenner
RESEARCH ARTICLE p. 49

POLICY FORUM

36 India's new National Education Policy: Evidence and challenges

Gains in schooling must be translated into gains in learning
By K. Muralidharan and A. Singh



NOMIS & Science YOUNG EXPLORER AWARD

Research at the intersection of the social and life sciences

Unconventional. Interdisciplinary. Bold.

The newly launched **NOMIS & Science** Young Explorer Award recognizes bold, early-career researchers with an M.D., Ph.D., or M.D./Ph.D. who ask fundamental questions at the intersection of the social and life sciences. It is awarded to scientists who conduct research with an enthusiasm that catalyzes cross-disciplinary collaboration and who take risks to creatively address relevant and exciting questions.

Award winners will receive a cash prize of up to \$15,000, and their essay submissions will be published in *Science*. They will also be invited to share their ideas with leading scientists in their respective fields at an award ceremony and to present their research to the NOMIS Board of Directors for potential future funding.

Apply by May 15, 2021
at www.sciencemag.org/nomis

Science
AAAS



NOMIS
FOUNDATION
Creating the Spark

BOOKS ET AL.

40 Stewart Brand's radical environmentalism

The counterculture icon continues to embrace an outside-the-box approach to the future *By W. P. McCray*

41 Stephen Hawking, celebrity scientist

A new biography considers the cosmologist's reputation while placing his work in context *By D. Fahy*

RESEARCH

IN BRIEF

43 From *Science* and other journals

REVIEW

46 Animal culture

The burgeoning reach of animal culture
A. Whiten
REVIEW SUMMARY; FOR FULL TEXT:
DOI.ORG/10.1126/SCIENCE.ABE6514

RESEARCH ARTICLES

47 Protein design

Designed proteins assemble antibodies into modular nanocages *R. Divine et al.*
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DOI.ORG/10.1126/SCIENCE.ABD9994

48 Human genomics

Haplotype-resolved diverse human genomes and integrated analysis of structural variation
P. Ebert et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DOI.ORG/10.1126/SCIENCE.ABF7117

49 Cancer

Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling *E. W. Weber et al.*
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DOI.ORG/10.1126/SCIENCE.ABA1786
PERSPECTIVE P. 34

**50 Neuroscience**

Gamma rhythm communication between entorhinal cortex and dentate gyrus neuronal assemblies *A. Fernández-Ruiz et al.*
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DOI.ORG/10.1126/SCIENCE.ABF3119

51 Neuroscience

Striatal dopamine mediates hallucination-like perception in mice *K. Schmack et al.*
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DOI.ORG/10.1126/SCIENCE.ABF4740
PERSPECTIVE P. 33; PODCAST

52 Transcription

Structure of the human Mediator-bound transcription preinitiation complex
R. Abdella et al.

56 Mars

Long-term drying of Mars by sequestration of ocean-scale volumes of water in the crust *E. L. Scheller et al.*
PERSPECTIVE p. 27

63 Paleoecology

Extinction at the end-Cretaceous and the origin of modern Neotropical rainforests *M. R. Carvalho et al.*
PERSPECTIVE p. 28

REPORTS

68 Interface physics

A van der Waals interface that creates in-plane polarization and a spontaneous photovoltaic effect
T. Akamatsu et al.

72 Topological optics

Nonlinear tuning of PT symmetry and non-Hermitian topological states *S. Xia et al.*
PERSPECTIVE p. 32

76 Catalysis

Isolated boron in zeolite for oxidative dehydrogenation of propane *H. Zhou et al.*

81 Pesticides

Applied pesticide toxicity shifts toward plants and invertebrates, even in GM crops
R. Schulz et al.

84 Marine conservation

At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts
C. C. O'Hara et al.

88 Quantum optics

Observation of a non-Hermitian phase transition in an optical quantum gas
F. E. Öztürk et al.

91 Neurogenomics

Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons
D. A. Reid et al.

DEPARTMENTS

8 Editorial

Reckoning with Asian America
By Jennifer Lee and Tiffany J. Huang

9 Editorial

Expanding the endless frontier
By Robert W. Conn

98 Working Life

Choosing from the heart
By Phil De Luna



ON THE COVER

Computationally designed proteins (green) direct the assembly of almost any antibody or Fc-fusion protein (yellow) into highly ordered nanostructures with symmetries ranging from dihedral to icosahedral (schematic of architecture in black). The antibodies contribute both form and function: The dimeric constant regions on the twofold symmetry axes are a key structural element of the nanocage, whereas the outward-facing variable regions endow the nanocage with a specific targeting or signaling function. See page 47.
Illustration: V. Altounian/Science; Data: PDB IDs 1IGT, 1L6X, and 4K0V; design models available at <http://doi.org/10.5281/zenodo.4599680>



Science Careers96

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. Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$2148; Foreign postage extra: 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 available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Reckoning with Asian America

Jennifer Lee

is the Julian Clarence Levi Professor of Social Sciences in the Department of Sociology at Columbia University, New York, NY, USA. lee.jennifer@columbia.edu

Tiffany J. Huang

is a graduate student in the Department of Sociology at Columbia University, New York, NY, USA. tjh2150@columbia.edu

It took the mass murder of six Asian women in Atlanta last week to draw national attention to what Asian Americans have been warning about since the wake of the pandemic: anti-Asian violence. The incident reflects an under-recognized history of anti-Asian violence and discrimination in this country that dates back more than 150 years. This needs to change. Asian Americans must become central to the discourse on race in America. For the country to “care” about the outcry by Asian Americans, the public needs to understand how America got to this point.

This moment of crisis has been building over the past pandemic year. Many consider the recent anti-Asian violence and harassment a consequence of the former Trump administration’s “China virus” and “Kung flu” rhetoric. Research shows that Americans exposed to such racist rhetoric are more likely to perceive Asian Americans as foreign and un-American, which can stoke greater hostility toward Asians. An AAPI Data survey conducted just after the Atlanta shooting shows that 71% of Asian American adults worry about COVID-19-related hate crimes, harassment, and discrimination, 21% of whom worry very often. The survey data also suggest that upwards of 2 million Asian American adults have experienced anti-Asian hate incidents since the onset of COVID-19: 1 in 8 Asian American adults in 2020, and 1 in 10 in the first quarter of 2021.

U.S. history is fraught with anti-Asian violence, misogyny, nativist discrimination, and legal exclusion, all of which are often absent in textbooks and university curricula. The 1882 Chinese Exclusion Act emerged from the earlier Page Act, which excluded presumed immoral Chinese women from immigrating. Without their wives, male Chinese laborers who helped build the transcontinental railroad segregated into tight-knit bachelor communities that became the precursors of today’s Chinatowns. Legal exclusion was coupled with violence. One of the largest mass lynchings in American history took place in Los Angeles in 1871, when 19 Chinese residents—10% of the city’s Chinese population—were killed by a white mob. In the Rock Springs Massacre of 1885, white miners killed 28 Chinese workers, wounded 15, and expelled hundreds more before setting their living quarters on fire.

The nativist prejudice that white settlers imparted, and the legal exclusion that Chinese endured, engineered the

very conditions that would shape societal perceptions of the Chinese in the 19th century as economic threats, clanish, untrustworthy, foreign, and immoral. These views would continue to mark Chinese Americans and evolve into the racial stereotypes of Asian Americans today—untrustworthy, passive, demure, hypersexual, and America’s insidious “model minority.”

Academia has not been immune to the effects of this history on institutional racism, bias, exclusion, and violence. COVID-19-related anti-Asian messages and harassment have been reported on college campuses across the country. Chinese American scientists have come under federal scrutiny for their associations with China under the 2018 China Initiative, which may jeopardize U.S.–Chinese scientific collaborations. And despite being

the group most likely to attend college, Asian Americans make up a mere 2% of college presidents. Asian Americans are the least likely among all women to be promoted to leadership positions and make up less than 1% of top earners at those universities engaging in the highest level of research activity. Anti-Asian bias also affects students. In one study, researchers sent emails with names signaling race and gender to 6548 professors, posing as prospective PhD students. Professors were the least likely to respond to

those who had Chinese and Indian names.

If universities and precollege schools fail to teach the history of Asian Americans in their curricula, we can expect bias and exclusion to perpetuate in our institutions. Asian American student activists in the 1960s understood this. They coined the term “Asian American” as a unifying political, pan-ethnic identity to advocate for Asian American Studies and build coalitions with African Americans, Latino/as, and women. Many Americans are unaware of this history, including Asian Americans, most of whom are immigrants who arrived after 1965. Today, Asian Americans encompass more than two dozen national origin groups with vastly different migration histories, languages, and socioeconomic statuses. Yet during the pandemic, they have shared a common fear of harassment, discrimination, and anti-Asian violence.

Violence and bigotry against Asian Americans have finally received national attention. We must make Asian Americans central to the country’s discussions of race, and reckon with the history of Asian America.

—Jennifer Lee and Tiffany J. Huang

“Asian Americans must become central to the discourse on race in America.”

Expanding the endless frontier

This month, the U.S. Senate is poised to consider legislation that would expand the National Science Foundation (NSF) and strengthen the U.S. science and technology research ecosystem. The heart of the legislation will be the Endless Frontier Act (EFA), a bipartisan and bicameral bill that was first introduced to the previous Congress in May 2020. With some modifications, this legislation could become a landmark achievement that bolsters U.S. competitiveness.

The bill would authorize \$100 billion over 5 years for a new Directorate for Technology to support basic science motivated by critical needs, often referred to as “use-inspired” basic research. The initial areas of focus would include artificial intelligence, quantum information sciences, and advanced materials. In addition to other provisions that protect NSF’s current mission and budget (\$8.5 billion in fiscal year 2021), the bill requires that the Directorate for Technology provide a minimum of 15% of its annual funds to enhance NSF’s existing areas of curiosity-driven research in its existing directorates. The new directorate also would fund academic research both by individuals and centers; offer support to undergraduates, graduate students, and postdoctorals; and enable universities to develop new ways to spin off and support companies on the basis of their discoveries and inventions.

Over the past 7 months, a group of scientific leaders that David Baltimore and I assembled* has been reviewing the bill and meeting with key people in Congress. We developed a short list of changes that would better ensure the success of the bill and its intentions. Our concerns are substantial but still minor in comparison with the benefits that the EFA would confer on the nation’s science and technology enterprise.

Maintaining NSF’s unity of structure is critical—a single director and board that makes certain that the agency’s work is greater than the sum of its parts. With this in mind, we do not support a provision in the May 2020 version that creates a Senate-confirmed head of the new Directorate for Technology who might be viewed as outranking the other directorate leaders and possibly competing with the NSF director. As well, a proposed new advisory board that includes congressionally appointed members would be equally disruptive.

The legislation also changes NSF’s name to the National Science and Technology Foundation. We recommend that the agency’s well-recognized name NSF be preserved unchanged, given its acclaimed history and position in science.

The bill does protect NSF’s existing programs, and these could be further strengthened. For example, it prevents the Directorate for Technology from making new grants if budgets for the existing directorates are not maintained. The bill could also stipulate additional funding for these directorates to ensure that NSF’s budget does not become too weighted toward use-inspired basic research. And Congress should design the ramp-up of the Directorate for Technology’s budget in a manner that gives stability and reasonable predictability commensurate with the long-term nature of basic scientific work.

The bill’s education language should be strengthened to encourage needed experimentation in the way that students are trained. With the country’s history of underrepresentation of many groups in science, technology, engineering, and mathematics (STEM), the new bill should promote new ideas garnered from experts in this area to attract diverse students into STEM fields. This is not only the

right thing to do but would address the losses that the United States suffers when a substantial portion of the population is not welcomed into the nation’s scientific enterprise. Also, the bill should encourage graduate programs to give students experience in industry and government as part of their training, and it should fund more professional, but not high-fee, professional master’s programs, especially in engineering.

We urge the scientific community to give the EFA its constructive attention and its vocal support. The Association of American Universities and the Association of Public and Land-Grant Universities both support the bill and provide helpful summaries as a guide. We have provided recommendations to congressional committees as well.

This is a rare moment, a once-in-a-generation opportunity to enhance the nation’s research enterprise and adapt it to current challenges for the benefit of the country and the world. We urge our colleagues to engage now.

—Robert W. Conn



Robert W. Conn is the past president and chief executive officer of The Kavli Foundation and dean emeritus and the Walter Zable Distinguished Professor of Applied Physics, Emeritus at the Jacobs School of Engineering, University of California, San Diego, San Diego, CA, USA. rconn@ucsd.edu

“This is a rare moment... to enhance the nation’s research enterprise...”

*This editorial represents the collective views of David Baltimore, Robert Conn, William Press, Thomas Rosenbaum, David Spergel, Shirley Tilghman, and Harold Varmus.

IN BRIEF

Edited by Jeffrey Brinard

ASTROPHYSICS

Black hole yields new, polarized look

The team that in 2019 used a global network of radio telescopes to reveal the first image of a black hole has offered a new twist on that iconic view: the same black hole in polarized light. The thin lines spiraling in toward the black hole's shadow (above) show areas of light that differ in their polarization—the direction in which the light waves vibrate. The light, from plasma near the black hole's edge, was polarized by magnetic fields, and so the new image, described last week in *The Astrophysical Journal* by the Event Horizon Telescope team, indicates their structure. Researchers hope to learn how the fields help accreting black holes funnel matter and energy into jets emanating from their poles.

Pandemic origin report blasted

PUBLIC HEALTH | The United States and 13 other countries this week criticized a report by a World Health Organization panel that had visited China to investigate how the COVID-19 pandemic started. The 300-page document says the most likely cause was a bat coronavirus that infected another, unidentified animal and then moved to humans, but it recommends further research. The report's

most definitive conclusion is also its most controversial: that it is “extremely unlikely” that SARS-CoV-2 came out of a Chinese laboratory. Scientists from China made up half of the 34-member international panel. A joint statement by other countries complained that the investigation was “significantly delayed and lacked access to complete, original data, and samples.” It called for a transparent, “rapid, independent, expert-led, and unimpeded evaluation of the origins.”

NSF tech directorate proposed

FUNDING | The science committee in the U.S. House of Representatives wants to more than double the budget of the National Science Foundation (NSF) in the next 5 years, from \$8.5 billion to \$18.3 billion. A sizable chunk of the extra money—\$5 billion by 2026—would go to a new directorate, Science and Engineering Solutions, that would accelerate the conversion of basic research into new technologies and products. Last year, Senate Majority Leader Chuck Schumer (D-NY) proposed growing NSF to \$100 billion over 5 years, with roughly one-third of that money going to a new technology directorate. Schumer's vision for NSF is part of still-evolving draft legislation affecting many federal agencies that pinpoints key technologies needed to address economic and security threats posed by China's growing technological prowess. In contrast, the House bill is limited to NSF's programs and is aimed at strengthening basic research across all disciplines that NSF supports. The House and Senate would need to agree on a vision for NSF, and other legislation would be needed to appropriate the money.

Satellite glow jeopardizes scopes

ASTRONOMY | Light pollution from space junk and satellites may have already robbed the entire Earth of the dark skies best for sensitive astronomical observations, an analysis has found. Researchers estimated the size and shininess of tens of thousands of objects in orbit as of 2020, before an onslaught of thousands more satellites that companies plan to launch in the coming years. Even at Earth's darkest sites, the sky glows from natural sources such as ionized particles; but the existing orbiting objects reflect and scatter about 10% more of this diffuse light back into the atmosphere, the research team calculates in a paper accepted this week by the *Monthly Notices of the Royal Astronomical Society*. That extra amount violates an International Astronomical Union standard for observing sites and could compromise observations of the dimmest galaxies, which scientists study for clues about the physics of galaxy formation and the nature of dark matter. To

gather such data, astronomers already need long exposures on the biggest telescopes at the darkest available sites.

Harvard rebukes Epstein associate

ETHICS | Harvard University last week penalized quantitative biologist Martin Nowak for his connections with disgraced financier Jeffrey Epstein. Epstein had donated \$6.5 million for Nowak's research in 2003; after being convicted in 2008 of soliciting prostitution from a minor, Epstein introduced Nowak to donors who provided an additional \$7.5 million. Nowak's actions after 2008—repeatedly hosting Epstein on campus, promoting Epstein on his program's web page, and providing false information about Epstein's support in a grant application—violated Harvard policies, and other actions showed “blameworthy negligence and unprofessional behavior,” Claudine Gay, dean of arts and sciences, wrote in an email last week to faculty members. Nowak will continue at Harvard as a math professor, but his

Program for Evolutionary Dynamics will be shut down and he will be barred for at least 2 years from serving as a principal investigator on grants. “I regret the connection I was part of fostering between Harvard and Jeffrey Epstein,” Nowak said in a statement last week. Epstein died by suicide in 2019.

China yields more unusual relics

ARCHAEOLOGY | Chinese archaeologists last week reported unearthing more than 500 artifacts, including gold ornaments, bronze heads, ivory and jade tools, and a gold mask dating back about 3000 years at the Sanxingdui archaeological site in southwestern Sichuan province. Sanxingdui, then ruled by the Shu kingdom, has already yielded thousands of bronze relics unlike anything found elsewhere in China, including at sites of the contemporaneous Shang dynasty in the Yellow River region. The new finds, retrieved from what are thought to be sacrificial pits, may shed light on how the Shu kingdom contributed to Chinese civilization.

IN OTHER NEWS

VACCINE LEADER FIRED Moncef Slaoui, who headed COVID-19 vaccine development during the Trump administration, has been fired as chairman of a medical research firm controlled by manufacturer GlaxoSmithKline after he was accused of sexual harassment. The company said an outside investigation substantiated the allegation by a female employee about Slaoui's behavior several years ago when he worked there. Slaoui also stepped down from leadership roles at two other pharmaceutical companies and issued a statement in which he apologized to the woman and his family.

RETURNING LOOTED ART Museums in Germany have pledged to return hundreds of artifacts, including bronze statues, looted during the colonial era from the kingdom of Benin in what is now Nigeria. The British Museum and others face growing pressure to join them.

PARDON SOUGHT The Australian Academy of Science issued a statement saying a court ignored new genetic evidence when it denied last week an appeal by a woman convicted of killing her four young children. Tests point to a natural cause of the deaths: Two of the children carried a mutation in the *CALM2* gene that is associated with sudden death by cardiac failure in infants and children. Prosecutors had accused Kathleen Folbigg of smothering the children but have not presented medical evidence that supports that position. Academy members have signed a petition asking New South Wales's governor to pardon her.

AI IN MEDICINE The Broad Institute has received \$300 million to study how machine learning can improve the prevention and treatment of disease. Half the sum is coming from a foundation of Wendy and Eric Schmidt, a member of Broad's board and former CEO of Google, and the rest from the Broad Foundation.

R&D SPENDING RISE The United States spent more than 3% of gross domestic product on R&D in 2019 for the first time. The 3.07% share is a record and met a goal set by former President Barack Obama a decade ago. Israel led globally with 4.9%, the Organisation for Economic Co-operation and Development said. Total U.S. spending was more than any other country's.



The golden lion tamarin (*Leontopithecus rosalia*) is among Brazil's many endangered species.

CONSERVATION

Brazil joins treaty on biodiversity and biotrade

Despite the antienvironmental policies of its current leadership, Brazil has become the 130th country to ratify the Nagoya Protocol, a part of the Convention on Biological Diversity that lays out measures to protect countries' biodiversity claims, the CBD announced last week. The ratification, first proposed by a previous administration in 2012, had languished until 2019, when rampant deforestation led pro-environment leaders to push for approval. The current government is seen as having consented because the protocol allows nations to impose rules on the international trade in its plant and animal products; by legitimizing the sales, the regulations are expected to increase exports and tax revenues. For example, money from sales of native plants such as açai (*Euterpe oleracea*) and Brazil nut (*Bertholletia excelsa*) could be returned to help Indigenous communities that use and harvest them. Observers question whether the ratification alone will protect Brazil's biodiversity, perhaps the world's greatest—but hailed the step as helpful.



IN DEPTH

Senior citizens wait in line at a COVID-19 vaccination center in Mumbai, India.

COVID-19

India speeds up vaccinations as cases soar again

Relaxed control measures, virus variants, and weather may drive powerful second wave

By Vaishnavi Chandrashekhar

Just over 1 month ago, many Indians believed the pandemic was winding down. Cases of COVID-19 had declined continuously and dramatically for five straight months, travel restrictions had been lifted, and wedding season was in full swing.

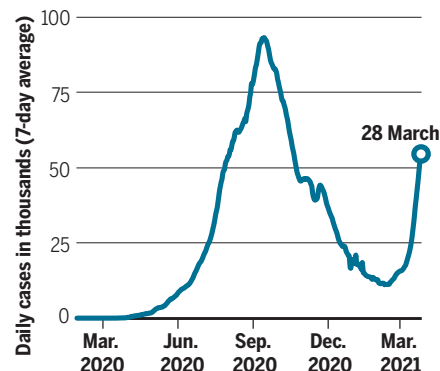
But now a second wave is hitting. Nationwide, cases have soared from just over 11,000 daily in mid-February to more than 60,000 per day as *Science* went to press, more than half of them in Maharashtra state, of which Mumbai is the capital. The remainder are concentrated in seven other states, but scientists worry the disease may soon surge across the country again. India is fighting the rise with new restrictions and efforts to step up vaccination. But although the country produces two authorized COVID-19 vaccines, its immunization campaign has yet to gather steam.

Coming after India's first giant wave of cases, which peaked in September 2020, the spectacular decline defied dire predictions. Antibody surveys, which suggested densely populated areas in cities such as New Delhi and Mumbai were near herd immunity, raised hopes that transmission was burn-

ing out. But the optimism may have been unfounded; a more recent survey across 700 districts found only about 22% of Indians had been exposed overall. Meanwhile, control measures such as wearing masks were loosened, travel and social gatherings increased, and testing and contact tracing stumbled. "We let our guard down too quickly," says virologist Shahid Jameel of Ashoka University.

Exponential growth returns

After a sharp 5-month decline, COVID-19 cases are rising rapidly in India. Most new cases are in eight states, but scientists worry about a nationwide surge.



Mutations may also be reigniting the pandemic. Just over 800 of more than 11,000 samples sequenced in recent months tested positive for B.1.1.7, a variant first discovered in the United Kingdom that is known to be more infectious. In Punjab, it was detected in 81% of 400 sequenced samples. Scientists are also investigating a variant with two mutations, E484Q and L452R, found in certain districts that are seeing an exceptional surge in cases. The two mutations are associated with "immune escape," or an ability to elude antibodies, and increased infectivity, health ministry officials said last week, although there is no evidence yet that this variant is causing the surge.

Climate could play a role as well. In Europe and the United States, the winter drives people indoors, where the virus spreads easily. In India, the increasing heat of spring may lead people to retreat to the fans and air conditioners of their homes, says epidemiologist Prabhat Jha, director of the Centre for Global Health Research, which has offices in India and Canada.

Meanwhile, less than 5% of India's 1.3 billion people have received at least one dose of vaccine. The government is striving to accelerate the pace, now about 2 million

to 3 million shots per day; on 23 March it announced that everyone over age 45 can get a shot starting 1 April. The AstraZeneca vaccine, manufactured by the Serum Institute of India, accounts for most of the shots delivered so far. The other locally produced vaccine, Covaxin, was developed by Bharat Biotech in collaboration with the Indian Council of Medical Research.

India has reportedly put on hold exports of the AstraZeneca vaccine to help meet domestic demand. Since January, India had exported 60 million doses to some 80 countries, through bilateral aid, commercial contracts, and the COVID-19 Vaccines Global Access Facility, a global scheme to increase access to the vaccine.

Vaccine coverage among the country's poor is lowest, because of low awareness and day workers' inability to take time off, social workers say. In Mumbai, authorities have begun to set up vaccination centers in slums. Also needed, says Arun Kumar, head of Apnalaya, a nonprofit that works in the city's slums, are "massive community-based programs to clear vaccine fears."

Some of those fears stem from Covaxin's hasty approval in early January, before data from phase 3 trials were available. "It created a doubt," says former federal health secretary K. Sujatha Rao. "Once trust is broken, it's not easily regained." (On 3 March, Bharat Biotech announced the vaccine had 81% efficacy, based on an as-yet-unpublished initial analysis of 43 cases.)

Reports of violations of informed consent in trials and inadequate transparency around adverse events may have also shaken confidence. On 16 March, a group of 29 doctors and researchers wrote a letter about reported deaths, about 100 so far, of vaccinees. Although the vaccines may not be responsible, the petitioners say, the government should investigate them and disclose its findings. Unlike at least 20 European countries, India has not paused use of the AstraZeneca vaccine after reports of serious clotting disorders (see p. 14); officials say they are reviewing the data.

In an attempt to slow the second wave, several states and cities have reintroduced curbs on social gatherings, imposed temporary lockdowns, and stepped up testing and tracing. In Mumbai, once again a hot spot of the pandemic, the city banned public celebrations of the spring festival of Holi. "We were lucky compared to what might have been," says epidemiologist Giridhar Babu of the Public Health Foundation of India. "But the story is not over. The virus keeps surprising us." ■

Vaishnavi Chandrashekar is a journalist in Mumbai, India.

COVID-19

Pandemic scientists fight burnout

Running on empty at bedside, bench, and beyond

By Meredith Wadman

When not caring for COVID-19 patients—her latest was a man with bacterial lung and blood infections superimposed on SARS-CoV-2 pneumonia—Krutika Kuppalli has been helping oversee the rollout of pandemic vaccines at the Medical University of South Carolina (MUSC), where she's an infectious disease physician. She has also been meeting with vaccine-hesitant hospital staff, sitting on a committee that reviews all planned COVID-19 clinical trials at MUSC, applying for funding to study patients with Long COVID, and handling online harassment that has followed her numerous media appearances and two rounds of congressional testimony last summer.

Asked recently during a Zoom interview how she is doing, she paused for nearly 20 seconds, struggling to regain her composure. "We have been busting our butts for 12, 14 months," she says. "I just feel I'm empty."

From academic research centers to intensive care units (ICUs) to scientific journals to government agencies, scientists fighting the pandemic say they are hitting a wall, 15 months after the first report of a cluster of cases of pneumonia in Wuhan, China, introduced the virus that would upend their lives. "The pace that led to the incredible generation of knowledge on SARS-CoV-2 and COVID-19 has put enormous demands on the people who are expected to generate that knowledge," says David O'Connor, a viral sequencing expert at the University of Wisconsin, Madison, who has been tracking the spread of the virus, doing Zoom Q&A sessions with the vaccine hesitant, and helping neighborhood schools set up diagnostic testing. "This is a terrible time and we should all do what we can to help. But is it going to be sustainable?"

Throughout higher education many are feeling a strain from campus closures, remote teaching, disrupted research, work-from-home challenges, and more. For example, a survey of more than 1100 U.S. faculty members found 55% had seriously considered changing careers or retiring early because of the pandemic. The survey, conducted in Oc-

tober 2020 by *The Chronicle of Higher Education* and Fidelity Investments, also found that 69% of respondents felt stressed, 68% felt fatigued, and 35% felt angry—more than double the 2019 numbers. An international survey by the publishing house De Gruyter found a similarly bleak picture among medical and life scientists, specifically, although the numbers were small: Of 116 respondents, 76% said the pandemic had impacted their well-being; 30% said the impact was "severe."

Physicians have borne much of the burden, says Mona Masood, a Philadelphia-area psychiatrist and founder of the Physician Support Line, a free call-in service staffed by volunteer psychiatrists. The help line has fielded calls from more than 2000 people since it launched 1 year ago; calls have peaked during pandemic surges, Masood says. Some are from doctors on the front lines—for example, an ICU physician who had just lost his 20th patient and broke down on the phone

after a 48-hour shift. Others come from physician-scientists who, despite their work on vaccines or variants, feel guilty that they are not caring for COVID-19 patients alongside their colleagues.

For junior scientists, the crisis has magnified stresses already present in the academic system.

"Everyone is working nights, weekends, every spare minute of their lives. There's no extra pay. There's no guarantee of any extra recognition," says Emma Hodcroft, a computational biologist and postdoc at the University of Bern who has been tracking SARS-CoV-2 evolution for the project Nextstrain. "I am precariously employed; I don't have a long-term job. I feel a lot of pressure that this is my opportunity and I cannot waste that," says Hodcroft, who has been an author on 18 SARS-CoV-2 papers and preprints since February 2020.

Some academic scientists—especially those with young children—say their institutions have done little to alleviate their stress. "A few 'atta-boys' are tossed by the Provost to thank faculty for their flexibility with coping with challenging times, but no real differences implemented," one senior lecturer commented in a recent National Academies of Sciences, Engineering, and Medicine report that found a disproportion-

Science's COVID-19 reporting is supported by the Heising-Simons Foundation.

ate, deleterious impact of COVID-19 on the careers of women in academic science, technology, engineering, math, and medicine.

The demands of reviewing or editing an unceasing glut of coronavirus papers have also been backbreaking for some. “They did replace me with two people. That tells you what my workload was,” says one biology manuscript editor who recently left a top-tier journal and asked not to be identified for fear of career repercussions.

Scientific societies have, in some cases, tried to step into the breach. The American Association for Anatomy (AAA) in October 2020 launched a website of self-care resources called THRIVE in response to “cries for help” from members, largely Ph.D. anatomists who teach and conduct research in medical schools, says Shawn Boynes, AAA’s executive director and the prime mover behind THRIVE. The website, which is open to all, has had roughly 3000 monthly visitors, but Boynes says it is at best a Band-Aid. “Why isn’t more of this addressed at the institutional level on a regular basis? The pandemic pulled the curtain back and you can see how unbelievably challenging it really is with people who choose academia as a career.”

The challenges extend beyond academia. At the Centers for Disease Control and Prevention (CDC), for example, the burden of being the front-line agency responding to the pandemic has taken a major toll. “This question of burnout, personal and professional, is the No. 1 thing I talk about with friends” at the agency, says a senior CDC epidemiologist who has worked at a high level in the pandemic response. Some mid-career CDC scientists are talking about early retirement—a choice that was almost unheard of before COVID-19, they add.

Still, other scientists grappling with the pandemic say that despite the pressure, they have never felt more fulfilled. These past 15 months have been “the hardest of

my life,” says Sarah Schmedes, lead bio-informatician at Florida’s Bureau of Public Health Laboratories. Schmedes is adjusting to being a single mother to a 1-year-old son after her husband died of a heart attack in December 2019. “It definitely helps that I love my job. Being in this field at this time is incredibly rewarding. I’m very honored.”

Marion Koopmans, chief of the viroscience department at Erasmus Medical Center in Rotterdam, Netherlands, says she has been working at least 8 a.m. to 11 p.m., 6 or 7 days a week since the pandemic began. Recently, she traveled to Wuhan as part of a World Health Organization team investigating the pandemic’s origins (see p. 10), while still managing a 150-person lab in Rotterdam. “I don’t feel like I’m burning out,” she says. “I can actually do things that contribute to getting through the pandemic. At least that’s what it feels like. And that helps.” Still, she says, she is taking measures to protect a modicum of personal time so she can watch *Anne with an E* on Netflix or share a nightcap with her husband. “I ask the strictest secretary to block my calendar and route my emails through her.”

But O’Connor emphasizes the need for the community to take broader steps. “A year into this, we need to assess: What does the future look like? Does it look like the same workforce being asked to do twice as much as they were doing before? I don’t know what the right answer is, but the number of times recently I’ve heard ‘I’m just done with this’ uttered in frustration from friends and colleagues is really concerning.”

Kuppalli for her part does not foresee an end to her own exhaustion, in part because whenever she is asked to do one more thing to fight the pandemic, “I don’t feel like I can say no. Because it’s larger than me and I feel lucky to be in a position to contribute.” ■

With reporting by Charles Piller.

COVID-19

Side effect worry grows for AstraZeneca vaccine

Some nations limit shot to older people, as probe of clotting disorders continues

By Gretchen Vogel and Kai Kupferschmidt

It’s been one step forward, two back for AstraZeneca’s COVID-19 vaccine. Even as the company rebutted criticism of its efficacy claims last week, a bigger problem loomed for the vaccine and the many millions depending on it. Evidence continues to accumulate that an unusual clotting disorder seen in dozens of European recipients is a real, albeit rare, side effect. A preprint has detailed a proposed mechanism, and multiple scientific groups have said the worry is legitimate and must be seriously weighed against the vaccine’s COVID-19 protection.

This week, Canada and Germany joined Iceland, Sweden, Finland, and France in recommending against the vaccine’s use in younger people, who seem to be at higher risk for the clotting problem and are less likely to develop severe COVID-19. The approach makes sense given that other vaccines are available, says Sandra Ciesek, a virologist at Goethe University Frankfurt. “We do not have just one vaccine. We have several.”

AstraZeneca’s vaccine incorporates the spike gene from SARS-CoV-2 into another, nonpathogenic virus. Last month, many countries suspended its use following initial reports of the clotting issues in recipients, which have led to at least 15 deaths in Europe. Some researchers dismissed the cases as normal background levels of blood clots. And most countries resumed vaccinations after the European Medicines Agency (EMA) said the vaccine’s benefits outweigh any risks, even though it couldn’t rule out that the clotting symptoms were connected to the vaccine.

In parallel, the company’s initial report of results from a key vaccine trial in the Americas drew unexpected criticism. In a press release, AstraZeneca claimed the trial had shown the vaccine had 79% efficacy in pre-



ILLUSTRATION: ROBERT NEUBECKER



Italy resumed administering AstraZeneca's vaccine, but some other countries now restrict its use by age.

venting symptomatic disease. But the independent board overseeing the trial protested that the data the company put out were “outdated” and potentially misleading. Two days later, AstraZeneca revised the efficacy down to 76%, leaving observers baffled by the dispute, but mostly reassured.

The potential side effect, on the other hand, does not seem to be going away. An EMA expert group discussed it on 29 March, but the agency issued no immediate public update; EMA's risk assessment panel will evaluate the issue next week.

The highly unusual combination of symptoms—widespread blood clots and a low platelet count, sometimes associated with bleeding—has so far been reported from at least seven countries. Medical societies around the world have warned members to be on the lookout for the clotting disorder in vaccine recipients and report it. Estimates of the incidence range from one in 25,000 people given the AstraZeneca vaccine in Norway to at least one in 87,000 in Germany. “People are absolutely working like crazy behind the scenes to provide more clarity,” says Saskia Middeldorp, a vascular internist at Radboud University Medical Center in the Netherlands.

So far, most of the clotting cases have been observed in women under age 65. That may be because many European countries used the shot only in younger, prioritized populations, such as health care workers and teachers, a majority of whom are women. They initially hesitated to give it to people older than 65 because the company's early clinical trial data included few in that group.

The United Kingdom remains a puzzle. Despite administering more than 11 million AstraZeneca doses, it has so far reported only a handful of suspicious clotting cases.

But the U.K. did not limit the vaccine to younger groups, so the average age of recipients there may be older. AstraZeneca had not commented on the clotting cases as *Science* went to press, except to say the rare set of symptoms did not appear in the company's vaccine trials.

Researchers in Germany have proposed that some component of the vaccine triggers a rare immune reaction like one occasionally seen with the blood thinner heparin, in which antibodies trigger platelets to form dangerous clots throughout the body. This week the team posted case descriptions of what they call vaccine-induced prothrombotic immune thrombocytopenia (VIPIT) on the preprint server Research Square. The team, led by Andreas Greinacher at the University of Greifswald, also recommends a way to test for the disorder and a treatment, which they say should help ease worries about the vaccine.

Even if VIPIT isn't the whole story, multiple other researchers told *Science* they are now convinced the vaccine somehow causes the rare set of symptoms. If true, that could be a serious blow to a vaccine that is central to the World Health Organization's push to immunize the world. AstraZeneca is working with partners around the globe to make and distribute billions of doses in low- and middle-income countries.

Discussion of this possible side effect is likely to stoke short-term vaccine hesitancy, says Michael Bang Petersen, a political scientist at Aarhus University in Denmark. He stresses, however, that clear, transparent communication about possible risks is crucial for maintaining public trust in all COVID-19 vaccines. “It is very important that we do not lose the war because we are too eager to win the short-term battle,” he says. ■

ASTRONOMY

Underwater neutrino traps take shape

Detectors aim to trace high-energy particles to violent cosmic accelerators in deep space

By Daniel Clery

Since 2010, IceCube, a detector frozen in the ice beneath the South Pole, has snared neutrinos from deep space. The universe is awash with these fleeting, almost massless particles, but IceCube is after a rare subset. They are messengers from distant cosmic accelerators such as supernovae, neutron stars, and black holes. IceCube has caught about 300 in its cubic kilometer of ice, but has had less success tracing them to their probable source—just two so far. Now, it is poised to get help from new detectors that trade Antarctic ice for deep northern waters.

This month, researchers will begin to drop sensor strings into the Mediterranean Sea off the coast of Sicily, as they embark on building the Cubic Kilometre Neutrino Telescope (KM3NeT). Meanwhile, a Russian team has been working on the frozen surface of Lake Baikal in Siberia, the world's deepest lake, to drop detector strings into its depths. The Gigaton Volume Detector (Baikal-GVD) is already half complete and taking data. A third effort, the Pacific Ocean Neutrino Explorer (P-ONE) hopes to deploy one or more prototype strings off the west coast of Canada next year.

“We're really looking forward to having a worldwide network,” says Olga Botner, an astroparticle physicist at Uppsala University and IceCube team member. “With three detectors we'll get more neutrinos and more likelihood of identifying sources.”

Trillions of neutrinos stream unnoticed through your body every second, most of them low-energy neutrinos from local sources like the Sun. IceCube and the other “neutrino telescopes” study the rare high-energy neutrinos produced when charged particles—cosmic rays—accelerated to ultrahigh energies in the distant universe

smash through a cloud of gas. The cosmic rays can also reach Earth, but can't easily be traced back to their source because they follow a twisting journey through the universe's magnetic fields. Chargeless neutrinos offer a truer flight path that reveals their source. But only if researchers can catch them.

Very occasionally a passing neutrino will collide with an atomic nucleus, spawning other particles. In water or ice, those particles emit a flash of light as they slow down. IceCube contains more than 5000 light detectors watching the deep, transparent ice to pin down the timing and brightness of the flash, from which researchers can reconstruct the neutrino's energy and path.

IceCube catches about 30 high-energy neutrinos per year that are presumed to be extragalactic. That's about the number expected to come from supernovae in starburst galaxies—young galaxies that forge huge, fast-burning stars tens of times faster than the Milky Way. When these stars die and explode, they are thought to fling out cosmic rays that produce neutrinos when they crash through dense clouds of star-forming gas near the supernovae.

The rate at which IceCube detects extragalactic neutrinos is “a strong hint that these are the sources,” says theorist Eli

Waxman of the Weizmann Institute of Science. Yet so far, the two neutrinos to be traced back to likely sources seem to have come from supermassive black holes (SMBHs) in galactic cores, not starbursts. One seemed to come from a blazar, a jet from a SMBH pointing at Earth, and another, announced earlier this year, from a tidal disruption event—an SMBH tearing

“We want to see the parts of the universe that cannot be seen with photons.”

Olga Bottner, Uppsala University

apart a star (*Science*, 26 February, p. 872). To resolve the issue, Waxman says, researchers need bigger detectors and better pointing. “With this next generation we will identify individual starburst galaxies,” he says.

Constructing IceCube took 5 years of drilling into the Antarctic ice cap with hot-water jets. Building a detector deep underwater has its own challenges. Each KM3NeT string, studded with detectors 40 meters apart, is dropped from a ship as a ball and unspools as it sinks to the floor of the Mediterranean 3.4 kilometers down.

Buoys keep the strings upright, while a remotely operated submersible anchors them and connects them into power and communication networks. The team is preparing to install 18 strings by September. “It's a major step forward,” says spokesperson Paschal Coyle of the Center for Particle Physics of Marseille. The aim is to have 230 strings and more than 4000 light detectors in place by 2026 to make a detector slightly larger than IceCube.

Baikal-GVD researchers have an easier job. For now, they can safely drive onto the frozen lake, erect winches, and lower strings into the water. Working on the ice “really makes it easier and cheaper to deploy things,” says Dmitry Zaborov of the Russian Academy of Sciences's Institute for Nuclear Research. The team has installed 56 strings so far and is aiming for another 40 by 2024, to cover a volume about 70% the size of IceCube.

Using water instead of ice will give the new detectors an edge. Light scatters less in water, so particle tracks can be mapped more precisely, giving a sharper view of the neutrinos' origin. KM3NeT estimates it can achieve a top angular resolution of less than 0.1°, compared with IceCube's 0.5°, which is about the size of the full Moon.

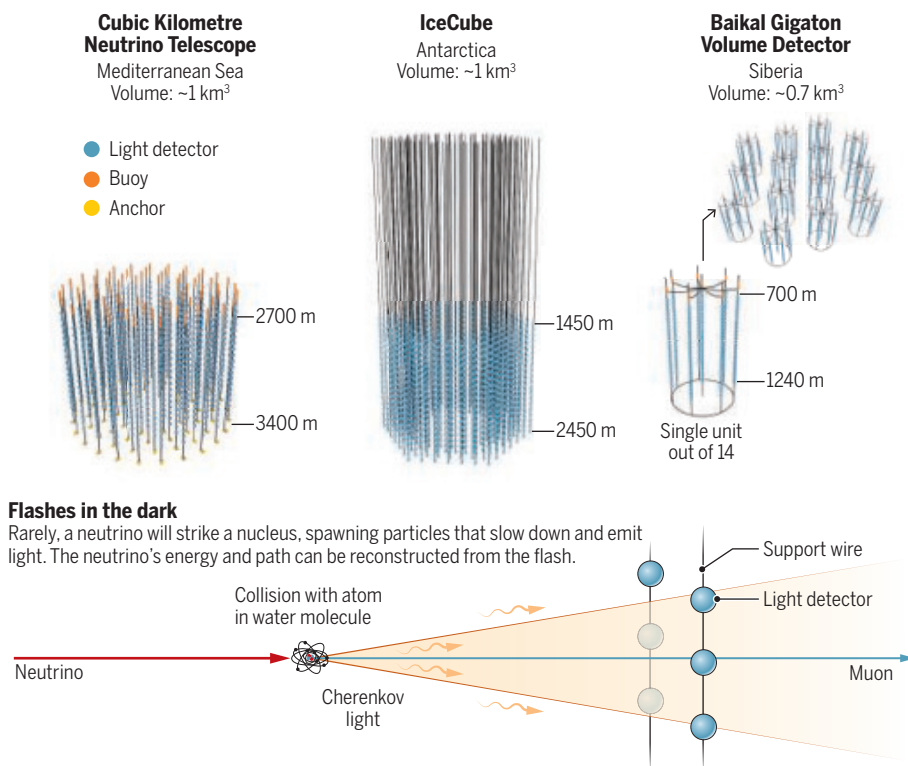
The telescopes' location in the Northern Hemisphere is also a plus. Neutrino detectors look down rather than up, watching for neutrinos that have passed through Earth, which acts as a shield against many background particles. As a result, IceCube's view takes in the northern sky. The northern detectors, in contrast, will look south, into the heart of the Milky Way, the most likely home for neutrino sources such as magnetized neutron stars, the galaxy's SMBH, or, if astronomers are lucky, a new supernova.

Later in the decade they could be joined by P-ONE, which is taking advantage of a network of existing power and data cables installed for oceanographic experiments off the coast of British Columbia. “It's a plug-and-play operation,” says team leader Elisa Resconi of the Technical University of Munich. With three widely spaced telescopes in the north, “we'll see nearly the entire sky all the time,” Resconi says. “It will bring the field to a new level.”

The ultimate aim, once researchers can link neutrinos of particular energies to different types of sources, is to do true neutrino astronomy: viewing the universe not with photons, but with neutrinos, which bear news about violent corners of the universe otherwise hidden from view. As Bottner puts it: “We want to see the parts of the universe that cannot be seen with photons.” ■

Traps of water and ice

Neutrino telescopes need huge detection volumes to catch the elusive particles. Two underwater detectors should offer better pointing than IceCube, enabling astronomers to trace the particles back to their origins.





Visitors gawk at Chang'e-5 lunar samples on display at the National Museum of China in Beijing.

SCIENCE DIPLOMACY

China's scientific treasures tempt foreign collaborators

Outside researchers are offered telescope time and access to Moon rocks, but some are uneasy about the relationship

By **Dennis Normile**

For a generation, China played scientific catch-up to more advanced nations, but the tables are turning. China has the world's largest radio telescope and the first Moon rocks in 45 years. Now, it is offering foreign researchers access to those scientific treasures. Many are eager, but others are uneasy about what they see as collaborating with an authoritarian regime.

In December 2020, the Chang'e-5 mission returned 1.7 kilograms of rock and soil from the Moon—the first lunar samples since 1976, and a chance for researchers to obtain dates that could help unravel Solar System history (*Science*, 20 November 2020, p. 900). On 18 January, the China National Space Administration (CNSA) confirmed it would encourage “joint international research” on the samples, and it may begin to review applications this month.

Also opening up is the Five-hundred-

meter Aperture Spherical radio Telescope (FAST), the world's most sensitive single-dish radio telescope since its completion in 2016. After several years of limited observations by domestically led teams, the Chinese Academy of Sciences's National Astronomical Observatories (NAOC), FAST's operator, will this month start to accept proposals from foreign principal investigators. FAST Chief Scientist Li Di expects tens of applications for the roughly 400 hours of foreign observing time. “It will be severely oversubscribed, so it will be a competitive process,” Li says.

NAOC Director General Chang Jin says a major objective in sharing the resources is simply to do the best science. Getting foreign ideas about how to use FAST “is definitely beneficial to advancing research in radio astronomy,” he says. Generosity is also seen as befitting a space power. “China has benefited a lot from international space cooperation; it's natural for China to give back to the world when it can,” says Zhang Ming,

a space policy expert at the Shanghai Academy of Social Sciences.

David Burbach, a security and space policy expert at the U.S. Naval War College, says China's science diplomacy “can promote domestic legitimacy [and] project a global image of being a cooperative and nonthreatening power.” But some see less benign motives. “The Chinese government is always looking for opportunities to convert scientific collaboration into political advantage,” says Clive Hamilton, an ethicist at Charles Sturt University, Canberra. For scientists this sets “an ethical trap of lending legitimacy” to an authoritarian regime, he says.

Some researchers agree. “Even if FAST was the perfect instrument to pursue my work, I would not be willing to work in China in a way that contributed to Chinese prestige,” says Joanna Rankin, a radio astronomer at the University of Vermont. She points to human rights concerns and the erosion of freedoms in Hong Kong.

For others, working with China is an exercise in scientific diplomacy, in the same spirit as U.S.-Soviet scientific collaborations of decades past. “In my opinion, working with China on scientific matters does not imply condoning its political practices,” says Sun Kwok, a Hong Kong-born astronomer and former dean of science at the University of Hong Kong now at the University of British Columbia, Vancouver, who previously participated in the Chang'e program. “Such interactions certainly contributed positively during the Cold War,” says Carl Heiles, a radio astronomer at the University of California, Berkeley. Invoking a hard line on cooperation would isolate China and reinforce disagreements, says Heiles, who is already on a FAST team observing the interstellar medium.

Legal and diplomatic obstacles may get in the way for U.S. researchers. Since 2011, Congress has barred NASA from using its funding for any bilateral activities “with China or any Chinese-owned company.” The language, originally added because of concerns over human rights and to protect advanced space technologies (*Science*, 29 April 2011, p. 521), could prevent U.S. lunar researchers from using NASA funds to study the samples.

China sees it as an impediment as well. Whether China will share lunar samples with U.S. scientists “depends on the policy of the U.S. government,” Wu Yanhua, CNSA deputy director, said at a 17 December 2020 press briefing. Bradley Jolliff, a planetary scientist at Washington University in St.

Louis, is frustrated but understands China's stance. "We cannot loan Apollo samples to the Chinese; why should they loan Chang'e samples to U.S. scientists?" he asks.

An international consortium might "break down [the barriers] that the politicians have put in place," says Clive Neal, a lunar scientist at the University of Notre Dame who is in the early stages of developing a multilateral approach. Another budding effort is the International Lunar and Planetary Research Center, under the Chinese Academy of Geological Sciences's Institute of Geology, which is studying the possibility of arranging international visits to the laboratories holding samples, says Alexander Nemchin, a geologist at Curtin University and a co-chair of the group.

Scientists seeking to use FAST face fewer hurdles. "Practically anybody can put in a request," Li says. An English-language application template has been posted on the FAST website that solicits proposals for observations up to 100 hours long. International referees will review and rank the proposals, and telescope time will be allocated by August 2021.

The untimely demise of the Arecibo Observatory in Puerto Rico (*Science*, 15 January, p. 225), previously the world's largest single radio dish, adds to the allure of FAST. It won't replace all of Arecibo's capabilities: It covers a narrower range of frequencies, and lacks the active radar system that Arecibo used to map the surfaces of planets and asteroids. But with twice Arecibo's sensitivity, FAST is discovering faint and unusual pulsars and fast radio bursts. Li also hopes FAST will help fill Arecibo's shoes in the International Pulsar Timing Array, a network of telescopes seeking to detect gravitational waves by looking for tiny timing variations in signals from fast-spinning pulsars.

For foreign researchers, the opportunities are just beginning. This month or next, CNSA is expected to launch the core module of China's space station, and within the next few years it will add two modules for experiments in microgravity, physics, and space weathering that will be open to international researchers. Around 2024, China is planning to launch an orbiting telescope with a 2-meter mirror—slightly smaller than the Hubble Space Telescope's—that will be able to dock with the station for servicing. On Earth, the Chinese Academy of Sciences's Institute of High Energy Physics is planning a \$5 billion particle accelerator that would dwarf the world's top facility, CERN's Large Hadron Collider.

"China is planning to implement many other big space exploration and science projects," Zhang says. The dilemmas will multiply along with the opportunities. ■

SYNTHETIC BIOLOGY

Researchers coax stripped-down cells to grow normally

Restoring seven removed genes rescues normal division

By **Mitch Leslie**

Five years ago, researchers announced to great fanfare that they had engineered a stripped-down microbial cell able to survive with fewer genes than any known organism. But that "minimal cell" often divides abnormally. Now, by putting back only seven genes, a team has corrected the cells so they grow like the natural versions.

The discovery could sharpen scientists' understanding of which functions are crucial for normal cells and what these still-mysterious genes are doing, says synthetic biologist Kate Adamala of the University of Minnesota, Twin Cities. "This is a significant step forward that maybe can help identify the functions of these unknown genes."

Pinning down essential genes could also benefit efforts to craft cells or cell-like objects that could produce chemicals, sense environmental conditions, deliver drugs, and perform other tasks in industry and medicine. "We need to know what is the minimal parts list we need to put together to restore life," says microbiologist Anthony Vecchiarelli of the University of Michigan, Ann Arbor. Minimal cells could also provide insight into the origin of life by illuminating which capabilities were essential for primordial cells.

Genome sequencing pioneer J. Craig Venter of the J. Craig Venter Institute (JCVI) and colleagues created the first minimal cells. They started with *Mycoplasma* microbes, parasites that are already pretty minimal—one variety gets by with 525 genes, compared with the roughly 4000 of the common intestinal bacterium *Escherichia coli*. In 2010, the team reported that replacing the 985-gene genome of one type of *Mycoplasma* with a 901-gene synthetic genome kept the cell, dubbed syn1.0, purring. The scientists continued to remove chunks of DNA from syn1.0's genome, and in 2016, they unveiled an even sparer version, known as syn3.0. It could metabolize and reproduce with a measly 473 genes.

But this cell also has a quirk: Many of its

progeny are misshapen. To check whether lab conditions might be stressing the delicate synthetic cells, a group led by synthetic biologist Elizabeth Strychalski of the National Institute of Standards and Technology cosseted the cells in chambers on microfluidic chips. These deluxe quarters shielded the cells from currents in the nutrient medium that might harm them and allowed the researchers to watch as they divided.

This gentle treatment didn't help, however. "When we looked at the individual cell level, it was absolute mayhem," says Strychalski, who worked with colleagues from JCVI and three universities. The cells should have been small orbs, but some were behemoths about 25 times the normal girth. Others looked like

threads or strings of pearls. Rough handling wasn't the problem, the researchers concluded; instead, wonky division stemmed from the removal of genes that help control reproduction and cell shape.

It wasn't obvious which missing genes were to blame, but a clue was sitting in a lab freezer. To create syn3.0, Venter and colleagues had generated strains of cells that lacked portions of syn1.0's genome. When Strychalski and her team thawed one of these strains,

which was missing 76 of syn1.0's genes, it also produced abnormally shaped progeny. "It helped us narrow the genes from 400 to 76," says co-author James Pelletier, a biophysicist at the Massachusetts Institute of Technology.

By adding back combinations of genes to see whether the resulting cells divided normally, the researchers shrank the number required to 19 and then even further. This week in *Cell*, they report they could restore normal division by adding just seven genes to syn3.0.

Two of the genes were already known to play a role in cell division, but the involvement of the other five came as a surprise—and their roles in cleaving the microbes remain unknown. The corrected minimal cells could help elucidate this still-mysterious process, Strychalski says: "We still don't know the mechanism by which these things divide. That blows my mind—it's one of the basic aspects of life." ■

"We need to know what is the minimal parts list we need to put together to restore life."

Anthony Vecchiarelli,
University of
Michigan, Ann Arbor



Artificially veiling the Sun might reverse global warming, but carries risks.

CLIMATE

U.S. needs solar geoengineering research program, report says

Proposed \$200 million effort could include limited field studies of ways to dim sunlight

By **Paul Voosen**

An influential panel of scientists has recommended the United States pursue a robust research program into a controversial technological bandage for climate change. Solar geoengineering—deliberately altering the atmosphere to reflect more sunlight—might forestall some of the worst effects of global warming, but fears of tinkering with climate systems and the technology's potential for misuse have slowed research.

Unabated global warming is changing the calculus, however, and a new report from the National Academies of Sciences, Engineering, and Medicine (NASEM) urges the government to fund a cautious but expansive effort that could total \$200 million over 5 years.

The panel's recommendation is “thrilling,” says David Keith, a Harvard University energy and climate scientist and longtime proponent of geoengineering research. The report sets multiple conditions: Any experiments would require substantial oversight, risk assessment, and public outreach. Field tests should only move forward if they provide observations that could not be made in the lab or by observing sunlight-altering natural events, such as volcanic eruptions. And, says Chris Field, a climate scientist at Stanford University and chair of the committee, “Learning more about these technologies shouldn't be seen as a step toward deployment.”

The 25 March report also makes clear that no research should occur without

strong government commitments to reducing emissions. Reflecting sunlight without curbing carbon dioxide emissions would do nothing to slow the acidification of the oceans, for example. Solar geoengineering “only makes sense in the context of a primary societal and government focus on mitigation and adaptation,” says Peter Frumhoff, chief climate scientist at the Union of Concerned Scientists and co-author of the report. “It makes no sense otherwise.”

In preparation since 2019, the report takes a close look at three proposed solar geoengineering strategies: stratospheric aerosol injection, which would release long-lived reflective particles into the upper atmosphere; marine cloud brightening, which seeks to thicken low-lying clouds over the ocean; and cirrus cloud thinning, which would alter wispy high-altitude ice clouds, allowing more infrared radiation to escape to space. Each has its own risks and uncertainties: Particles released into the stratosphere, home of the ozone layer, could have long-lasting, global effects. Cloud thickening and thinning, though taking place on more regional scales, would change cloud properties with unpredictable results.

The report recommends that the U.S. Global Change Research Program, which coordinates federal climate research from the White House, lead the effort and establish a standing advisory body that would include scientists, policymakers, and representatives from civil society. Any research agenda should include studying the interplay between solar geoengineering and public perceptions, poli-

tics, and socioeconomic trends, along with its ethical dimensions, the panel wrote. And experiments should be designed to explore risks as well as feasibility, Frumhoff says. “They should potentially have outcomes that could say, ‘This is not a good idea.’”

That aligns with “how we hope to see this research evolve in the future,” says Sarah Doherty, an atmospheric scientist at the University of Washington, Seattle. Doherty is program director of the Marine Cloud Brightening Project, which plans within 1 year to test a nozzle that will spray sea salt particles to thicken ocean clouds, making them more reflective.

The \$200 million budget suggested by NASEM would be enough to support several aircraft campaigns, which might include the deliberate release of particles into the atmosphere. But the panel says any such releases should be minute, 100 times smaller than the smallest amount that could theoretically influence the environment or global temperature.

Current U.S. research into solar geoengineering is fractured and ad hoc. The last two federal spending bills provided \$13 million to the National Oceanic and Atmospheric Administration to pursue such studies—the first such funding. The agency is planning to fly a lunch box-size spectrometer into the stratosphere by balloon to capture a high-resolution view of long-lived, light-reflecting natural particles, with the first flight scheduled later this year. The rest of the money has, so far, gone to improving instrumentation, modeling the stratosphere, and studying how fine particles influence marine clouds, a long-standing research question.

Without a strong U.S. government role, philanthropy has filled the void. Bill Gates has supported a project co-led by Keith: the Stratospheric Controlled Perturbation Experiment, which aims to release 2 kilograms of light-reflecting chalk in the upper atmosphere and observe the resulting plume. A proposal to test its experimental balloon rig above Sweden this year, without releasing particles, is pending before an independent review board, after a delay following an outcry from Swedish environmental groups. Meanwhile, last year SilverLining, a nongovernmental organization dedicated to supporting solar geoengineering research, announced \$3 million in support to five research labs, largely focused on modeling.

As pleased as Keith is with the NASEM report, he notes that other bodies have issued similar, though less ambitious, suggestions stretching back to the early 1990s. “The big question,” Keith says, “is will it happen?” ■

A PLATEFUL OF MEDICINE

Special diets might boost the power of drugs to vanquish tumors



When New York City medical oncologist Vicky Makker meets a patient with endometrial cancer that has spread or recurred, she knows the outlook isn't good. Even after radiation and drug treatments, most women with advanced disease die within 5 years.

But this spring, Makker is helping launch two clinical trials she hopes will change the picture. The drug patients will receive, called a phosphatidylinositol 3-kinase (PI3K) inhibitor, has already failed in multiple can-

By **Jocelyn Kaiser**

cer trials. But the new studies are taking an unconventional tack to resurrect the drug: putting patients on a ketogenic diet, a low-carbohydrate regimen that typically involves loads of meat, cheese, eggs, and vegetables. The researchers hope the diet will render tumors more vulnerable to the drug, which blocks a growth-promoting pathway in cells. "It's very outside of the mainstream thinking," says Makker, a researcher at Memorial Sloan Kettering Cancer Center.

The trials are the brainchild of cell metabo-

lism researcher Lewis Cantley of Weill Cornell Medicine (WCM). Decades ago, he discovered the PI3K signaling pathway, which the drugs aim to target. More recently, his lab showed in mice that a ketogenic diet can counter tumors' resistance to those drugs.

Cantley isn't the first to suggest that a particular diet, such as fasting or selectively reducing certain nutrients, can make cancer treatments work better. For at least a century, doctors and self-styled nutrition experts have touted the idea in bestselling books and, more recently, on popular websites. "There's a big industry there, but it's not based on a



real understanding of what's going on in a tumor cell," says cancer biologist Karen Vousden of the Francis Crick Institute in London. Still, some early clinical trials showed hints of an effect. Now, studies from high-profile labs are spawning a new wave of trials with more rigorous underpinnings.

Scientists including Vousden, who co-founded a company with Cantley to test diet-drug combinations in cancer trials, are unraveling the molecular pathways by which slashing calories or removing a dietary component can bolster the effects of drugs. In mice with cancer, "the effects are often-

times on the same order of magnitude as those from the drugs that we give patients. That's a powerful thing to think about," says physician-scientist Matthew Vander Heiden of the Koch Institute for Integrative Cancer Research at the Massachusetts Institute of Technology (MIT) and the Dana-Farber Cancer Institute. And the idea appeals to patients, he adds. "Diet is something that people feel like they can control."

STILL, COMPELLING RESULTS in patients will be needed to overcome some oncologists' view of special diets as fringy alternative medicine. The doubts often focus on a pioneer in the field, biochemist Valter Longo of the University of Southern California and the Italian Foundation for Cancer Research's Institute of Molecular Oncology, who has built a huge popular following with his fasting research. Critics worry the media attention encourages cancer patients to diet without adequate evidence. Longo agrees patients should not improvise and says fasting needs more clinical testing.

His labs in Los Angeles and Milan are full of hungry mice. Longo began his career studying caloric restriction, which can extend the life spans of diverse species and has been shown to reduce the incidence of cancer in rodents and monkeys. Because few people can stay on low-calorie diets in the long term, Longo shifted his focus to fasting, a treatment offered for various ailments as far back as ancient Greece. In two key papers in 2008 and 2012, his team reported that reducing nutrients in the medium used to grow cells in a dish protected normal cells from the toxic effects of chemotherapy drugs such as cyclophosphamide and doxorubicin, yet made cancer cells more likely to die. In mice with cancer, fasting—drinking only water for 2 or 3 days—helped the drugs curb tumor growth and boosted the animals' survival.

Longo's explanation is that fasting, which lowers levels of glucose in the blood, causes healthy cells to hunker down in a protective mode. But cancer cells need to keep growing, which puts them at risk of starvation. Fasting also reduces the body's production of hormones, such as insulin, that can drive tumor growth. Both effects may make the cancer cells more susceptible to chemotherapy.

Hoping to make fasting easier on cancer patients, Longo's team showed that merely limiting calories for a few days has similar effects on blood hormones and other biomarkers. A company Longo started in 2009, L-Nutra, supplies that "fasting-mimicking" diet for clinical trials: packets of crackers, soups, teas, and nut bars. The company also sells the meal kits online to the public, touting them as a way to combat aging. After fac-

ing criticism for profiting off a product that hasn't been fully validated, Longo began to donate profits from his shares in the company to charity in 2017.

Animal studies supporting the benefits of fasting-mimicking diets in cancer are now plentiful. Last year, Longo's team reported that restricting calories enhanced the effects of hormone therapies in mice with breast cancer. Another team reported a synergistic effect with immunotherapies, again in mice. "There are probably 100 papers on this, almost all positive," Longo says.

In small, preliminary clinical trials, Longo's team and others showed that the fasting-mimicking diet may protect against some side effects of chemotherapy. Carolina Sandoval of Pasadena, California, age 40, participated in one of Longo's intermittent fasting trials for 2 months last fall during four cycles of chemotherapy to fight breast cancer. "It was really hard," she says, especially when the chemo affected her taste buds. "I couldn't bear the taste of some of the food," and she lost weight, she says.

But Sandoval also says she thinks the diet spared her from some of the nausea and fatigue of chemotherapy and allowed her to avoid taking days off from her job teaching high school online. She hopes the fasting "put my good cells to sleep, and the chemotherapy was able to attack more of my cancer cells," she says. "I would do it again."

However, the first trial aiming to rigorously test whether a fasting-mimicking diet can make chemotherapy work better faltered, partly because participants found the diet unappetizing. The study, launched in 2014 and led by oncologist Judith Kroep of Leiden University Medical Center, monitored 131 Dutch women with early stage breast cancer who were slated to receive chemotherapy before surgery. They were randomly assigned to follow either Longo's fasting-mimicking diet or a regular diet for 4 days leading up to each round of chemo. But many women disliked the taste and lack of choices in the diet, and just 20% completed all eight cycles. In part because of the dropout rate, Kroep's team couldn't go on to compare biomarkers predicting overall survival in the two groups.

Yet the trial did yield hints that chemotherapy was more potent and less toxic to healthy cells in women who completed at least two cycles of the fasting-mimicking diet. Scans showed their tumors were more likely to shrink, and immune cells in their blood had less DNA damage from chemotherapy, Kroep, Longo, and colleagues reported last year in *Nature Communications*. Longo calls that "remarkable evidence" that the diet worked. But without definitive evidence that dieters were more likely to survive longer, other researchers found the trial inconclusive.

And when Kroep and a patient from the trial were featured on a Dutch TV program in late 2019, an uproar ensued from physicians and dietitians worried women with cancer would fast on their own. Medical groups and Kroep's own institute released cautionary statements. "I agree that confirmation is needed before we can advise patients to fast, also because it is not always easy," says Kroep, who is planning a new trial with changes to make the diet more appealing.

Longo hopes for a bigger test. His team has applied to the U.S. National Cancer Institute for a \$12 million grant to run a 460-patient clinical trial at 11 hospitals of a fasting-mimicking diet and chemotherapy for breast cancer. The agency includes whether fasting works on a list of "provocative questions" in cancer. "If it happens, it's going to be very exciting," Longo says.

PATIENTS WHO BALK at cutting calories may have an easier time with the popular low-carb ketogenic diet that, like fasting, lowers glucose and hormone levels in blood. "Generally speaking, the ketogenic diet and fasting are two roads to a similar metabolic state," says Princeton University biochemist Joshua Rabinowitz. A ketogenic diet also forces the liver to turn excess fat into molecules called ketone bodies that glucose-craving cancer cells struggle to burn for energy, some scientists suggest. The approach has been used to treat epilepsy since the 1920s, when researchers discovered the seizure-reducing effects of the ketogenic diet on brain metabolism.

Animal studies going back a decade suggest a ketogenic diet can enhance the effects of chemotherapy and radiation. Case reports and some small clinical trials hint that the diet may extend the lives of cancer patients—particularly those with the brain cancer glioblastoma, which tends to use large amounts of glucose.

Cantley turned to a ketogenic diet because of a major disappointment: Drugs based on his discovery of the PI3K pathway that drives growth in many tumors largely flopped in trials in the 2010s. Except for blood and breast cancer studies that led to drug approvals, the trials were a disaster, Cantley says. "Billions of dollars went into efforts that failed."

He thinks he knows why. The drugs cause a side effect—a rise in blood sugar—that doctors often treat with insulin. But insulin stimulates the PI3K pathway in tumors and cancels out the cancer drug's effects. In a study published in 2018, Cantley's team fed a ketogenic diet designed to lower the body's natural insulin production to cancer-afflicted mice receiving a

Matching meals to meds

Studies in mice have found that removing calories or specific nutrients from the diet can boost the power of cancer drugs. The diet-drug pairings vary with tumor type and genetics.

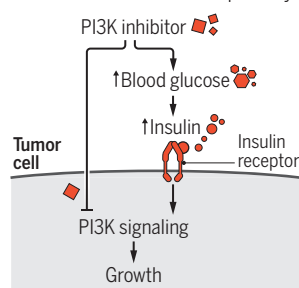


Endometrial cancer

In some tumors, mutations in the genes *PIK3CA* or *PTEN* ramp up PI3K signaling, a key growth pathway.

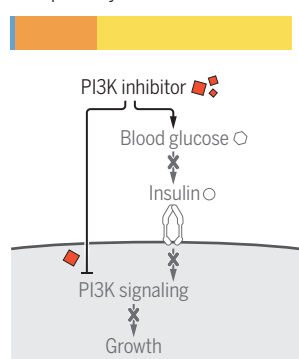
PI3K inhibitor treatment

The drug stops working because it raises blood glucose, boosting insulin levels to reactivate the PI3K pathway.



Plus ketogenic diet

Restricting carbohydrates lowers insulin and glucose levels so the PI3K pathway remains blocked.

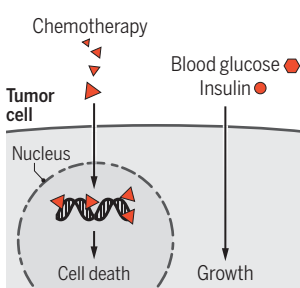


Breast cancer

When tumors lack changes in the gene *HER2*, doctors rely on chemotherapy to shrink them before surgery.

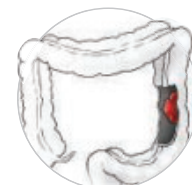
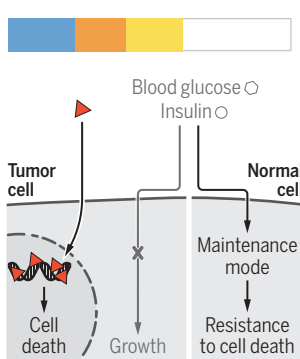
Chemotherapy treatment

Drug kills tumor cells by damaging DNA and blocking replication, but nutrients enable some cells to keep growing.



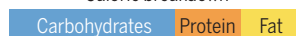
Plus fasting-mimicking diet

Normal cells survive the drop in glucose, while tumor cells become more vulnerable.



Normal diet

Caloric breakdown

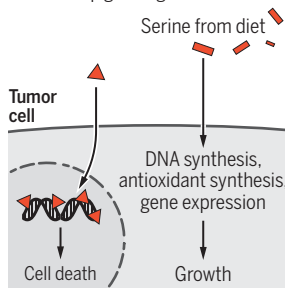


Colon cancer

Tumors missing the *P53* gene can't easily make their own supply of the amino acid serine.

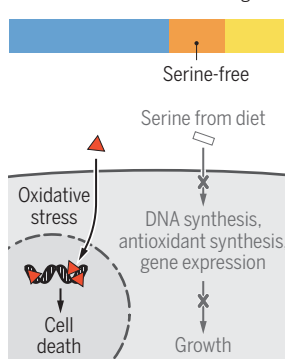
Chemotherapy treatment

Drug kills tumor cells by damaging DNA and blocking replication, but nutrients enable some cells to keep growing.



Plus serine-free diet

Serine deficiency hinders cancer cell growth and causes oxidative stress that boosts DNA damage.



PI3K drug. The researchers found that the diet allowed the drug to keep working and curbed tumor growth.

The two trials Makker is co-leading will soon test whether that hypothesis holds up in people with endometrial and some other cancers bearing a mutation in one of two genes that rev up the PI3K pathway. In one trial, participants will shop for and prepare meals according to instructions. In the other, the company Faeth Therapeutics that Cantley co-founded will ship meals to patients to help them stay on track.

Memorial Sloan Kettering and WCM are

already running a feasibility test, supplying 4 weeks of packaged meals to about 30 women with endometrial cancer awaiting surgery. "They like the food, which is really great to see," Makker says. And blood tests showed a drop in insulin and other changes that "mirrored the mouse model," says WCM endocrinologist Marcus Goncalves, a co-investigator for the study and the two new trials.

If those trials show the ketogenic diet helps curb tumor growth for a year or two longer than the PI3K inhibitor otherwise would, the diet "could become the standard

of care,” Cantley says. “That will be what physicians will tell patients to do.”

A ketogenic diet may enhance other cancer treatments, too. Immunologist Laurence Zitvogel of the Gustave Roussy Institute in France recently studied mice with skin, kidney, or lung cancers receiving a drug known as a checkpoint inhibitor that helps the immune system’s T cells kill tumors. In animals on a ketogenic diet, the ketone bodies they produced boosted the T cells’ power; her team reported in January. Rabinowitz and collaborators have begun enrollment for a 40-person trial to see whether the diet can enhance the impact on pancreatic cancer of a chemotherapy cocktail.

But researchers warn that a ketogenic diet could backfire and fuel the growth of fat-loving cancers such as those of the breast and prostate and others with certain mutations. Cantley found the diet spurred tumor growth in mice with leukemia. In a recent study, researchers found that, contrary to prevailing thought, glioblastoma tumors can get around the dearth of glucose by feeding on ketone bodies. To safely harness a ketogenic diet as treatment, “you need to really understand how and where it works,” Vander Heiden says.

OTHER RESEARCHERS ARE exploring an even more precise dietary limitation: cutting out specific amino acids, best known as the building blocks of proteins but also key to many other metabolic processes. Vousden unexpectedly veered into that line of research while studying a cancer-preventing gene called *p53*. The protein it encodes can trigger cells that have DNA damage to self-destruct, stopping them from turning cancerous. The gene is mutated in many tumors, allowing unrestrained growth.

But in 2005, a U.S. lab reported a surprising finding: The intact *p53* protein helps healthy cells survive when glucose is scarce, suggesting *p53*-mutated cancer cells are especially vulnerable to glucose limitation. Vousden wondered whether the protein also helps cells survive a shortage of other, less explored nutrients, such as amino acids—and whether the mutated *p53* in cancer would make the cells less resilient.

To find out, her postdoc Oliver Maddocks methodically removed various amino acids from cancer cells’ culture medium. Many types of cancer cells grew more slowly when deprived of two related amino acids, serine and glycine, and deleting *p53* ramped up that effect. The scientists then tested the effects of a serine- and glycine-free diet in mice. To some colleagues, doing so seemed “kind of a pointless experiment,” Maddocks says, because the molecules are nonessential amino acids, meaning the body can make

them even if they are absent from the diet.

But the experiment wasn’t pointless. Several tumor types, it turned out, could not efficiently make serine on their own. Maddocks and Vousden reported in 2013 and 2017 that the special diet slowed cancer growth and extended the lives of mice implanted with colon cancer cells lacking *p53* as well as in mice engineered to develop lymphoma or colon tumors. Cells need serine or glycine to make a compound that sops up DNA-damaging free radicals, and the deprivation made tumor cells more sensitive to that oxidative stress. Radiation and some

In mice, “effects are oftentimes on the same order [as] the drugs that we give patients. That’s a powerful thing to think about.”

Matthew Vander Heiden,

Koch Institute for Integrative Cancer Research

chemotherapies kill cells by generating free radicals, so the results suggested the diet could prime tumors for those treatments.

Similar findings have emerged for other amino acids. Limiting the essential amino acid methionine appears to amplify the effects of radiation and chemotherapy in mice with colon cancer and sarcomas. And removing asparagine, an amino acid abundant in asparagus, from mouse diets curbed the spread of metastatic breast cancer, suggesting the diet could enhance drug treatments.

As with serine, depriving mice of one of those amino acids apparently disrupts metabolic cycles by which cancer cells respond to oxidative stress, synthesize DNA, and turn genes off and on. The idea isn’t totally new. Cancer drugs known as antifolates developed in the 1940s also “broadly affect pathways that are sensitive to the levels of nutrients in our diet,” says cancer biologist David Sabatini of MIT’s Whitehead Institute.

Removing a specific amino acid from a person’s diet won’t be easy. Patients will have to eliminate all protein-rich foods, including grains, meat, and beans, and drink a specially formulated shake lacking the relevant amino acid. But that approach is not unprecedented: People born with the metabolic disorder phenylketonuria, for whom the amino acid phenylalanine is a neurotoxin, stay healthy by getting amino acids from a phenylalanine-free drink.

The company Vousden and Cantley founded, Faeth (Welsh for nutrition), is gearing up to test amino acid-depleted diets in two clinical trials this year. Faeth, also co-

founded by Maddocks, will combine chemotherapy with a shake lacking specific amino acids, delivered to participants’ homes alongside other meal components, such as salads. The researchers got support from private investors after failing to win research grants for their idea, says Maddocks, now at the University of Glasgow. “It’s quite out of the box.”

LOOMING OVER ALL those trials is the question of whether patients will be able to stick to the diets. Many people struggle with the high fat content of a strict ketogenic diet, for example. “It’s like eating butter” all the time, Vander Heiden says. Some very sick patients may be too underweight to follow any diet that restricts their caloric intake. And people getting treatments such as PI3K inhibitors or hormone therapy for months or years may have to keep up a special diet for just as long—a daunting prospect.

If a diet turns out to work long term, the cost of delivering meals to patients could add up. But Maddocks notes that meals should still be cheaper than many cancer drugs.

A simpler alternative to diet changes might be drugs that have some of the same effects—for instance, an enzyme that blocks cells from synthesizing a specific amino acid. In Zitvogel’s mouse study, adding ketone bodies to the animals’ chow worked as well as a ketogenic diet at boosting immunotherapies; she now plans to compare the two in a clinical trial.

Another concern is that tumors may harbor some cells that resist the effects of special diets. Longo argues that’s unlikely with a fasting-mimicking diet because “it takes away so many things from the cancer that some of them are bound to be important.” But work led by Vousden and Maddocks showed that some tumors with a mutation in a gene called *KRAS* can get around a serine-free diet by ramping up their own serine synthesis. And a serine-free diet would have little effect on breast cancer cells that spread to the pancreas, which is naturally awash in that nutrient, Vander Heiden’s group reported. “Every tissue has its own metabolic flavor,” says bioengineer Christian Metallo of the University of California, San Diego.

Maddocks expects the cancer-diet field will take years to move from “piecemeal forays” to a clear understanding of each diet’s pros and cons. Establishing that a specific diet works well enough to become part of routine clinical care also will take time. But Zitvogel says fighting cancer with diet is no longer a fringe idea. The field is at the start of “a new era where people will really take diet seriously into account,” she says. “The time is ripe.” ■

INSIGHTS



LETTERS

NEXTGEN VOICES: ASK A PEER MENTOR

Forging remote relationships

As we enter year 2 of the COVID-19 pandemic, many scientists continue to work remotely, including those starting new jobs. We called on young scientists to submit advice for the researcher below, who grapples with the challenges of working remotely in a new lab. Responses included reflective questions, stories about personal experiences, and advice. Read a selection of the peer mentors' thoughts here. Follow NextGen Voices on Twitter with hashtag #NextGenSci. Read previous NextGen Voices survey results at <https://science.sciencemag.org/collection/nextgen-voices>. —Jennifer Sills

Dear NextGen VOICES peer mentors,

I am excited to be joining a new lab for my first postdoc. It has been a huge relief to get a job despite the even-tougher-than-usual job market. I have moved to a new city and am settling in, but because of the pandemic almost everyone is working remotely most of the time and all meetings are virtual. With limited in-person opportunities for collaboration and none for socializing, I am having trouble forging relationships with my new colleagues. This isolation is taking its toll on the scientific exchange I need for my project to succeed as well as my comfort as part of the team. How can I integrate myself into the lab environment?

Sincerely, New on the Quaran-team

Be proactive

Have you found a mentor? In my experience, integrating in a post-PhD world can be very difficult because you are the new person in a group of people whose lives are already settled. Try asking a senior member of the lab if they are willing to be your mentor and to help you navigate the new environment. Schedule a meeting as soon as possible to discuss the best way to succeed in the lab. (Don't wait for them to invite you—they are busy people!) I also suggest that you identify the most socially engaged person in meetings and ask that person for a private meeting to introduce yourself. The beating heart of the lab always has time for everyone and will know who is best to talk to about work-related issues.

Elvira Sojli

School of Banking and Finance, University of New South Wales, Kensington, NSW 2052, Australia.
Twitter: @esojli

Have you considered that your arrival is an opportunity for other group members? Your unique set of skills and scientific

background are important to the group. Why don't you organize [together with your principal investigator (PI), if appropriate] a round of talks or team-building events that are followed by spontaneous chats? Encourage people to attend the talks and create a friendly, safe, and non-competitive environment. I have found that initiating group get-togethers fosters interaction not only on a professional level but also a personal one.

Luca Soattin

Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine, and Health, The University of Manchester, Manchester, M13 9WU, UK.
Email: luca.soattin@manchester.ac.uk

Do you think there might be characteristics of this lab environment other than the imposed isolation that are preventing you from integrating successfully? Although the isolation you describe differs from my experience joining a lab before the pandemic, I faced roadblocks in being relatively young and inexperienced. It was intimidating, at first, to be thrust into a world where I had to learn on the spot and meet the high expectations of the existing lab workflow. These challenges created a relationship disconnect much like the one you describe. Just as during normal times, a healthy conversation can spark a healthy relationship, so begin there. Schedule virtual meetings with your colleagues to provide project updates and ask for any assistance you need. By taking initiative, you invite them to form stronger connections with you, scientifically and socially.

Shaan Patel

Vagelos Molecular Life Science Program, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: spatel20@sas.upenn.edu

Are you comfortable being flexible and trying new approaches? I recently applied to medical residency training programs. In normal times, I would have had the opportunity to meet would-be colleagues in person. Instead, networking took place online. Although I was concerned that it would be difficult to tell whether I fit in, I was pleasantly surprised by how committed program directors and residents were to chatting online. In the end, I probably met more people than if I had been limited by travel. I suggest being proactive and adaptable. Conversations might not happen organically, but you can still reach out to your lab members for a quick Zoom call to discuss a question. Lab socials may no longer take place at the local bar, but there are increasingly fun online games that have been designed for video calls.

My research group has a channel on our Slack account dedicated to sharing updates about our hobbies, such as a recent meal we made or the view from our bike ride. Despite the pandemic, we still need to collaborate, so we must be resilient in how we meet this moment.

Cody Lo

Faculty of Medicine, University of British Columbia, Vancouver, BC V5Z 1M9, Canada.
Twitter: @cody_lo

Show interest in others' work

Have you considered approaching a colleague to start a new project that builds on their research agenda? Your current priority may be to have scientific exchanges to complete your own projects, but in my experience, providing meaningful insights into a colleague's work is an effective approach for establishing yourself as a valuable member of the team. After carefully reading a colleague's work, which is ideally closely related to your own, have an in-depth discussion with them on the subject. Connecting their work with your own can open the door to receiving valuable knowledge and feedback from your more experienced peer. Your insights into their research can also pave the way for future collaborations. Moreover, this discussion will help you forge a social connection and demonstrate your collegiality, ultimately making you more comfortable.

Samuel Nathan Kirshner

School of Information Systems and Technology Management, University of New South Wales, Sydney, NSW 2052, Australia.
Email: s.kirshner@unsw.edu.au

Have you invited others to share their thoughts with you? I'm a senior student in my lab group, and I wasn't sure how to connect with a new student. One day, after a lab meeting, he emailed me to follow up on something I had said that resonated with him. This made me feel like we had a shared experience that we could build on. I would suggest trying this strategy to find common ground with your new colleagues.

Theresa B. Oehmke

Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. Twitter: @t_oehmke

Is fear your primary obstacle? As a young researcher in a new lab, I found myself self-conscious about my lack of experience and knowledge. I worried that my colleagues didn't truly consider me part of the lab, and I struggled to make

connections. What helped me overcome this fear was vocalizing my curiosity. I talked to the postdoc about the lab's publications, asked the grad students about their projects, and emailed the PI about new project ideas. In every good lab, researchers are eager to share their work. When I expressed my desire to learn, the lab welcomed me and my questions without any hesitation.

Junwon Kim

Vagelos Molecular Life Science Program, University of Pennsylvania, Philadelphia, PA 19104, USA.
Email: jkim0312@sas.upenn.edu

Start small

Have you tried working on a small project together? I have noticed that the postdocs who have joined us during the pandemic integrated fairly well by handling projects such as online webinars, which gave us the opportunity to exchange ideas and work as a group. If there is no ongoing project, you could organize one. It does not need to be something grand and fancy, just a small endeavor that you think would bring out the core strengths of your team. This will not only bring everyone closer but also highlight to others that you are passionate about what you do.

Khor Waiho

Higher Institution Centre of Excellence, Institute of Tropical Aquaculture and Fisheries, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, 21030, Malaysia. Email: waiho@umt.edu.my

Have you tried asking for help? Even a question about logistics or policies will open the door to communication. Colleagues will likely be happy to lend a hand, and once you start talking, you will be able to learn about your colleagues' academic experience, interesting news in the group, recent research progress, and academic difficulties. This will help you develop stronger relationships with other lab members.

Jian Zhang

School of Public Administration, Central South University, Changsha, Hunan, 410075, China.
Email: zhangjian3954@126.com

Do you get along well with anyone on the team in particular? I've managed big teams of people and noticed that someone who tries to connect with everyone tends not to make deep relationships with anyone. If you're on a big team, it can be overwhelming; if you're on a small team, it can seem like the group is already formed and has no place for you. I suggest that you pick just one person

with whom you think you can connect. If you find someone who is willing to have one-to-one chats online, then you can build on that. If you two can find things to do together, others will want to join in. That way you can grow your connections. It may seem slow at first, but it will be worthwhile. Of course, picking someone who is going to be on the team for a while is a good idea, too, as people move on alarmingly quickly!

Timothy L. Easun

School of Chemistry, Cardiff University, Cardiff, CF10 3AT, UK. Twitter: @TimEasun

Network creatively

Have you invited your new colleagues to informal, remote, one-on-one meetings to find out about each person's project, skills, and specific roles in the lab? These meetings could be the beginning of new collaborations and meaningful friendships. Schedule video calls with a variety of people beyond your lab as well, such as other PIs in your new Institute, researchers working in facilities that you may use in the future, and students happy to profit from your experience. If you don't know them, drop them an email and ask if they want to meet. Don't be surprised if you find out that they are also willing to talk about science and to discuss future projects with you.

Ana Neves-Costa

Instituto Gulbenkian de Ciência,
2780-156 Oeiras, Portugal.
Email: ananevescostaana@gmail.com

Have you tried joining virtual meet-ups tailored to your goals? My lab is a computational biology group that worked semi-remotely even before the pandemic, so it took a long time for me to make connections. Try signing up for department or postdoc-specific email listservs that advertise virtual social opportunities such as coffee hours. There are often only a few people there, which I think makes for more meaningful conversations. To help with productivity, members of my lab started scheduling virtual co-working sessions. At set times during the week, we meet on Zoom and chat for a few minutes about what we're planning to work on. Then we mute ourselves and work together for a couple of hours. It's a great way to have some accountability and structure in the day while staying connected with co-workers.

Beth Adamowicz

Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55401–2605, USA.
Twitter: @BethA_z

Have you thought about socializing outside of your laboratory group? When I moved across the country to start my postdoc, I knew no one at my new institution. Through becoming involved in peer mentoring groups (we meet via Zoom), international societies, and community service organizations, I built a community that I can interact with on a social basis. Now is not a time to mourn the loss of working with lab mates but rather an opportunity to engage with the broader community in ways you haven't before.

Mark Martin Jensen

Department of Surgery, Massachusetts General Hospital, Boston, MA 02114, USA.
Twitter: @MMJensen3

Have you connected with your new colleagues on social media? When I was the freshman in the lab, I found myself in a similar situation. I read my colleagues' social media posts to learn what we had in common. Try to find someone who shares a hobby of yours. If possible with social distancing, invite that person to accompany you (maybe you could play chess online or take a hike outside). If practicing the hobby is inadvisable, at least you have a topic you can use to start a conversation, which could in turn make it easier to break the ice with the others.

Wagner Eduardo Richter

Department of Chemical Engineering, Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brazil. Email: richter@utfpr.edu.br

Bond through shared experience

Have you acknowledged the ways in which the pandemic is exacerbating the feelings of isolation and uncertainty that always come with forging connections with new colleagues? After my project mentor left the lab, I also had to navigate change in addition to pandemic restrictions. It was important to remind myself that any perceived detachment, difficulty in collaboration, or feelings of isolation were not a result of a personal deficit or failure but rather the consequences of unprecedented challenges. All labs are experiencing a struggle with collaboration and socialization. Ask if the lab has a messaging platform to facilitate casual and quick communication, suggest virtual bonding events, and contact colleagues directly to learn more about their projects or offer and request advice. You could even use the mutual experiences of pandemic life as a starting point to forge individual relationships.

Renée Louane Barbosa

Vagelos Molecular Life Science Program, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: renee22@sas.upenn.edu

Have you considered that there are most likely others in your lab in the same situation as you? I have noticed that whenever individuals feel isolated, there are always other people who do as well. This was especially true this past semester, when my freshman fall was spent at home rather than on campus. Instead of a classroom full of people, I was met with dozens of faceless names on Zoom. However, I realized that this experience was universal and though making connections was more difficult, it was worthwhile. I suggest embracing the discomfort that meeting people in a work setting always entails. It is important to remember that you are not a singular island in the ocean of the pandemic, but instead part of an archipelago. Find those around you and begin the difficult task of making connections. In video calls about a project, greet each person by name and ask them about their day. These interactions may seem minor, but they are how workplace relationships begin. Don't be afraid to reach out to people individually for a video call or socially distanced coffee. Everyone is doing their best to get through this time and will appreciate any effort you make.

Laura Baeyens


Vagelos Molecular Life Science Program, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: lbaeyens@sas.upenn.edu

Shared common experiences are the foundation of friendships and relationships. Although the challenges you face seem to be stifling your ability to build relationships, the experiences you are sharing with your colleagues right now are a powerful bonding opportunity. Years from now, you'll always have a "Remember when..." story to recall and share with one another. How will you be remembered and what will you be remembered for? Make it known to your colleagues that you are there for them. Volunteer to assist someone who is falling behind. Actively look for ways to make the lives of others in the lab better. Your colleagues will come to see you as being the person who remained optimistic, positive, reliable, and willing to step up in a moment of need. You'll build goodwill, social capital, and meaningful relationships along the way.

Bradley J. Cardinal

School of Biological and Population Health Sciences, Oregon State University, Corvallis, OR 97331, USA.
Email: brad.cardinal@oregonstate.edu

10.1126/science.abi4726

An illustration of the MAVEN spacecraft in orbit around Mars. The spacecraft is shown from a side-on perspective, with its large white dish antenna and solar panels extended. Mars is a large, reddish-orange sphere in the lower right, and the background is a dark space filled with stars.

Instruments on the MAVEN spacecraft investigated the atmosphere of Mars.

PERSPECTIVES

ASTRONOMY

Hydrated crust stores Mars' missing water

Mars' liquid water might not have been lost to space as thought

By **Hiroyuki Kurokawa**

Mars is an arid and cold planet. Its surface water is found mainly as polar ice because of the low temperature. The amount of ice is far smaller than that of Earth's seawater by about three orders of magnitude. In a unit used in the community, the total volume is ~20 to 40 m global equivalent layer (GEL). However, Mars exploration missions have found plenty of evidence for a large amount of liquid water that shaped the Mars landscape. Such evidence includes valley networks, paleo-shorelines, conglomerates, and hydrated minerals. The geomorphologically estimated volume of missing water is on an ocean scale, ~100 to 1500 m GEL (1). On page 56 of this issue, Scheller

et al. (2) report a theoretical modeling study to quantify the amount of hydrated minerals in the surface crust that stores water. This is possibly the largest sink of Mars' missing ancient water.

Hydrogen (H) escape to space has been thought to be the most promising mechanism to account for the loss of water from the Mars surface. Surface water evaporates as water vapor, which then dissociates in the atmosphere into H that eventually escapes into space. Mars' small size (and thus, a gravity weaker than that of Earth) and/or the lack of the global magnetic field to shield its atmosphere against the solar wind can cause its atmospheric escape rate to be greater than that of Earth. However, the present-day escape rate, measured recently by using instruments aboard the Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft, is (if integrated over 4.5 billion years) too small to remove the ancient water volume (3). Because atmo-

spheric escape chiefly removes H and leaves deuterium (D), the integrated H loss is recorded as an increase in the D/H ratio of surface water. However, estimating the integrated loss by using the D/H ratio of the remnant water only gives the lowest estimate of ancient water volume. This has led previous studies to propose the existence of missing water reservoirs, such as undiscovered subsurface ice (4).

Scheller *et al.* demonstrate that a model that considers crust hydration can reconcile the gap between the volumes of ancient water and present-day ice, the observed H escape rate, and D/H constraints. Chemical reactions of water with rock form hydrated minerals such as smectite, which was found on Mars with orbital remote sensing and in situ measurements. The model also shows that H originally in Mars' water is mostly incorporated into the hydrated minerals. The smaller amount of remnant H is released to the atmosphere and then escapes

Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan.
Email: hiro.kurokawa@elsi.jp

to space. In this scenario, no hypothetical missing reservoir is needed.

The crust hydration scheme has several implications for Mars' evolution. Global crust hydration may require a warm and wet climate to sustain liquid water. H released from the water-rock reactions itself could possibly sustain the warm climate because of its greenhouse effect in a dense carbon dioxide (CO₂) atmosphere (the Mars atmosphere is 96% CO₂) (5). Whereas H escape from water would lead to oxidation of the Mars surface (Mars is 0.2% O₂), crust hydration could possibly induce the accumulation of H in Mars' atmosphere (which would ultimately escape into space).

Crust hydration is common on Earth, but plate tectonics recycle the crust to the underlying mantle, and volcanism returns water to the hydrosphere. The absence of plate tectonics on Mars causes irreversible crust hydration. The traditional view holds that the difference in planetary sizes and the presence or absence of global magnetic fields led to the divergent fates of the two planets. The notion of crust hydration on Mars supports the importance of plate tectonics for the sustainability of liquid water on terrestrial planets.

The crust hydration scenario does not mean that atmospheric escape is not a major factor in Mars evolution. Surface oceans, which may be required for global crust hydration, need to be sustained by the greenhouse effect of a dense atmosphere, which is thought to have been chiefly lost through atmospheric escape processes. The D/H ratio recorded in an old (~4.1 billion years) martian meteorite suggests that substantial water loss predates the periods of geomorphologically recorded liquid water (4, 6). The atmospheric escape rate could have been higher during this earlier period, when solar extreme ultraviolet radiation and solar wind were more intense. Future studies will need to quantify the contribution of crust hydration on water loss and how it changed throughout martian history. Nevertheless, Scheller *et al.*'s study highlights the importance of the aqueous alteration of crust as a potential driver of the climate change and the potential role of plate tectonics to control the sustainability of surface water, both of which are crucial for understanding planet evolution. ■

REFERENCES AND NOTES

1. M. H. Carr, J. W. Head, *J. Geophys. Res. Planets* **108**, E5 (2003).
2. E. L. Scheller *et al.*, *Science* **372**, 56 (2021).
3. B. M. Jakosky *et al.*, *Icarus* **315**, 146 (2018).
4. H. Kurokawa *et al.*, *Earth Planet. Sci. Lett.* **394**, 179 (2014).
5. R. Wordsworth *et al.*, *Geophys. Res. Lett.* **44**, 665 (2017).
6. J. P. Greenwood, S. Itoh, N. Sakamoto, E. P. Vicenzi, H. Yurimoto, *Geophys. Res. Lett.* **35**, L05203 (2008).

10.1126/science.abh4469

PALEONTOLOGY

The impactful origin of neotropical rainforests

A mass extinction event led to vast diversity and structural complexity of neotropical rainforests

By **Bonnie F. Jacobs¹** and **Ellen D. Currano^{2,3}**

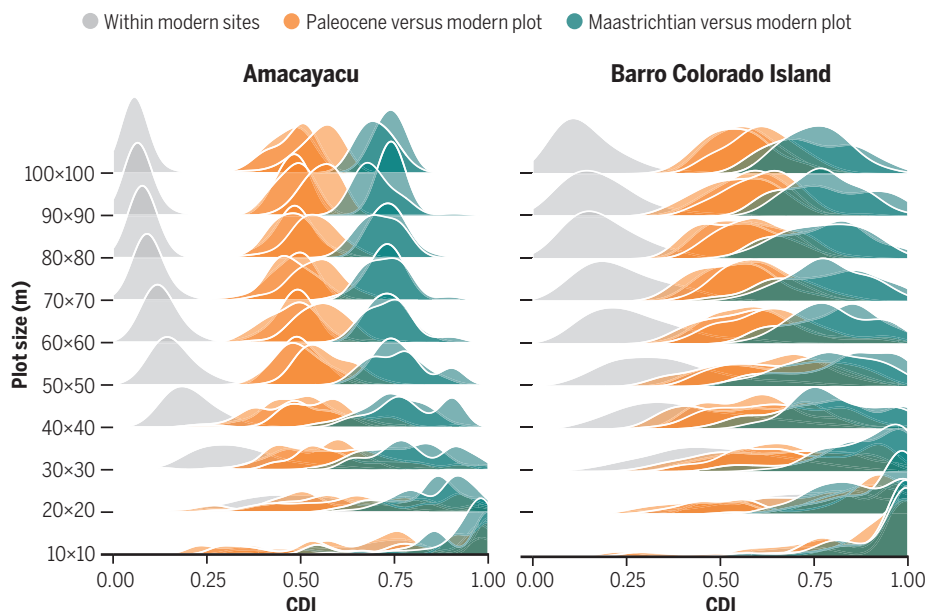
About 66 million years ago (Ma), at the boundary of the Cretaceous-Paleogene (K/Pg) period, a sudden mass extinction was triggered by the impact of a bolide, destroying an estimated three-quarters of Earth's plant and animal species. The long-term effects of this event varied across Earth, and little is known about the outcome in low-latitude regions of the world. On page 63 of this issue, Carvalho *et al.* (1) report analyses of fossil pollen and leaf data across the K/Pg boundary, ~1500 km south of the Chicxulub crater left behind by the impact. They assessed plant diversity and structure in the lowland tropics before and after the catastrophe, put their interpretations into the broader context of flowering plant (angiosperm) evolution, and answer one of the biggest questions in paleobotany:

When and how did the diverse, angiosperm-dominated, stratified tropical forests of South America emerge?

Before about the mid-20th century, the low-latitude regions of the world were severely underexplored paleontologically, owing in part to perceived hindrances, including surface vegetative cover and the extreme weathering (and decomposition) of organic matter in soils of the wettest areas. In contrast to this were the obvious benefits of working in the mid-latitude temperate zones, where sediments were easy to see, excavate, and core into and known to produce fossils. A dearth of paleontologists in tropical countries exacerbated this relative neglect because developing nations were just beginning to train and hire their own scientists. Today, the situation is much improved, but to address big questions through paleontology, it takes time to collect and study samples,

Leaf comparison

Leaves from unbiased fossil census sites are compared with leaves from trees in 50 randomly selected areas of varying size within a 25-ha forest plot at Amacayacu (Colombia) and a 50-ha forest plot at Barro Colorado Island (BCI, Panama) (1). Dissimilarity at the family level is shown as density plots (Chao-Sørensen dissimilarity index, CDI). The Paleocene sites are more similar in family composition to the living forests of Barro Colorado Island (Panama) and Amacayacu (Colombia) than they are to the Cretaceous census sites.



GRAPHIC: N. CARY/SCIENCE BASED ON M. R. CARVALHO ET AL. (1)

and big gaps remain regarding the evolution of Earth's tropical biodiversity and its role in global climate dynamics through time.

Well-preserved leaves from the Cretaceous Guaduas and the mid-Paleocene (the earliest epoch of the Paleogene, 66 to 56 Ma) Cerrejón and Bogotá floras of Colombia and a compilation of pollen samples from 39 cores spanning the K/Pg boundary provided Carvalho *et al.* with hundreds of samples from which forest composition and structure could be determined (see the figure). A suite of methods was applied, including leaf vein density measures and determining the range of leaf carbon isotope values, both of which served as proxies of light variations. Leaf insect damage diversity and intensity were analyzed to assess ecosystem connectedness and stability. In addition, numerical analyses of >600 pollen samples through time provided a view of diversity, composition, and rates of change across the K/Pg boundary.

The well-supported findings of Carvalho *et al.* reveal that angiosperms were well on the way to becoming dominant and diverse members of forest communities by the latest Cretaceous—a previously accepted consensus (2, 3). But it is now clear that in the tropics, as is true at higher latitudes, gymnosperms (cone-bearing plants, for example) and ferns were important members of forested communities before the K/Pg boundary. In addition, light availability in these forests, representing forest openness, was greater than in the early Paleogene. Across the boundary, diversity declined and became more dominated by angiosperms but did not exceed preboundary levels until about 60 Ma (an evolutionary recovery time of ~6 Ma). Although some angiosperm families found today in rich tropical forests were present in the latest Cretaceous (~69 Ma), they mingled with an (approximately equal) abundance of gymnosperms and ferns in a more open, nonstratified forest structure. It took a bolide smashing into Earth to reset these low-latitude ecosystems, putting them on a path to the rich, stratified tropical forests of today in the most productive terrestrial region of the world.

How do these findings differ from what is



A fossil leaf from a 58 to 60 million years ago (Paleocene) tropical rainforest in Colombia was among the samples assessed by Carvalho *et al.* to determine forest composition and structure.

known elsewhere? The K/Pg extinction event demolished nonavian dinosaurs and killed off >75% of all species worldwide. What has become clearer more recently is that plant extinction and turnover varied, depending upon proximity to the impact location, the nature of the preimpact flora, and climatic conditions in the millennia afterward.

Until recently, K/Pg floras were reported primarily from the US western interior basins of North Dakota, Montana, Colorado, and New Mexico. Here, some 3000 km from the impact, loss in diversity of pollen and macrofossil taxa across the boundary is reported to range from 30 to >50%. Insect damage also reflects the strong imprint of the boundary event at this latitude. The telltale signs of high diversity, such as the presence of specialized leaf mines, decline across the boundary. Not far from these interior basins, and up against the newly emerging Rocky Mountain front, is the Denver Basin's Castle Rock site, which, in a wet, warm climate, preserves a highly diverse flora only about 1.6 Ma after the boundary event (4). This unusual flora was referred to as a rainforest, but its structure is unknown. Yet although floral diversity is high, insect damage diversity is among the lowest reported in the Paleocene. Combined high plant and insect damage diversity, as occurred in the latest Cretaceous, does not recur among the western US basins for some 10 Ma into the Paleocene (5). Thus, there is substantial heterogeneity among Paleocene plant localities even among the western US basins.

More than 8000 km from the impact

crater, paleofloras from Patagonia document a very different pattern of change across the boundary. Latest Cretaceous pollen and spore assemblages from Patagonia indicate dominance by angiosperms, ferns, and gymnosperms, but, despite a large loss in overall diversity across the boundary, conifers extirpated elsewhere on Earth survived through to the earliest Paleocene (6). Patagonian macrofloras document a decline in specialized leaf miners across the boundary event but here took only 4 Ma (as opposed to 10 Ma) from which to recover (5, 7). Similarly, major loss of leaf species (45%) across the boundary and the emergence of a different pattern of leaf shapes (8) reflect substantial ecosystem change. But palynological (pollen and spore) samples show change of little consequence at higher taxonomic levels. Thus, Patagonia, >8000 km from the impact crater, experienced less severe consequences of the impact than

nearer sites to the north. As expected, palynological samples recently reported from New Zealand (9), ~12,000 km from the impact, show only moderate changes across the boundary. Gymnosperms declined about 10%, and although angiosperms increased to 60%, substantial dominance to 80% takes place later, in the Eocene (56 to 43 Ma).

The Carvalho *et al.* study and others show that a global catastrophe involving a mass extinction produces a different world, which recovers in a spatially and temporally heterogeneous way. At the K/Pg boundary, a bolide impact caused a mass extinction event, but the effects were heterogeneous—the consequences depended on proximity to the crater and local conditions such as climate. Today, the world is experiencing a sixth mass extinction event, but this time, there is no place on Earth far from the ultimate cause—humans. It seems that proximate perturbations are and will be substantial everywhere, even if they vary. ■

REFERENCES AND NOTES

1. M. R. Carvalho *et al.*, *Science* **372**, 63 (2021).
2. S. H. Lidgard, P. R. Crane, *Palaeobiology* **16**, 77 (1990).
3. P. R. Crane, S. Lidgard, *Science* **246**, 675 (1989).
4. K. R. Johnson, B. Ellis, *Science* **296**, 2379 (2002).
5. P. Wilf, C. C. Labandeira, K. R. Johnson, B. Ellis, *Science* **313**, 1112 (2006).
6. V. D. Barreda *et al.*, *PLOS ONE* **7**, e25455 (2012).
7. M. P. Donovan, A. Iglesias, P. Wilf, C. C. Labandeira, N. R. Cúneo, *Nat. Ecol. Evol.* **1**, 0012 (2016).
8. E. Stiles, P. Wilf, A. Iglesias, M. A. Gandolfo, N. R. Cúneo, *Palaeobiology* **46**, 445 (2020).
9. J. G. Prebble *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **566**, 110207 (2021).

10.1126/science.abh2086

¹Roy M. Huffington Department of Earth Sciences, Southern Methodist University, P.O. Box 750395, Dallas, TX 75275, USA. ²Department of Botany, University of Wyoming, 1000 E. University Avenue, Laramie, WY 82071, USA. ³Department of Geology and Geophysics, University of Wyoming, 1000 E. University Avenue, Laramie, WY 82071, USA. Email: bjacobs@mail.smu.edu

MEDICINE

Achieving fairness in medical devices

Learning from computer science can make medical devices fair for all races and sexes

By Achuta Kadambi^{1,2}

The hardware or software that operates medical devices can be biased. A biased device is one that operates in a manner that disadvantages certain demographic groups and influences health inequity. As one measure of fairness, reducing bias is related to increasing fairness in the operation of a medical device. Initiatives to promote fairness are rapidly growing in a range of technical disciplines, but this growth is not rapid enough for medical engineering. Although computer science companies terminate lucrative but biased facial recognition systems, biased medical devices continue to be sold as commercial products. It is important to address bias in medical devices now. This can be achieved by studying where and how bias arises, and understanding these can inform mitigation strategies.

Bias in medical devices can be divided into three broad forms (see the figure). A medical device can exhibit physical bias, where physical principles are biased against certain demographics. Once data are collected, computational bias, which pertains to the distribution, processing, and computation of data that are used to operate a device, must be considered. Subsequent implementation in clinical settings can lead to interpretation bias, where clinical staff or other users may interpret device outputs differently based on demographics.

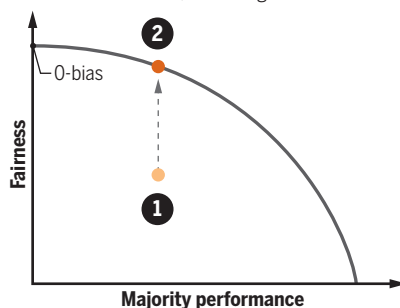
The physical working principle of a medical device is biased when it exhibits an undesirable performance variation across demographic groups. An example of physical bias occurs in the context of optical biosensors that use light to monitor vital signs. A pulse oximeter uses two colors of light (one in near-infrared and the other in visible light) to measure blood oxygenation. Through the pulse oximeter, it is possible to diagnose occult hypoxemia, low levels of arterial oxygen saturation that are not detectable from symptoms. However, a recent study found that Black patients had about three times the frequency of undiagnosed occult hypoxemia as measured by pulse oximeters (1). Dark skin tones respond differently to these wavelengths of light, particularly visible light.

Because hypoxemia relates to mortality, such a biased medical device could lead to disparate mortality outcomes for Black and dark-skinned patients.

Physical bias is not restricted to skin color. For example, the mechanical design of implants for hip replacement exhibits a potentially troubling gender disparity. The three-dimensional models used to design hip-joint implants sometimes do not account for the distinct bone structure of female hips (2). This could lead to alignment issues and relatively poor outcomes for affected females. This problem was one motivation for the development of gender-specific implants.

Measuring fairness

Fairness can be quantified based on ϵ -bias. Fairness is maximized when $\epsilon = 0$, achieving a state of 0-bias.



1 Pareto inefficient

Improving fairness is possible to do without changing performance

2 Pareto optimal

Reaching the Pareto frontier, where it is not possible to improve fairness without decreasing performance and vice versa

Fortunately, physical challenges can also be addressed through unexpected technical innovation, such as in the example of the remote plethysmograph. This device measures heart rate through visual changes in skin color. Because visual cues are biased, researchers developed an alternative approach using motion cues to estimate heart rate. Because motions are visible on the surface of skin, the technique is less biased by subsurface melanin content (3). With the goal of promoting fairness, an exciting technical direction of studying motion cues instead of color cues has been advanced.

Computational workflows are becoming more tightly coupled with devices, which increases the number of entry points where computational bias can invade medical tech-

nologies. An aspect of computational bias is dataset bias. Consider the following example from x-ray imaging: Diagnostic algorithms can learn patterns from x-ray imaging datasets of thoracic conditions. However, these imaging datasets often contain a surprising imbalance, where females are underrepresented. For example, despite having a sample size of more than 100,000 images, frequently used chest x-ray databases are ~60% male and ~40% female (4). This imbalance worsens the quality of diagnosis for female patients. A solution is to ensure that datasets are balanced. Somewhat unexpectedly, balancing the gender representation to 50% female boosts diagnostic performance not only for females but also for males (4). Despite best efforts, demographic balancing of a dataset might not be possible. This could be due to conditions that present more often in one sex than the other. In such cases where balancing a dataset is truly infeasible, transfer learning can be used as a step toward a longer-term solution (5). Transfer learning could repurpose design parameters from task A (based on a balanced dataset) to task B (with an unbalanced dataset). In the future, it might be possible to balance a dataset using a human digital twin. These are computational models that can be programmed to reflect a desired race, sex, or morphological trait.

Another form of computational bias is algorithm bias, where the mathematics of data processing disadvantages certain groups. Now, software algorithms are able to process video streams to detect the spontaneous blink rate of a human subject. This is helpful in diagnosing a variety of neurological disorders, including Parkinson's disease (6) and Tourette syndrome (7). Unfortunately, traditional image-processing systems have particular difficulty in detecting blinks for Asian individuals (8). The use of such poorly designed and biased algorithms (9) could produce or exacerbate health disparities between racial groups.

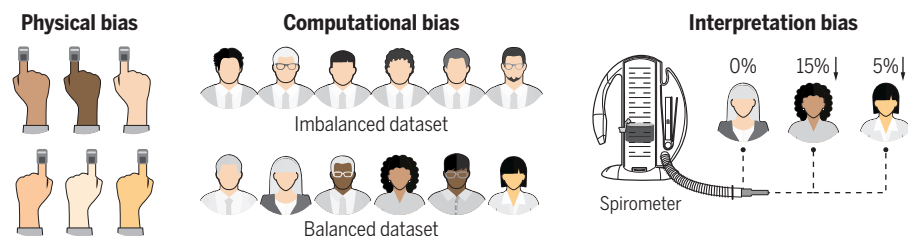
Interpretation bias occurs when a medical device is subject to biased inference of readings. An example of a misinterpreted medical device is the spirometer, which measures lung capacity. The interpretation of spirometry data creates unfairness because certain ethnic groups, such as Black or Asian, are assumed to have lower lung capacity than white people: 15% lower for Black people and about 5% lower for Asian people. This assumption is based on earlier studies that may

¹Department of Electrical and Computer Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ²Department of Computer Science, UCLA, Los Angeles, CA, USA. Email: achuta@ee.ucla.edu



Bias in medical devices

A device can be biased if its design disadvantages certain groups on the basis of their physical attributes, such as skin color. For example, pulse oximeters (see the photo) detect changes in light passed through skin and are less effective in people with dark skin. Computational techniques are biased if training datasets are not representative of the population. Interpretation of results may be biased according to demographic groups, for example, with the use of “correction factors.”



have incorrectly estimated innate lung capacity (10). Unfortunately, these “correction factors,” based on questionable assumptions, are applied to the interpretation of spirometer data. For example, before “correction,” a Black person’s lung capacity might be measured to be lower than the lung capacity of a white person. After “correction” to a smaller baseline lung capacity, treatment plans would prioritize the white person, because it is expected that a Black person should have lower lung capacity, and so their capacity must be much lower than that of a white person before their reduction is considered a priority.

However well intentioned, errors in “correction” for race (or sex) can disadvantage the groups it seeks to protect. In the spirometer example, the device designers conflated a racial group’s healthy lung capacity with their average lung capacity. This assumption does not account for socioeconomic distinctions across race: Individuals who live near motorways exhibit reduced lung capacity, and these individuals are often from disadvantaged ethnic groups. The spirometer is just one of several examples of systemic racism in medicine (11).

If our society desires fair medical devices, it must reward a fair approach to innovation. It is inspiring to observe the speed at which the artificial intelligence (AI) community has recognized fairness in its endeavors. Authors can be encouraged by journals to address the societal implications of their technologies and include a “broader impacts” statement that is considered in peer review. This has already been introduced at an AI journal to encourage consideration of the diversity of potential users of their software (12). Fairness research in AI is increasingly garnering scholarly acclaim. For example, a seminal report highlighted the widespread problem of bias in face recognition, which found that darker-skinned females are misclassified at rates up to 34.7% while the maximum error rate for lighter-skinned males is only 0.8% (13). In response to concerns of fairness, action is being taken. For example, Amazon Inc. has re-

cently banned the use of its facial-recognition products by police until bias concerns can be resolved. There is still a long way to go in addressing bias in AI, but some of the lessons learned can be repurposed to medical devices.

A “fairness” statement for the evaluation of studies of medical devices could use the three categories of bias as a rubric: physical bias, computational bias, and interpretation bias. A medical-device study does not need to be perfectly unbiased to be reported. Indeed, it may not always be possible to remove all sources of bias. For example, an oximeter reliant on an optical sensor is likely to remain biased against dark skin (1). The fairness statement can consist of technical explanations for how attempts to mitigate bias failed and suggest technical compensations for disadvantaged groups (e.g., collect additional data points for dark-skinned people). This is consistent with the introduction of “positive biases,” where race-aware and gender-aware methodologies are explicitly designed to counteract negative bias (14).

Additionally, the inclusion of fairness metrics in studies of medical devices could be considered. Choosing the right fairness metric of an algorithm is a quantitatively challenging computer science exercise (15) and can be abstracted here as “ ϵ -bias,” where ϵ quantifies the degree of bias across subgroups. For example, 0-bias would be seen as perfectly fair. Achieving 0-bias on its own is trivial: Simply return a measurement that is consistently useless across demographics. The problem is to maximize performance and minimize ϵ -bias. This may present a Pareto trade-off, where maximizing the performance and minimizing bias are objectives at odds with each other. A Pareto curve can quantitatively display how changing device configuration varies the balance between performance and fairness (see the graph). Such analyses might be a useful inclusion in medical-device studies.

Achieving fairness in medical devices is a key piece of the puzzle, but a piece nonetheless. Even if one manages to engineer a fair medical device, it could be used by a clinical

provider who has conscious or subconscious bias. And even a fair medical device from an engineering perspective might be inaccessible to a range of demographic groups, owing to socioeconomic reasons. Several open questions remain. What is an acceptable trade-off between device performance and fairness? It is also important to consider how biases that are not easy to predict or easy to observe at scale can be dealt with. Race and sex are also part of human biology. How can positive biases be properly encoded into medical-device design? Diversity and inclusion have gained increasing attention, and the era of fair medical devices is only just beginning. ■

REFERENCES AND NOTES

1. M. W. Sjoding *et al.*, *N. Engl. J. Med.* **383**, 2477 (2020).
2. C. W. Hartman *et al.*, *Semin. Arthroplasty* **20**, 62 (2009).
3. G. Balakrishnan, F. Durand, J. Guttag, in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, 2013), pp. 3430–3437.
4. A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, E. Ferrante, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12592 (2020).
5. S. Jabbour *et al.*, in *Proceedings of the Fifth Machine Learning for Healthcare Conference*, F. Doshi-Velez *et al.*, Eds. (Proceedings of Machine Learning Research, 2020), pp. 750–782.
6. R. Sanddy, *Int. J. Neurosci.* **51**, 99 (1990).
7. C. N. Karson *et al.*, *J. Nerv. Ment. Dis.* **173**, 566 (1985).
8. J. Zou, L. Schiebinger, *Nature* **559**, 324 (2018).
9. Z. Obermeyer *et al.*, *Science* **366**, 447 (2019).
10. L. Braun, *Breathing Race into the Machine: The Surprising Career of the Spirometer from Plantation to Genetics* (Univ. of Minnesota Press, 2014).
11. A. H. Wingfield, *Science* **369**, 351 (2020).
12. B. Hecht *et al.*, “It’s time to do something: Mitigating the negative impacts of computing through a change to the peer review process,” *ACM Future of Computing Blog*, 29 March 2018; <https://acm-fca.org/2018/03/29/negativeimpacts/>.
13. J. Buolamwini, T. Gebru, in *Proceedings of the Conference on Fairness, Accountability and Transparency*, S. A. Friedler, C. Wilson, Eds. (Proceedings of Machine Learning Research, 2018), pp. 77–91.
14. D. Cirillo *et al.*, *NPG Digi. Med.* **3**, 81 (2020).
15. J. Kleinberg, S. Mullainathan, M. Raghavan, in *Proceedings of the Eighth Innovations in Theoretical Computer Science Conference*, C. H. Papadimitriou, Ed. (Schloss Dagstuhl, 2017), pp. 43:1–43:23.

ACKNOWLEDGMENTS

I thank P. Chari, L. Jalilian, K. Kabra, M. Savary, M. Majmudar, and the Engineering 87 class at UCLA for constructive feedback. I am supported by a National Science Foundation CAREER grant (IIS-2046737), Google Faculty Award, and Sony Imaging Young Faculty Award.

10.1126/science.abe9195

PHOTONICS

Death and rebirth through nonlinear control

Nonlinearity in complex photonic systems uncovers implications beyond optics

By **Piotr Roztocki**¹ and **Roberto Morandotti**^{1,2}

Nonlinear systems, characterized by outputs that are not proportional to their inputs, form the bulk of real systems in nature and applied science. Nonlinearity in a platform can substantially extend the range of its accessible functionalities, as seen in digital electronics and artificial neural networks, for example. However, there are many systems for which the impact of nonlinearity is not yet clear. This includes a complex class of systems that exhibits both topological and non-Hermitian aspects—i.e., a set of interesting properties invariant under continuous deformations and unconventional operators that describe the total system energy, respectively. In such systems, nonlinear control is, thus far, underexplored. On page 72 of this issue, Xia *et al.* (1) bridge the investigation of non-Hermitian topological physics with nonlinearity, leading to implications for complex systems in acoustics, plasmonics, polaritronics, and ultracold atoms.

The field of topological photonics was founded from developments in condensed-matter physics, particularly from ideas relating to the celebrated quantum Hall effect (2) and topological insulators (3). Analogous effects were shown to be accessible in a variety of photonics platforms, which enabled major fundamental and engineering advances, including the demonstration of electromagnetic states protected from scattering (4, 5), topological insulator lasers (6, 7), and others. In turn, studies of non-Hermitian physics

and parity-time (PT) symmetry countered the conventional belief that quantum Hamiltonians had to be Hermitian. Indeed, a weaker condition (commutation with the PT operator) was determined sufficient to enable real spectra (8). Such systems interact with their environments through complex potentials, the imaginary part of which determines whether a system is gaining or

losing energy. The exploration of analogous systems in optics, with photons experiencing well-designed gain and loss, has enabled rich investigations of fundamental physics (9–12) while also establishing new engineering paradigms through PT-based devices and sensors (12).

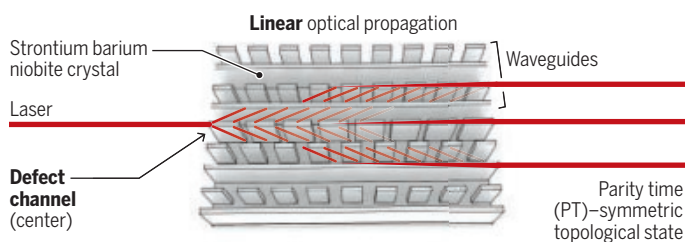
However, although the interplay of topology and non-Hermitian physics is already subject to active and productive study (6, 7, 13, 14), nonlinear effects have been underexplored in the joint context of these disciplines. Investigations dedicated specifically to the interplay of all three properties are rare to nonexistent.

Xia *et al.* aim to address this research gap and explore a direction in nonlinear non-Hermitian topological photonics. Toward this end, the authors introduce an easy-to-access, versatile platform for the exploration of nonlinearity in such systems, based on the reconfigurable writing of an optical waveguide lattice into a biased photorefractive crystal (see the figure). Structures of this type, called Su-Schrieffer-Heeger lattices (15), are an established tool for topological studies in photonics. Using a bias field, Xia *et al.* are able to modify the waveguide nonlinear response, causing optical self-focusing and defocusing effects that change how light propagates throughout the lattice. Although this nonlinearity affects only the real part of the bulk refractive index, this equivalently tunes both the real and imaginary parts of the waveguide potential function.

This degree of freedom enables Xia *et al.* to demonstrate destruction and restoration of non-Hermitian topological states through nonlinear control. In particular, the authors show that a state initially PT-symmetric in the linear regime can be destroyed by introducing optical nonlinearity. The

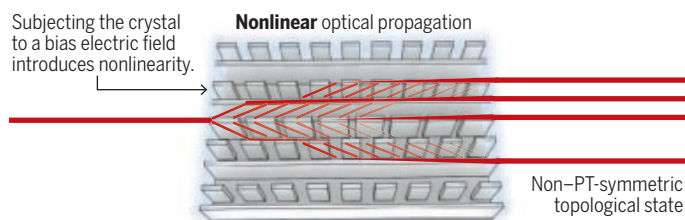
Controlling states in a complex system

An optical platform uses waveguides to explore nonlinear control of a complex system. The ability to use local control to change the global character of the system makes it a flexible strategy for investigating complex phenomena.



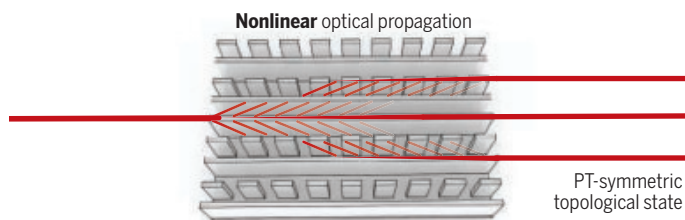
Linear and symmetric

When a light probe is injected into the so-called defect channel of the optical lattice, it couples to other waveguides as it propagates through. Here, it supports a light state exhibiting symmetry.



Nonlinear tuning brings destruction

Applying a bias electric field introduces nonlinearity into the system. This brings about optical effects that change how the light probe couples to other modes. The result is the destruction of the symmetric state.



Nonlinear tuning brings restoration

If the supported state is not symmetric in the linear regime, the introduction of nonlinear effects can modify light propagation such that the symmetric state is restored.

¹INRS-Énergie, Matériaux et Télécommunications, Varennes, Québec, Canada. ²Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China. Email: piotr.roztocki@inrs.ca; morandotti@emt.inrs.ca

inverse also holds, as an initially non-PT-symmetric state can be nonlinearly tuned to restore PT symmetry. Counterintuitively, these results show that topology and PT symmetry, which broadly describe global properties of complex systems, can be tuned using optical nonlinearity, which is a local effect. This demonstrates, in a sense, an interplay of local and global effects.

Xia *et al.* also study another pair of seemingly antagonistic effects—the sensitivity of their system near exceptional points (EPs) versus the robustness of topologically protected states. EPs are degenerate points specific to the operation of non-Hermitian systems and characterized by enhanced sensitivity to perturbation (12). EP physics are thus especially interesting when combined with the protection from perturbations that topological systems can offer (3, 7). Xia *et al.* find that although the nonlinear system “inherits” topological protection from its linear counterpart, this stability wears off away from the protected mode in a way that is dependent on how close the system is to the EP.

Xia *et al.* open the door for investigating this overlap of disciplines, but many questions remain. There is a clear need for a

“This demonstrates, in a sense, an interplay of local and global effects.”

general theoretical framework to describe the nonlinear driving of non-Hermitian topological systems, as well as for an extension of current symmetry classification methods to such systems. The potential impact of the authors’ work on the development of new devices is also exciting. Such nonlinear control schemes may be beneficial in many other complex systems outside optics. ■

REFERENCES AND NOTES

1. S. Xia *et al.*, *Science* **XXX**, 72 (2021).
2. D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).
3. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).
4. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljacic, *Nature* **461**, 772 (2009).
5. M. C. Rechtsman *et al.*, *Nature* **496**, 196 (2013).
6. M. A. Bandres *et al.*, *Science* **359**, eaar4005 (2018).
7. G. Harari *et al.*, *Science* **359**, eaar4003 (2018).
8. C. M. Bender, S. Boettcher, *Phys. Rev. Lett.* **80**, 5243 (1998).
9. K. G. Makris, R. El-Ganainy, D. N. Christodoulides, Z. H. Musslimani, *Phys. Rev. Lett.* **100**, 103904 (2008).
10. A. Guo *et al.*, *Phys. Rev. Lett.* **103**, 093902 (2009).
11. C. E. Rüter *et al.*, *Nat. Phys.* **6**, 192 (2010).
12. R. El-Ganainy *et al.*, *Nat. Phys.* **14**, 11 (2018).
13. C. Poli, M. Bellec, U. Kuhl, F. Mortessagne, H. Schomerus, *Nat. Commun.* **6**, 6710 (2015).
14. J. M. Zeuner *et al.*, *Phys. Rev. Lett.* **115**, 040402 (2015).
15. W. P. Su, J. R. Schrieffer, A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979).

10.1126/science.abg5441

NEUROSCIENCE

How dopamine leads to hallucinations

An increase of dopamine in the striatum virtualizes a nonexistent auditory signal in mice

By Miriam Matamalas

The ability to detect external stimuli rapidly and accurately by building internal sensory representations is a central computation of the brain that is critical to guide behavior. Such expectations (or priors) may be acquired throughout the lifetime of an individual and are important to influence perception, particularly when incoming sensory signals are ambiguous (1). But this process is not exempt from failure. Hallucinations (perceptual experiences without external stimuli) seen in conditions such as schizophrenia are thought to result from giving too much weight to priors, creating an imbalance at the expense of actual sensory evidence (2, 3). Sustained high-dopamine tone in the striatum has been proposed to contribute to this imbalance (4); however, it has remained unclear how the dopaminergic perturbation leads to the generation of hallucinations. On page 51 of this issue, Schmack *et al.* (5) uncover the neurobiological mechanisms that underlie dopamine-dependent auditory hallucinatory states, with therapeutic implications.

In the laboratory, it has been consistently shown that hallucination-like perceptual experiences (percepts) can be experimentally elicited by establishing associations between stimuli (6). For example, repeated visual-auditory stimulus pairings lead to auditory percepts when visual stimuli are presented alone, an effect that appears widely across the animal kingdom. On the basis of this effect, Schmack *et al.* developed a sensory detection task that is conducted under variable levels of perceptual uncertainty to maximize the proportion of hallucination-like percepts in mice. In their task, the animals learned a visual cue (a light) that was presented either with an auditory signal (a tone) that was embedded in background noise or with the background noise alone. After repeated presentations of this set of associations (light and tone above noise versus light and noise), mice could appropriately choose a response according to what was presented.

However, when Schmack *et al.* increased the uncertainty in the task by introducing auditory signals hardly distinguishable from

the noise, conditioned hallucinations started to emerge: the light-evoked representation of the tone established early in training biased the mice toward the “tone present” response even when the tone was largely attenuated or completely absent. The bias was even greater when the expectancy of hearing the auditory signal was manipulated by increasing the number of light-tone pairings during training, in line with the idea that learned—perhaps aberrant—associations may seed the establishment of overweight priors that can ultimately distort perception (7).

A key feature of hallucination-prone individuals is that they show inadequate overconfidence in decision-making tasks, particularly when engaged in an erroneous choice (8). Schmack *et al.* recorded the amount of time an animal is willing to wait for the reward after a response and used this measure as a behavioral proxy for confidence (9). On the basis of this postdecision time investment, Schmack *et al.* found a fraction of false alarm responses (tone choice in the absence of a tone) in which mice had a high conviction that the auditory signal had been presented, and therefore the authors demonstrated that high-confidence hallucination-like percepts (HALIPs) could indeed be modeled in mice (see the figure). In support of the hallucinogenic nature of this response, administering ketamine (a drug that precipitates psychogenic symptoms) specifically increased HALIP rate in mice. Moreover, the behavioral task could be readily translated to nonclinical human participants, whose self-reported auditory hallucinations were positively correlated with HALIPs (false alarm responses in the task).

A critical unresolved question is whether dopamine dysregulation at specific striatal loci is at the root of psychotic symptoms. Schmack *et al.* addressed this by studying the dynamics of dopamine release in the tail of the striatum (TS), an associative region implicated in the perception and memory of auditory signals in which cortico-, thalamo- and nigro-striatal projections converge (10).

Decision Neuroscience Laboratory, School of Psychology, University of New South Wales, Sydney, Australia. Email: m.matamalas@unsw.edu.au

ing signals render T cells progressively more inert, coupling their continuing recognition of tumor or viral antigens to an inability to lyse antigen-expressing cells, secrete effector molecules, or proliferate. The magnitude and duration of inhibitory signals 2 and 3 regulate the intensity of T cell exhaustion, and the effects can initially be reversed—for example, by disrupting negative signal 2 with antibodies against PD-1 or PD-1 ligand (PD-L1) (5).

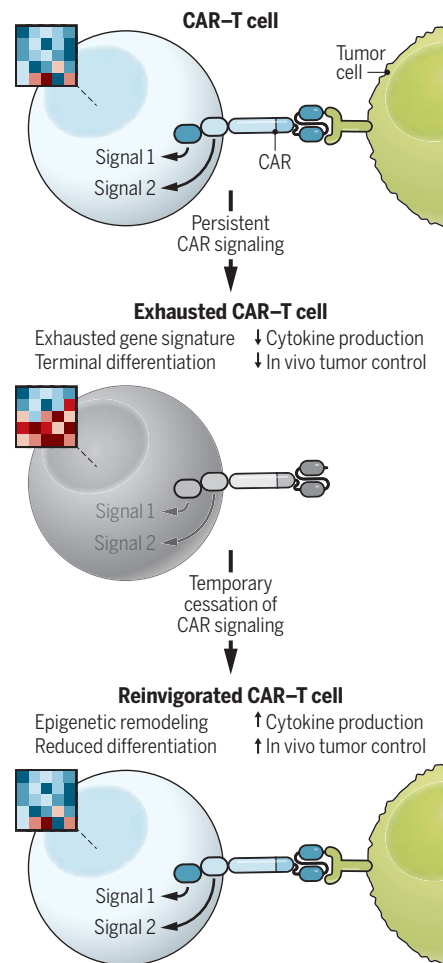
CARs bypass both the endogenous TCR specificity (signal 1) and the dependence on a separate and positive costimulatory signal 2 by incorporating both signals in one artificial molecule, which is triggered upon binding to a target antigen that it is engineered to recognize. High CAR expression and structural attributes of the receptor can promote their spontaneous, persistent, and ligand-independent signaling (7, 8). Such tonic stimulation produces a state resembling persistent T cell stimulation by cognate antigen. Prolonged tonic CAR signaling alone is often sufficient to accelerate terminal T cell differentiation to short-lived effector subsets and induce a gene expression signature associated with exhausted T cells. These changes substantially reduce the antitumor activity of CAR-expressing T cells in preclinical models (8).

The T cell dysfunction induced by tonic CAR signaling was thought to be similar to the prolonged and hard-to-reverse functional exhaustion observed in tumor- and virus-specific T cells. It was therefore assumed that only extensive and time-consuming CAR optimization could avert the problem. Weber *et al.* challenge this concept, demonstrating that the transcriptional, and even the epigenetic, changes induced in engineered T cells by unrelenting CAR activation can be reversed by temporarily blocking CAR signaling (see the figure).

The authors controlled spontaneous signaling either by adjusting turnover of the CAR itself or by pharmacologically inhibiting key kinases involved in proximal TCR and CAR signaling with dasatinib. In both models, transient rest from CAR signaling restored their gene expression profile and phenotype. The rescued CAR-T cells had increased antitumor function in mouse xenograft models of human leukemia and osteosarcoma. Although ceasing CAR signaling for up to 3 weeks provided the greatest benefit, even brief inhibition of CAR signaling for 3 days prior to in vivo administration significantly improved anti-leukemia activity of CAR-T cells. Moreover, Weber *et al.* demonstrated the benefit of in vivo pulsatile inhibition of CAR signaling: An intermittent 3-days ON/4-days OFF regimen produced superior leukemia control compared with conventional, always-ON CAR-T cells. This observation is surprising because suppressing

Reinvigorating engineered T cells

Chimeric antigen receptor (CAR)-expressing T cells are engineered to recognize antigens expressed by tumor cells. Upon chronic stimulation of the CAR, T cells can become functionally exhausted. This can be reversed by temporary cessation of CAR signaling, which induces enhancer of zeste homolog 2 (EZH2)-dependent chromatin remodeling.



CAR-mediated target cell killing could potentially facilitate tumor growth during the OFF period. Given intermittent signal-starvation-enhanced tumor control even in T cells expressing a CAR without tonic signaling, functional reinvigoration induced by interrupting constant antigenic stimulation may have additional but unidentified benefits.

Weber *et al.* revealed that the epigenetic changes (mostly transcriptional silencing) observed in hypofunctional CAR-T cells is primarily associated with histone methylation, rather than the more stable DNA methylation associated with “conventionally” exhausted T cells (9). The mechanism of these histone methylation changes is unknown, but they were reversible with “signal starvation” and required the activity of the histone methyltransferase, enhancer of

zeste homolog 2 (EZH2). This may indicate that exhaustion resulting from hyperactive CAR signaling may differ from conventional exhaustion associated with overstimulated TCR and negative costimulation with immune checkpoints and inhibitory cytokines. Whether the authors’ approach can be more broadly applied to tumor- and virus-specific T cells that have become exhausted in situ, or whether it will be used to prevent and reverse dysfunction primarily in CAR-T cells should be investigated in clinical studies. It will also be important to ascertain if reversal occurs even in the tumor microenvironment, where inhibitory checkpoint and cytokine signals are abundant.

The study by Weber *et al.* has important immediate implications for CAR-T cell therapy, providing a rationale for intermittently inhibiting CAR-T cell signaling. For example, licensed drugs such as dasatinib that interrupt the tyrosine kinase pathways of CAR signaling could be given as pulses after CAR-T cell administration. Preclinical studies showed the feasibility of this approach and demonstrated that in vivo administration of dasatinib would mitigate excessive CAR-T cell-mediated inflammation (10, 11). Any ON/OFF schedule must, however, avoid excessive periods of unopposed tumor growth or waves of inflammatory cytokines similar to the immune reconstitution syndrome observed in patients with some chronic infections, such as HIV and tuberculosis, where reinvigoration of T cell immunity may cause lethal hyperinflammation. These concerns may be addressable by alternating inhibition of different CAR-T cell populations using orthogonal systems of conditional CAR signaling. If T cell exhaustion due to tonic CAR signaling can indeed be avoided in cancer patients merely by inducing pulsatile CAR activation, accelerated development of optimized CAR-T cells and substantive benefits to their in vivo potency are anticipated. ■

REFERENCES AND NOTES

1. C. H. June, M. Sadelain, *N. Engl. J. Med.* **379**, 64 (2018).
2. C. M. Bollard, H. E. Heslop, *Blood* **127**, 3331 (2016).
3. S. Guedan *et al.*, *Annu. Rev. Immunol.* **37**, 145 (2019).
4. L. M. McLane *et al.*, *Annu. Rev. Immunol.* **37**, 457 (2019).
5. M. Hashimoto *et al.*, *Annu. Rev. Med.* **69**, 301 (2018).
6. E. W. Weber *et al.*, *Science* **371**, abal786 (2021).
7. D. Gomes-Silva *et al.*, *Cell Rep.* **21**, 17 (2017).
8. A. H. Long *et al.*, *Nat. Med.* **21**, 581 (2015).
9. H. E. Ghoneim *et al.*, *Cell* **170**, 142 (2017).
10. E. W. Weber *et al.*, *Blood Adv.* **3**, 711 (2019).
11. K. Mestermann *et al.*, *Sci. Transl. Med.* **11**, eaau5907 (2019).

ACKNOWLEDGMENTS

M.K.B. has equity or advisory board interests in Allovir, Allogene, Marker Therapeutics, Tessa Therapeutics, Walking Fish Therapeutics, Abintus, Memgen Kuur, and Poseida Therapeutics. M.M. has licensing or advisory board interests with Fate Therapeutics, Allogene, and Xenetic Biosciences. M.K.B. and M.M. receive research support from NIH/NCI P50 CA7019-19, LLS 126752, and SU2C-AACR-DT-29-19.

10.1126/science.abh0583

POLICY FORUM

EDUCATION

India's new National Education Policy: Evidence and challenges

Gains in schooling must be translated into gains in learning

By **Karthik Muralidharan¹** and **Abhijeet Singh²**

The global expansion of schooling in the past three decades is unprecedented: Primary school enrollment is near-universal, expected years of schooling have risen rapidly, and the number of children out of school has fallen sharply. Yet the greatest challenge for the global education system, a “learning crisis” per the World Bank, is that these gains in schooling are not translating into commensurate gains in learning outcomes. This crisis is well exemplified by India, which has the largest education system in the world. Over 95% of children aged 6 to 14 years are in school, but nearly half of students in grade 5 in rural areas cannot read at a grade 2 level, and less than one-third can do basic division (1). India's new National Education Policy (NEP) of 2020 (the first major revision since 1986) recognizes the centrality of achieving universal foundational literacy and numeracy. Whether India succeeds in this goal matters intrinsically through its impact on over 200 million children and will also have lessons for other low- and middle-income countries. We review the NEP's discussion of school education in light of accumulated research evidence that may be relevant to successfully implementing this ambitious goal.

GOVERNANCE AND PEDAGOGY

India has made tremendous progress on access to schooling since the 1990s. Yet multiple nationally representative datasets suggest that learning levels have remained largely flat over the past 15 years. A large body of evidence has shown that increasing “business as usual” expenditure on education is only weakly correlated with improvement in learning (2). Two key constraints that limit the translation of spending (of time and money) into outcomes are

weaknesses in governance and pedagogy.

Governance challenges are exemplified by high rates of teacher absence in public schools, with nearly one in four teachers absent at the time of surprise visits (3). Even when teachers are present, instructional time is low for a variety of reasons, including large amounts of administrative paperwork.

Further, teacher recognition for performance and sanctions for nonperformance are low. Studies in India and elsewhere have shown that even modest amounts of performance-linked bonus pay for teachers can improve student learning in a cost-effective way (4). By contrast, unconditional increases in teacher pay (the largest component of education budgets) have no impact on student learning (4, 5). Overall, improving governance and management in public schools may be a much more cost-effective way of improving student learning than simply expanding education spending along default patterns.

An even greater challenge in translating school attendance into learning outcomes may be weaknesses in pedagogy. Even motivated teachers primarily focus on completing the textbook, without recognizing the mismatch between the academic standards of the textbook and student learning levels. The rapid expansion of school enrollment has brought tens of millions of first-generation learners into the formal education system who lack instructional support at home and often fall behind grade-appropriate curricular standards. The mismatch is clearly illustrated in the figure, which presents the levels and dispersion of student achievement in mathematics in a sample of students from public middle schools in Delhi (6). There are three points to note about this figure: (i) The vast majority of students are below curricular standards (represented by the blue line of equality), with the average grade 6 student 2.5 years behind; (ii) the average rate of learning progress is much flatter than that envisaged by the curricular standards, resulting in widening

learning gaps at higher grades; (iii) there is enormous variation in learning levels of students in the same grade, spanning five to six grade levels in all grades.

The figure captures many features that we think are central to understanding the Indian education system. It suggests a curriculum that targets the top of the achievement distribution and moves much faster than the actual achievement level of students. Coupled with social promotion—grade retention is forbidden by law until grade 8—this leads to student achievement being widely dispersed within the same grade and most students receiving instruction that they are not academically prepared for. Similar patterns likely exist in many other developing countries (6).

The figure may also help explain why increased expenditures on items such as teacher salaries and school infrastructure may have little impact on learning. Students, having fallen so far behind the curriculum, may not gain much from the default of textbook-linked instruction. By contrast, pedagogical interventions that target instruction at the level of students' academic preparation can be highly effective (6–8).

The figure also highlights the stark inequality in Indian education. The true inequality is likely even greater because the figure does not reflect the large number of students in private schools. A comparison of data from two Indian states to countries included in an international learning assessment found that learning inequality in India is second only to South Africa (9). Thus, although the academically strongest Indian students are internationally competitive, with many ultimately achieving world-renowned success, most Indian children fail to acquire even basic skills at the end of their schooling.

To better understand the Indian education system, it is useful to recognize that education systems have historically served two very different purposes: (i) to impart knowledge and skills (a “human development” role) and (ii) to assess, classify, and select students for higher education and skill-intensive occupations (a “sorting and selection” role). The Indian education system primarily serves as a “sorting and selection” or a “filtration” system rather than a “human development” system. The system focuses primarily on setting high standards for competitive exams to identify those who are talented enough to meet those standards, but it ends up neglecting the vast majority of students who do not. Thus, a fundamental challenge for Indian education policy is to reorient the education system from one focused on sorting and identifying talented students to one

¹Department of Economics, University of California San Diego, La Jolla, CA, USA. ²Stockholm School of Economics, Stockholm, Sweden. Email: kamuralidharan@ucsd.edu

that is focused on human development that can improve learning for all.

RESEARCH INTO POLICY

The NEP, released in 2020, does an excellent job of reflecting key insights from research. Three points are especially noteworthy.

First, and most important, is the centrality accorded to universal foundational literacy and numeracy, which the NEP calls an “urgent and necessary prerequisite for learning.” This represents a substantial shift in the definition of education “quality” from inputs and expenditure to actual learning outcomes. Relatedly, the NEP recognizes the importance of early childhood care and education and brings preschool education into the scope of national education policy alongside school education. The NEP’s focus on stronger and universal preschool education is consistent with global recognition of the importance of “the early years” in developing cognitive and socioemotional skills.

Second, consistent with the evidence, the NEP aims to strengthen teacher effectiveness through a combination of improving their skills, reducing extraneous demands on their time, and rewarding performance. Notably, the NEP highlights the need for “a robust merit-based structure of tenure, promotion, and salary structure.” This is a meaningful departure from the status quo that does not reward good performance. If implemented well, improving teacher motivation and effort can be a force multiplier for the effectiveness of other input-based spending. School inputs on their own do not seem to translate into learning gains (2), but inputs can be highly effective when teachers and principals are motivated to improve learning outcomes (10).

Third, the NEP recognizes that improving school effectiveness may require changes to how schools are organized and managed. Large-scale school construction in the 1990s played an important role in promoting universal school access by providing a school in every habitation. However, as of 2016, over 417,000 government primary schools (~40% of schools) had fewer than 50 students across grades 1 to 5 (11). Small and spread-out schools present challenges for governance (by making supervision difficult), pedagogy (by requiring teachers to simultaneously teach students in multiple grades), and

infrastructure quality (by being too small for libraries and computer laboratories), as well as cost-effectiveness. The NEP, therefore, recommends investing in larger school complexes and also recognizes the importance of school management, emphasizing the need for customized school development plans to anchor a process of continuous school improvement. Given large improvements in rural road construction, it will be viable to provide buses or other transport to ensure universal school access for all children while also obtaining the benefits of larger-scale schools.

dent achievement of an ambitious reform that aimed to improve school management, largely through the type of school development plans that are recommended in the NEP (12). Yet, this model is perceived to be successful and has been scaled up to over 600,000 schools nationally (and aims to reach 1.6 million schools). Our work suggests that this perception is based primarily on completion of paperwork (such as school assessments and improvement plans), even though there was no change in management, pedagogy, or learning outcomes.

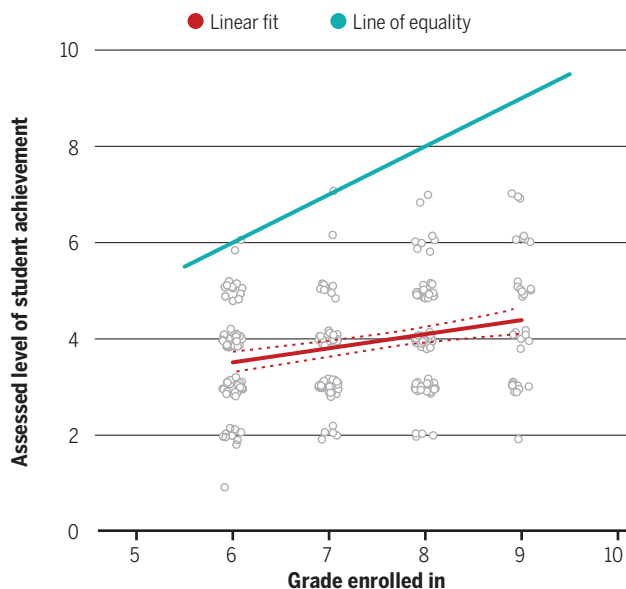
The second example illustrates how even measuring learning outcomes accurately is challenging. The state of Madhya Pradesh administers an annual state-level standardized assessment to all children in public schools from grades 1 to 8. This has been declared a national “best practice” and the NEP recommends a similar assessment for students in all schools in grades 3, 5, and 8. Yet, an independent audit that administered the same test questions to the same students a few weeks after the official tests showed that levels of student achievement are severely overstated in official data (13). The audit found that a large fraction of students did not possess even basic skills even though most of these students were shown as having passed the test.

In light of such challenges, we highlight three key principles that may increase the likelihood of success. The first is measurement. India’s success in achieving universal enrollment shows that the system is capable of delivering on well-defined goals that are easily measured. A similar approach needs to be implemented for delivering universal foundational literacy and numeracy. Although the challenge of data integrity is real, one reason for optimism is that there is evidence that using technology-based independent testing sharply reduced the extent to which data on learning was inflated (13). Thus, investing in independent ongoing measurement of learning outcomes in representative samples to set goals and monitor progress will be a foundational investment.

The second key principle is ongoing evaluations of policy and program effectiveness. An important lesson from the past two decades of research on education is that many commonly advocated interventions for improving education (such as in-

Achievement versus curricular standards

The estimated level of student achievement (determined by a computer-aided instruction program) in mathematics in public middle schools in Delhi is plotted against the grade in which students are actually enrolled. See (6) for details and data. Most students are below curricular standards (line of equality), average progress in learning is flatter than curricular standards, and there is substantial variation in achievement.



IMPLEMENTATION CHALLENGES

Although the NEP is an excellent document that reflects research and evidence, delivering on its promise will require sustained attention to implementation. The glaring gaps between the high quality of policy and program design on one hand, and the low quality of implementation on the other, are widely recognized in India across many dimensions of public policy.

Preliminary findings from two of our recent projects illustrate this challenge in relation to policy recommendations in the NEP. First, in a large-scale randomized controlled trial covering over 5000 schools in the state of Madhya Pradesh, we found no notable effects on school functioning or stu-

creasing teacher salaries, providing school grants, or giving out free textbooks) may have very little impact on learning outcomes, whereas other interventions (such as teaching at the right level) may be highly effective. Even in the same class of policies, different interventions may have widely varying effectiveness; for instance, in the case of education technology, the impact of providing hardware alone is zero or even negative, but personalized adaptive learning programs have been found to be highly effective (6, 7). Yet, use of rigorous, experimental evidence in education policy-making remains more an exception than the rule. Disciplining interventions under the NEP with high-quality evaluations can accelerate the scaling up of effective programs as well as course corrections of ineffective ones.

The third key principle is cost-effectiveness. Evidence has shown pronounced variation in the cost-effectiveness of education interventions, with many expensive policies having no impact and inexpensive ones being very effective. Given limited resources and competing demands on them, cost-effectiveness is not only an economic consideration but also a moral one. The World Bank and the UK Foreign and Commonwealth Development Office recently synthesized a large body of evidence on the most cost-effective education interventions (14). India would do well to heed these recommendations (suitably modified to its context) when allocating scarce public resources.

CONFRONTING COVID-19

Education has been sharply disrupted around India and the world by the COVID-19 shock. Public schools in India have been mostly closed and are likely to remain so for the entire academic year. This presents one major threat and two opportunities.

The threat is that the learning crisis will worsen. Children who have missed a year of school—especially those without educated parents—are likely to have regressed in their learning and suffer long-term learning losses. Thus, the challenges (see the figure) are likely to have worsened, making it imperative to provide high-quality supplementary instruction when schools reopen, including perhaps through reducing holidays and vacation days.

Yet, there may also be two important longer-term opportunities. The first is the rapid acceleration in the use of education technology by both households and the government. Given evidence of strong positive effects of personalized instruction, the

widespread adoption of education technology may help accelerate the NEP's stated goal of reducing the digital divide and leveraging potential benefits of technology for education, such as opportunities to increase student engagement and personalize instruction to individual student needs.

The second is increasing engagement with parents and families. Households play a critical role in education. Yet, education policy has mostly focused on school-based interventions, reflecting a belief that it is more feasible to improve schools than

“Although the NEP is an excellent document that reflects research and evidence, delivering on its promise will require sustained attention to implementation.”

to intervene in households at scale. The COVID-19 crisis and the resulting growth in the use of mobile phones for engaging children have sharply increased educators' engagement with parents, with approaches ranging from text-message reminders to check their child's homework to parent groups for peer coaching and motivation. Work is under way to evaluate the impacts of these promising approaches. The benefits of increased parental engagement may persist even after schools reopen.

NOTHING INEVITABLE

Effective reform will require a confluence of ideas, interests, institutions, and implementation. Our focus has been on the ideas of the NEP and the extent to which they are supported, or may be refined by, research evidence. The NEP also pays attention to institutional infrastructure needed to deliver on this vision and acknowledges the centrality of implementation. However, both the NEP and our discussion are silent on the interests, specifically on political and bureaucratic constraints. We remain optimistic that substantial improvements are possible. In particular, backing the intent of the NEP with a commitment to regular independent measurement and reporting of learning outcomes in a representative sample of all children—as envisaged by the NEP in setting up a quasi-independent national testing agency—may help to provide an institutionalized focus on learning to both political and bureaucratic leadership. The NEP's proposal to provide such information to parents directly, if implemented in easily accessible formats, may catalyze improvements in both public and private schools.

Such reforms are particularly urgent given India's demographic transition. In

many states, especially in South India, total fertility rates are already below replacement levels, and cohort sizes in primary schooling are shrinking. Thus, much of the country has already passed the peak of potential demographic dividend without having solved the learning crisis. Some large populous states in Northern India, such as Uttar Pradesh and Bihar, still have a window for intervention, but this window is shrinking. The one silver lining is that declining cohort sizes may increase resources per student in coming years, thus freeing up fiscal space for cost-effective investments.

There is nothing inevitable about low learning levels in Indian schools. Other developing countries, such as Vietnam, have been able to achieve substantially superior learning outcomes at very similar levels of per capita incomes. Research suggests that a key explanation is the greater productivity of Vietnam's schooling system, which focuses attention on ensuring that even the weakest students reach minimum standards of learning (15). The NEP provides an important opportunity to move Indian education from “sorting and selection” to “human development,” enabling every student to develop to their maximum potential. India, and the world, will be better off if this vision is realized in practice. ■

REFERENCES AND NOTES

1. Pratham, Annual Status of Education Report 2018, Pratham, New Delhi, 2019.
2. P. Glewwe, K. Muralidharan, “Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications” in *Handbook of the Economics of Education* (Elsevier, 2016), vol. 5, pp. 653–743.
3. K. Muralidharan, J. Das, A. Holla, A. Mohpal, *J. Public Econ.* **145**, 116 (2017).
4. K. Muralidharan, V. Sundararaman, *J. Polit. Econ.* **119**, 39 (2011).
5. J. de Ree, K. Muralidharan, M. Pradhan, H. Rogers, *Q. J. Econ.* **133**, 993 (2018).
6. K. Muralidharan, A. Singh, A. Ganimian, *Am. Econ. Rev.* **109**, 1426 (2019).
7. A. V. Banerjee, S. Cole, E. Duflo, L. Linden, *Q. J. Econ.* **122**, 1235 (2007).
8. A. Banerjee et al., *J. Econ. Perspect.* **31**, 73 (2017).
9. J. Das, T. Zajonc, *J. Dev. Econ.* **92**, 175 (2010).
10. I. Mbiti et al., *Q. J. Econ.* **134**, 1627 (2019).
11. G. G. Kingdon, *J. Dev. Stud.* **56**, 1795 (2020).
12. K. Muralidharan, A. Singh, “Improving Public Sector Management at Scale: Experimental Evidence on School Governance in India,” NBER Working Paper, 2020.
13. A. Singh, “Myths of Official Measurement: Auditing and Improving Administrative Data in Developing Countries,” Tech. Rep., RISE Programme, Oxford, 2020.
14. Global Education Evidence Advisory Panel, “Cost Effective Approaches to Improve Global Learning: What does recent evidence tell us are ‘Smart Buys’ for improving learning in low- and middle-income countries?” World Bank, Washington, DC, 2020.
15. A. Singh, *J. Eur. Econ. Assoc.* **18**, 1770 (2020).

10.1126/science.abf6655

NEWS FROM Science

Up-to-the-minute research and policy news you won't find in print

Visit us online to read all the news coverage that there just wasn't enough room to print in this issue.



ScienceMag.org/news



Brand visits a vault in Siberia where seeds are stored as a safeguard against future biodiversity loss.

BOOKS *et al.*

FILM

Stewart Brand's radical environmentalism

The counterculture icon continues to embrace an outside-the-box approach to the future

By W. Patrick McCray

Born in 1938, Stewart Brand studied systems biology at Stanford University before serving a stint in the army as a parachutist and photographer. In the 1960s, he was drawn into the nascent counterculture where beatniks rubbed shoulders and shared hot tubs with a younger cohort of hippies. After a psychedelic drug experience in 1966, Brand successfully lobbied NASA to release photographs taken from space of the entire planet. Such pictures, Brand claims, helped “blow away” the dark pessimism of the nuclear mushroom cloud that permeated 1960s popular culture.

In 1968, Brand created the *Whole Earth Catalog*, a counterculture periodical that was populated with articles and products designed to promote self-sufficiency and sustainability. Its runaway success enabled him to assume a decades-long role as a provocateur and cultural influencer.

Since the 1970s, Brand has catalyzed public debate about space settlements, personal computers, nanotechnology, the internet, and nuclear power. Central to all of these

is an abiding concern for environmental issues. Taken together, his activities reflect a talent for conceiving of a radically different future, helping build tools to make it happen, and then popularizing this vision (1).

The plenitude of Brand's projects presents a challenge for any biographer or filmmaker, but the documentary *We Are As Gods* provides a compelling introduction to his life. At the beginning of the film, Brand is compared to American icons ranging from Johnny Appleseed to P. T. Barnum. Each comparison captures a facet of his life, but, in the end, no single one suffices. The long arm of Brand's reach can be seen in the film's other voices, which include Doug Engelbart, Hunter Lovins, Steve Jobs, and Paul Ehrlich as well as Ken Kesey, Brian Eno, and members of the Grateful Dead.

Part biography and part meditation on the nature of time, *We Are As Gods* weaves together the disparate causes that Brand has championed over the past 60 years with his current fascination, the potential “de-extinction” of creatures such as the woolly mammoth and the American chestnut tree. The film also challenges viewers to rethink the stereotype that the hippie counterculture was “antitechnology” (2). In reality, young people a half-century ago success-

fully found ways to reconcile science and technology with an environmental sensibility and consumer hipness.

At the heart of all of Brand's activities is his profound desire to encourage people to see the world in new ways. The film's core idea as well as its title is drawn from the introductory pages of the *Whole Earth Catalog*: “We are as gods and might as well get good at it.” This message, an alloy of hope and hubris, revolves around the idea that the right tool, be it the psychedelic LSD or a clock built to last 10,000 years, can “concentrate consciousness,” alter our perception, and inspire new behavior.

The film's blend of enthusiasm and wariness is presented with another essential element—honesty. Brand, shown just as thin, rangy, and birdlike today as he was in the 1960s, is candid about a period of deep depression he experienced in the 1970s, brought about by a too-liberal regimen of recreational drugs combined with sudden celebrity and overwork. The emergence of personal computers, which Brand calls “a better drug,” and of online communities helped him recover, both professionally and personally.

Although doubters of de-extinction might disagree, many believe that the past loss of flora and fauna is, as the film says, a “tragedy that can be rewritten.” Much of the film concerns Brand's collaboration with geneticist George Church to bring back and then reintroduce the woolly mammoth to a region in the Siberian Arctic known as Pleistocene

We Are As Gods

David Alvarado and Jason Sussberg, directors
Structure Films, 2021. 94 minutes.

Park as a means of combating climate change. It is in this more recent effort that Brand's activities as a biologist, conservationist, and technologist are most tightly spliced together.

The filmmakers generously allow the viewer to draw their own conclusions

as to whether Brand is once again ahead of his time or blinded by techno-optimism.

Toward the end of the film, we see Brand in a greenhouse, surrounded by new shoots of American chestnut trees genetically altered to be blight-proof. As he places some in soil and waters them, he reflects on a dream he has had, in which the plants transform and, in time, become a forest. In the dream, he is flying over the forest, almost as a god. ■

REFERENCES AND NOTES

1. W. P. McCray, *The Visioneers: How a Group of Elite Scientists Pursued Space Colonies, Nanotechnologies, and a Limitless Future* (Princeton Univ. Press, 2012).
2. D. Kaiser, W. P. McCray, Eds., *Groovy Science: Knowledge, Innovation, and American Counterculture* (Univ. of Chicago Press, 2016).

The reviewer is at the Department of History, University of California, Santa Barbara, Santa Barbara, CA 93106, USA. Email: pmccray@history.ucsb.edu

10.1126/science.abh3991

SCIENCE LIVES

Stephen Hawking, celebrity scientist

A new biography considers the cosmologist's reputation while placing his work in context

By Declan Fahy

For decades, cosmologist Stephen Hawking was caught in a contradiction. In popular culture, he was portrayed as a pure mind roaming the cosmos to uncover fundamental truths of the Universe, the modern heir to Albert Einstein and Isaac Newton. In the physics community, he was respected as a productive theorist who made seminal contributions to black hole research, but many scientists considered his popular reputation to be ludicrously overblown. Veteran science writer Charles Seife seeks to resolve this contradiction in *Hawking Hawking*, the best biography yet published of the most famous scientist of recent decades. Seife presents Hawking as a complicated man and evaluates the cosmologist's scientific legacy, both of which became obscured by decades of self-promotion, marketing, and mythmaking.

The popular-culture image of Hawking arose largely as a result of the success of his 1988 cosmology book, *A Brief History of Time*, which became an unexpected nonfiction blockbuster, selling more than 10 million copies. Hawking published the book with Bantam Books, as he wanted to reach the largest possible audience and to earn money, in part to pay for his daughter's school fees.

As Seife recounts, Hawking was warned by a friend at Cambridge University Press, which had tried to acquire the rights to the book, that a trade publisher might highlight the scientist's physical condition to market the book. This observation proved astute: The cover of the book's US edition featured Hawking in his wheelchair, superimposed against a starry Universe, helping to fix Hawking's image in the public imagination as a symbol of disembodied scientific rationalism.

The reviewer is at the School of Communications, Dublin City University, Glasnevin, Dublin 9, Ireland, and is the author of *The New Celebrity Scientists: Out of the Lab and into the Limelight* (Rowman and Littlefield, 2015). Email: declan.fahy@dcu.ie

This image was solidified through endless repetition by uncritical journalists and the marketing of Hawking's subsequent books. But unlike those accounts, Seife's portrait in this unauthorized biography is often unflattering. Hawking is represented as neglectful and dismissive of his first wife, Jane, who bore most of the burden of caring for her husband after he was diagnosed with amyotrophic lateral sclerosis (ALS) at the age of 21. He comes across as having been reluctant to give due credit to his research collaborators. We learn that he erroneously accused (in print) two scientists of stealing an idea from his friend, physicist Andrei Linde, and lobbied (un-



Hawking makes a red carpet appearance before an event in 2015.

successfully) to the highest levels of the University of Cambridge to stop a student from pursuing a doctorate, because the proposed research topic would challenge his ideas. Far from floating in a cerebral realm, Hawking was actively engaged in the earthly business of protecting his intellectual capital.

As the author of several popular books on mathematics and physics, Seife is well positioned to determine Hawking's specific scientific contributions and to judge their quality and impact. He situates Hawking's work within the rich intellectual history of cosmology, which flourished in the 1960s and 1970s and has been referred to by

Hawking Hawking:
The Selling of a Scientific
Celebrity
Charles Seife
Basic Books, 2021. 400 pp.



physicist Kip Thorne as the “golden age of black holes.”

In this period, Hawking made what would become his signature contribution to cosmology. Contrary to the prevailing view at the time, he discovered that a black hole does not absorb everything in its vicinity. Instead, it emits a form of energy that would eventually be called Hawking radiation, the equation for which is inscribed on Hawking's tombstone in London's Westminster Abbey.

No lone theorist, Hawking collaborated with graduate students and physicists from around the world, and through his professional networks, he became a conduit between physicists in the East and West during the Cold War. Moreover, in Seife's evaluation, Hawking's research inspired a new generation of scientists and catalyzed the work of other physicists working on problems at the intersection of quantum theory and relativity.

The book's subtitle—*The Selling of a Scientific Celebrity*—suggests that it is focused on the construction of Hawking's stardom, but it is

more accurately described as a successful attempt to rescue the complicated scientist from fame's myriad distortions. Seife tells the story in reverse chronological order, starting with a description of Hawking's tombstone and ending with his birth, a structure that invites the reader to see the man beyond the flashbulbs. Yet the biography's main narrative is that of a fame-hungry physicist whose popularity grew over time, even as his greatest scientific achievements retreated further into the past. The book humanizes Hawking but reveals a tragic core to his celebrity. ■

10.1126/science.abg8058

CALL FOR PAPERS

Journal of Remote Sensing

The *Journal of Remote Sensing* is an online-only Open Access Science Partner Journal published in affiliation with **Aerospace Information Research Institute, Chinese Academy of Sciences (AIR-CAS)** and distributed by the **American Association for the Advancement of Science (AAAS)**. Like all partners participating in the Science Partner Journal program, the *Journal of Remote Sensing* is editorially independent from the *Science* family of journals and AIR-CAS is responsible for all content published in the journal. This journal covers multiple research areas that include theory, science, technology of remote sensing, and interdisciplinary research with earth science and information science. Particular topics of interest within the journal include radiative transfer modeling, biogeosciences remote sensing, remote sensing of energy, and more.

Submit your research to the *Journal of Remote Sensing* today!

Learn more at spj.sciencemag.org/remotesensing

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at spj.sciencemag.org



@SPJournals



@SPJournals

 OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED UNTIL JULY 2023

RESEARCH

IN SCIENCE JOURNALS

Edited by **Michael Funk**

Human activities—especially fishing—are disrupting marine life, like these Steller's sea lions, at an ever-increasing scale.



MARINE CONSERVATION

An ever-growing human footprint

Human activities are increasingly affecting the marine environment but understanding how much and in what ways is an extreme challenge given the vastness of this system. O'Hara *et al.* looked at a suite of human-induced stressors on >1000 marine species over the course of 13 years. They

found that species are experiencing increasing levels of stress over more than half of their ranges, with some species having an even higher proportion of their ranges affected. Fishing has the largest impact, but other stressors, such as climate change, are also important and growing. —SNV *Science*, this issue p. 84

NEUROGENOMICS

DNA repair within neurons

Humans have only a limited capacity to generate new neurons. These cells thus need to repair errors in the genome. To better understand this process, Reid *et al.* developed Repair-seq, a method to locate DNA repair within the genome of stem cell–derived neurons. DNA repair hotspots (DRHs) were more likely to occur within specific genomic features such as gene bodies as well as in genomic formations, open chromatin, and active regulatory regions. This method showed that repair was enriched

at sites involved in neuronal function and identity. Furthermore, proteomic data indicated that genes in DRHs are enriched in Alzheimer's disease and that DRHs are more active in aging. These observations link neuronal DNA repair to aging and neurodegeneration. —LMZ

Science, this issue p. 91

TRANSCRIPTION

Mediating transcription

The Mediator complex is recruited by transcription factors to all protein-coding genes in eukaryotes and helps to assemble the machinery necessary to

transcribe the gene. Abdella *et al.* present the cryo-electron microscopy structure of the human Mediator-bound preinitiation complex (Med-PIC). The structure shows how Mediator positions the long, flexible C-terminal domain of RNA polymerase II to be phosphorylated by the kinase CDK7, a crucial step for further processing of the RNA into a mature RNA. Most sites where transcription factors bind to Mediator are flexibly tethered to the complex, allowing the large Med-PIC to assemble at any gene. —DJ

Science, this issue p. 52

PALEOECOLOGY

The birth of modern rainforests

The origin of modern rainforests can be traced to the aftermath of the bolide impact at the end of the Cretaceous. Carvalho *et al.* used fossilized pollen and leaves to characterize the changes that took place in northern South American forests at this time (see the Perspective by Jacobs and Curran). They not only found changes in species composition but were also able to infer changes in forest structure. Extinctions were widespread, especially among gymnosperms.

Angiosperm taxa came to dominate the forests over the 6 million years of recovery, when the flora began to resemble that of modern lowland neotropical forest. The leaf data also imply that the forest canopy evolved from relatively open to closed and layered, leading to increased vertical stratification and a greater diversity of plant growth forms. —AMS

Science, this issue p. 63;
see also p. 28

QUANTUM OPTICS

A dissipative quantum gas of light

Our textbook understanding of quantum systems tends to come from modeling these systems isolated from the environment. However, an emerging focus is understanding how many-body quantum systems behave when interacting with their surroundings and how they subsequently become dissipative, or non-Hermitian, systems. Öztürk *et al.* formed a quantum condensate of light by trapping photons in an optical cavity, a system that is naturally dissipative. By altering the trapping conditions, they demonstrated that the system provides a powerful platform with which to explore the complex dynamics and phase transitions occurring in dissipative quantum systems. —ISO

Science, this issue p. 88

BIOSENSORS

A simple sweat test for cystic fibrosis

Cystic fibrosis is often diagnosed in infants using dried blood spot testing, but this method can have poor sensitivity and a high false-positive rate. Toward the goal of developing a noninvasive, simple test for cystic fibrosis, Ray *et al.* devised an adhesive microfluidic device, or “sweat sticker,” to capture and analyze sweat in real time with colorimetric readout. Elevated chloride concentrations in sweat are indicative of cystic fibrosis. Benchtop testing and validation

in patients with cystic fibrosis showed that smartphone imaging of sweat stickers adhered to the skin could monitor sweat chloride concentrations. Results support further testing of the sweat stickers in larger studies. —CC

Sci. Transl. Med. **13**, eabd8109 (2021).

HEART PHYSIOLOGY

Blocking the nuclear option for GRK5

Myocardial infarction results in pressure overload on the heart, which stimulates β -adrenergic signaling and the translocation of G protein-coupled receptor kinase 5 (GRK5) to the nucleus, where it promotes changes in gene expression that lead to hypertrophy. Coleman *et al.* investigated whether blocking the nuclear translocation of GRK5 could prevent the pathological signaling of this kinase. The N-terminus of GRK5 contains a sequence that is required for nuclear translocation, and mice expressing a peptide encompassing the GRK5 N-terminus developed less cardiac hypertrophy, fibrosis, and dysfunction after pressure overload. —WW

Sci. Signal. **14**, eabb5968 (2021).

CATALYSIS

Isolating and stabilizing boron

Oxidative dehydrogenation of propane can produce propene from shale gas and help to replace petroleum as a propene feedstock. Boron-based catalysts can have high selectivity to propene, but the water by-product can deactivate the catalyst by hydrolyzing boron. Zhou *et al.* synthesized boron-doped silicate zeolites containing isolated boron sites that were stable against hydrolysis. The catalyst could achieve one-pass propane conversions up to ~44% with selectivities for propene and >80% for ethene. They observed no deactivation after a 210-hour continuous test. —PDS

Science, this issue p. 76

IN OTHER JOURNALS

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



CHEMICAL ENGINEERING

Advancing chemical kinetic modeling

Quantitative prediction of the time evolution of chemical mixtures is the central problem of reactive chemical engineering. This is a challenging task because of the thousands of elementary chemical reactions that must be considered in chemical kinetic modeling. Usually, such models are developed and refined gradually over many years. Using broadband rotational spectroscopy, combined with automated ab initio transition state theory-based master equation calculations and high-level thermochemical parametrization, Zaleski *et al.* reveal an important role of a variety of radical substitution reactions in the flash pyrolysis of acetone that was previously omitted in the corresponding combustion mechanisms. Their unified combination of modeling, experiment, and theory is a promising approach in the development of comprehensive chemical kinetic models. —YS

J. Am. Chem. Soc. **143**, 3124 (2021).

TUBERCULOSIS

Cellulose as virulence factor

Infections with *Mycobacterium tuberculosis* commonly display antibiotic treatment failure and immune evasion. The pathogen itself has several physiological properties that contribute to these resistances. Chakraborty *et al.* show that several mycobacterial pathogens can produce cellulose-containing biofilms. Biofilms are formed by *M. tuberculosis* in vitro and within granulomas in lung samples of mice, macaques, and humans. When nebulized cellulase was administered to infected mice in combination with frontline drugs such as isoniazid and rifampicin, lung tissue damage was minimal compared with controls receiving heat-inactivated cellulase. Strains of the pathogen engineered to overexpress bacterial cellulases grew well but were sensitive to antibiotics, produced deficient biofilms, and had limited capacity to cause tissue damage. Mycobacterial

PHOTO: TIM PAGE/GETTY IMAGES



POLITICAL SCIENCE

Engaging more after war

Exposure to war has been linked to short-term increases in collective action and prosocial behavior, but the extent to which these effects persist across generations is unclear. Barceló examined records of bombing locations and civilians' provinces of residence during and after the Vietnam War and found that civilians who lived in areas that were heavily bombed were still more likely to volunteer and participate in social groups more than a quarter of a century later. These civilians were also more likely to support participatory values such as believing people should have more say in government decisions. These findings suggest that war may have long-lasting effects on civic participation. —TSR

Proc. Acad. Natl. Acad. Sci. U.S.A. **118**, e2015539118 (2021).

A bombing strike in Vietnam, 1968

cellulose may represent an Achilles heel to target for tuberculosis therapies. —CA

Nat. Commun. **12**, 1606 (2021).

NANOMATERIALS

Growing up with the twist

Twisted layers of two-dimensional materials can display a variety of electron-correlation effects, but their fabrication usually requires exfoliation, transfer, and alignment of the sheets. Yu *et al.* show that twisted grain boundaries can form when monolayer crystals of molybdenum disulfide grown with random orientations by chemical vapor deposition on a silica surface collide and coalesce. Electron microscopy and Raman and second-harmonic generation spectroscopy showed that misorientations at the shared grain boundary $>20^\circ$ were preserved and led to moiré twist angles between 20° and 55° . Bilayers with dimensions from 2 to 10 micrometers formed from the top layer climbing over the bottom layer and aligning with it through

kink nucleation and propagation. —PDS

ACS Nano **15**, 4504 (2021).

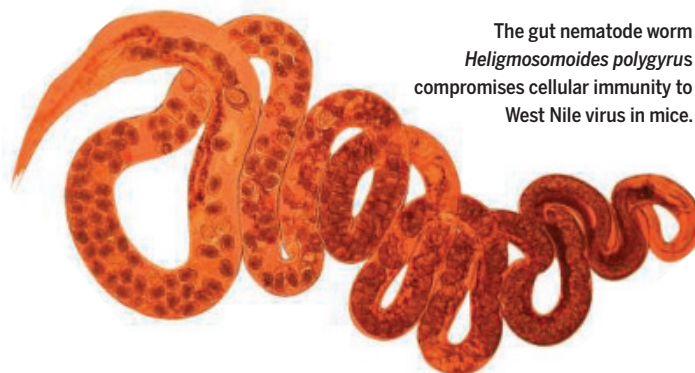
COINFECTION

Virus worms its way out of trouble

The immune system regularly encounters an array of bacteria, viruses, and multicellular organisms such as fungi and helminths. It remains unclear how the disparate responses mounted by the immune system are coordinated and the extent to which they affect one another. Desai *et al.* report that mice infected with the intestinal helminth *Heligmosomoides*

polygyrus bakeri were more prone to die during West Nile virus (WNV) coinfection. Coinfected animals had alterations in their gut mucosa that allowed translocation of commensal gut bacteria and induced failure of the anti-WNV CD8 T cell response. Helminth-derived succinate triggered tuft cells in the gut to produce type 2 cytokines. The cytokines were detected by intestinal epithelium and triggered gut barrier defects. Future studies are needed to tease out whether coinfections with other flavivirus–helminth combinations cause similarly detrimental immune synergies. —STS

Cell **184**, P1214 (2021).



The gut nematode worm *Heligmosomoides polygyrus* compromises cellular immunity to West Nile virus in mice.

NEUROSCIENCE

Framing anti-depressant action

Essentially, all antidepressant drugs increase the expression and signaling of brain-derived neurotrophic factor (BDNF). These drugs act through tyrosine kinase receptor 2 (TRKB), the receptor for BDNF, to regulate neuronal plasticity. Casarotto *et al.* investigated the potential interactions among TRKB, cholesterol, and antidepressants. The authors found that a dimer of TRKB forms a binding pocket, where several antidepressants from different drug classes bind with a low but physiologically meaningful affinity. This low-level binding depends on membrane cholesterol to stabilize the TRKB structure in synaptic membranes and thereby promotes BDNF signaling. Such direct binding to TRKB and promotion of BDNF-mediated plasticity may therefore be a common mechanism of action for antidepressant drugs. —PRS

Cell, **184**, 1299 (2021).

CANCER

T cell burn out

Tumors generally contain infiltrating immune cells such as T lymphocytes, but these cells are often dysfunctional and offer little or no antitumor benefit. Cancer immunotherapies, such as immune checkpoint inhibitors, aim to reactivate these inactive T cells, but this approach is often unsuccessful as well. Sanmamed *et al.* have discovered a distinct subset of intratumoral T cells that could help to explain these observations. This subset of cells, which the authors call “burned-out” effector T cells, are able to proliferate and accumulate in the microenvironment of human non-small-cell lung cancer but lack antitumor effects and contribute to therapeutic resistance. —YN

Cancer Discov. **10**, 1158/2159-8290.CD-20-0962 (2021).

REVIEW SUMMARY

ANIMAL CULTURE

The burgeoning reach of animal culture

Andrew Whiten

BACKGROUND: Culture—the inheritance of an array of behavioral traditions through social learning from others—was once thought specific to humans. Recent and accumulating evidence has shown that, to the contrary, culture permeates the lives of a great diversity of animals, with far-reaching implications for evolutionary biology, anthropology, and conservation. Early evidence for animal culture emerged in the mid-20th century in the discovery of regional birdsong dialects and the spread of provisioned sweet potato washing in Japanese monkeys. Stimulated by these discoveries, long-term studies of wild chimpanzees and orangutans later in the century revealed complex cultures composed of multiple traditions spanning diverse aspects of apes' lives, from tool use to social and sexual behavior.

In part through the accumulation of further long-term field studies, the present century

has witnessed an explosion in discoveries about social learning and culture, not only in primates but also in a rapidly growing range of animal species, from cetaceans to a diverse array of birds, fish, and even invertebrates.

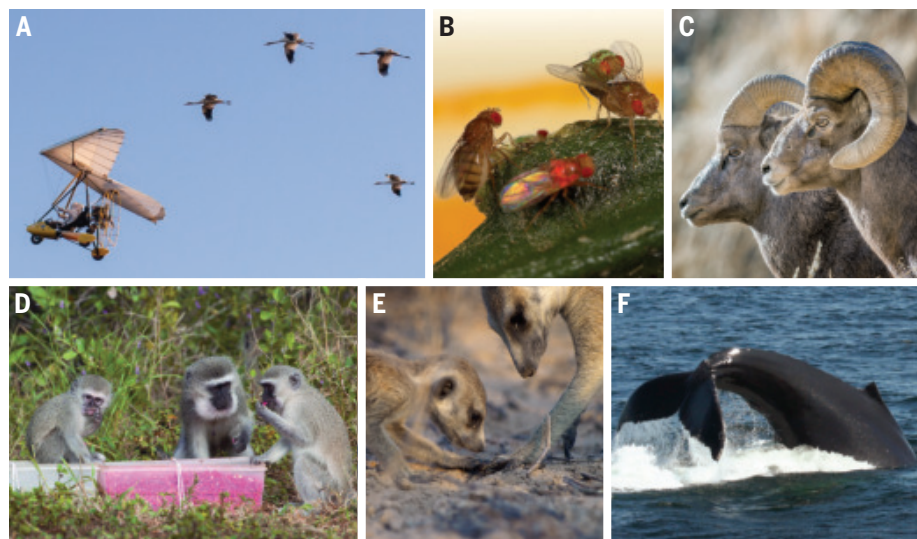
ADVANCES: Novel experimental designs have rigorously demonstrated the cultural transmission and spread of behavioral innovations introduced by researchers, both in the wild and in labs. New statistical methods have detected the signatures of behavioral innovations as they spread through social networks, identifying culture in species (e.g., whales) for which experiments are impractical. Through these and other methodological advances, the reach of cultural learning is now known to encompass an unexpected range of species, with surprising new discoveries extending even to insects, from bees to fruit flies.

The reach of culture has similarly been discovered to span diversity in behavioral domains, including foraging techniques, tool use, vocal communication, social customs and preferences for particular prey, migratory pathways, nesting sites, and mates. The revelation that cultural inheritance permeates many species' lives is increasingly recognized to have profound implications for evolutionary biology at large, because it provides a second form of inheritance that builds on the primary genetic inheritance system, facilitating cultural evolution. The two inheritance systems may generate rich interactive effects, as they have in humans.

A plethora of innovative experiments has further identified an array of cognitive processes involved in learning from others, ranging from simple and ubiquitous forms to specialized ones such as imitation and teaching. These forms of social learning have been shown to be further refined through a variety of selective biases, such as conforming to majorities or copying particularly skilled elders.

OUTLOOK: United Nations bodies operating under the aegis of international conventions have recently recognized the importance of all that has been discovered about animal cultures, for conservation policies and practices. Among sperm whales and chimpanzees, specific cultural entities, as opposed to genetically defined units, have been recognized as meriting conservation in their own right. This finding, in turn, urges a greater focus on understanding cultural phenomena in the wild. The task of rigorously identifying social learning has relied heavily on controlled experiments in captivity, but field experiments are increasingly carried out. These and other innovative methods to identify and trace animal cultures in the wild deserve to be developed and applied further to wild populations.

The wealth of methodological advances and empirical discoveries about animal cultures in the present century provides an exciting foundation from which to explore deeper questions. Do animal cultures evolve, cumulatively, as human cultures have done so impressively over past millennia? How profoundly does the lifetime reach of culture in animals' lives reshape our understanding of behavioral ecology and the fundamentals of evolution at large? How close are human and animal cultures now perceived to be, and where do the principal differences remain? ■



Diversity in cultural species and behavioral domains. (A) After filial imprinting on the costumed human pilot of a microflight aircraft, young cranes followed the flight path of this surrogate parent, adopting it as a traditional migratory route. (B) Female fruit flies (left) that witness a male marked with one of two colors mating (top right) later prefer to mate with similarly colored males. This behavior is further copied by others, initiating a tradition. (C) Bighorn sheep translocated to unfamiliar locations were initially sedentary, but spring migration and skill in reaching higher-altitude grazing grounds expanded over decades, implicating intergenerational cultural transmission. (D) Groups of vervet monkeys were trained to avoid bitter-tasting corn of one color and to prefer the other. Later, when offered these options with no distasteful additive, both naive infants and immigrating adult males adopted the experimentally created local group preference. (E) Young meerkats learn scorpion predation because adults initially supply live prey with stingers removed and later provide unmodified prey as the young meerkats mature. (F) A humpback whale innovation of slapping the sea surface to refine predation, known as "lobtail feeding," spread over two decades to create a new tradition in hundreds of other humpbacks. For reference citations, see the full article.

School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK. Email: a.whiten@st-andrews.ac.uk
Cite this article as A. Whiten, *Science* 372, eabe6514 (2021). DOI: 10.1126/science.abe6514

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abe6514>

RESEARCH ARTICLE SUMMARY

PROTEIN DESIGN

Designed proteins assemble antibodies into modular nanocages

Robby Divine, Ha V. Dang, George Ueda, Jorge A. Fallas, Ivan Vulovic, William Sheffler, Shally Saini, Yan Ting Zhao, Infencia Xavier Raj, Peter A. Morawski, Madeleine F. Jennewein, Leah J. Homad, Yu-Hsin Wan, Marti R. Tooley, Franziska Seeger, Ali Etemadi, Mitchell L. Fahning, James Lazarovits, Alex Roederer, Alexandra C. Walls, Lance Stewart, Mohammadali Mazloomi, Neil P. King, Daniel J. Campbell, Andrew T. McGuire, Leonidas Stamatatos, Hannele Ruohola-Baker, Julie Mathieu, David Veessler, David Baker*

INTRODUCTION: Antibodies that bind tightly to targets of interest play central roles in biological research and medicine. Clusters of antibodies, typically generated by fusing antibodies to polymers or genetically linking antibody fragments together, can enhance sig-

naling. Currently lacking are approaches for making antibody assemblies with a range of precisely specified architectures and valencies.

RATIONALE: We set out to computationally design proteins that assemble antibodies into

precise architectures with different valencies and symmetries. We developed an approach to designing proteins that position antibodies or Fc-fusions on the twofold symmetry axes of regular dihedral and polyhedral architectures. We hypothesized that such designs could robustly drive arbitrary antibodies into homogeneous and structurally well-defined nanocages and that such assemblies could have pronounced effects on cell signaling.

RESULTS: Antibody cage (AbC)-forming designs were created by rigidly fusing antibody constant domain-binding modules to cyclic oligomers through helical spacer domains such that the symmetry axes of the dimeric antibody and cyclic oligomer are at orientations that generate different dihedral or polyhedral (e.g., tetrahedral, octahedral, or icosahedral) architectures. The junction regions between the connected building blocks were optimized to fold to the designed structures. Synthetic genes encoding the designs were expressed in bacterial cultures; of 48 structurally characterized designs, eight assemblies matched the design models. Successful designs encompass D2 dihedral (three designs), T32 tetrahedral (two designs), O42 octahedral (one design), and I52 icosahedral (two designs) architectures; these contain 2, 6, 12, or 30 antibodies, respectively.

We investigated the effects of AbCs on cell signaling. AbCs formed with a death receptor-targeting antibody induced apoptosis of tumor cell lines that were unaffected by the soluble antibody or the native ligand. Angiopoietin pathway signaling, CD40 signaling, and T cell proliferation were all enhanced by assembling Fc-fusions or antibodies in AbCs. AbC formation also enhanced in vitro viral neutralization of a severe acute respiratory syndrome coronavirus 2 pseudovirus.

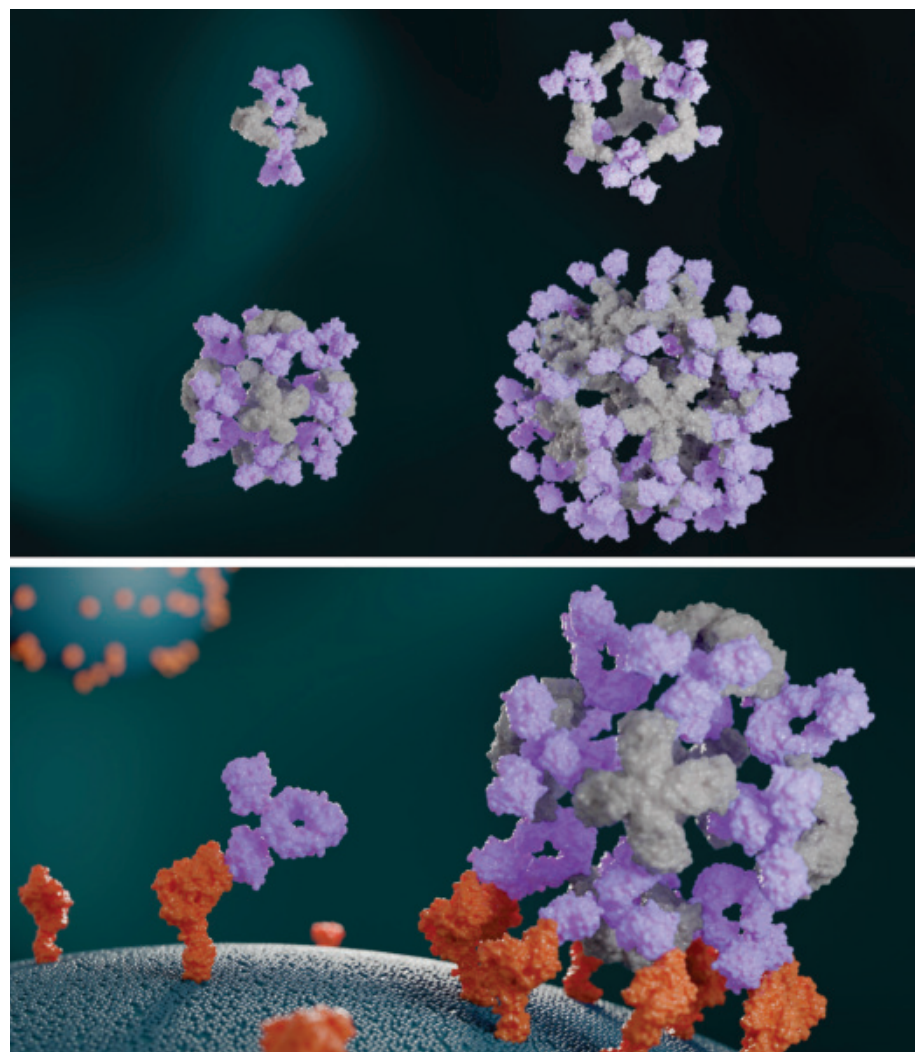
CONCLUSION: We have designed multiple antibody cage-forming proteins that precisely cluster any protein A-binding antibody into nanocages with controlled valency and geometry. AbCs can be formed with 2, 6, 12, or 30 antibodies simply by mixing the antibody with the corresponding designed protein, without the need for any covalent modification of the antibody. Incorporating receptor binding or virus-neutralizing antibodies into AbCs enhanced their biological activity across a range of cell systems. We expect that our rapid and robust approach for assembling antibodies into homogeneous and ordered nanocages without the need for covalent modification will have broad utility in research and medicine. ■

The list of author affiliations is available in the full article online.

*Corresponding author. Email: dabaker@uw.edu

Cite this article as R. Divine et al., *Science* 372, eabd9994 (2021). DOI: 10.1126/science.abd9994

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abd9994>



Designed proteins assemble antibodies into large symmetric architectures. Designed antibody-clustering proteins (light gray) assemble antibodies (purple) into diverse nanocage architectures (top). Antibody nanocages enhance cell signaling compared with free antibodies (bottom).

IMAGE: IAN HAYDON, INSTITUTE FOR PROTEIN DESIGN

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

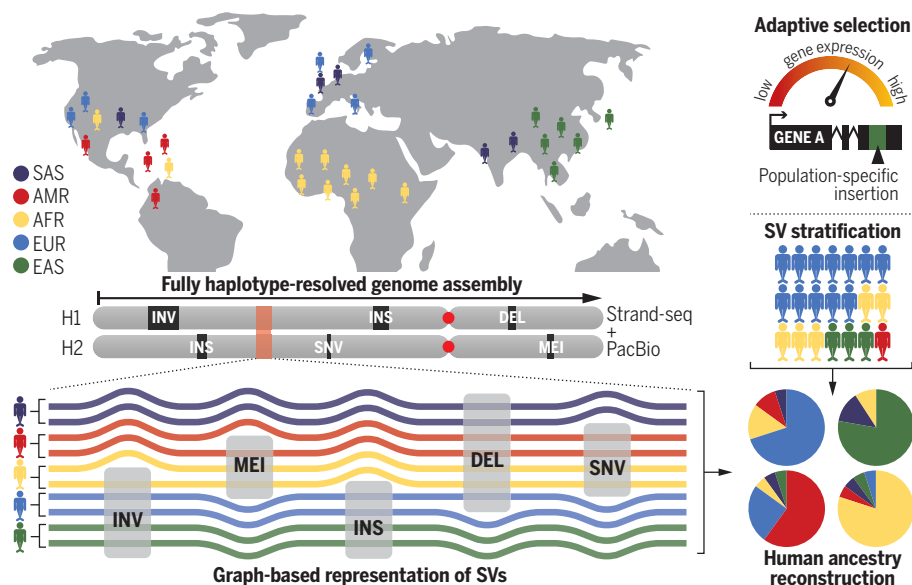
Haplotype-resolved diverse human genomes and integrated analysis of structural variation

Peter Ebert*, Peter A. Audano*, Qihui Zhu*, Bernardo Rodriguez-Martin*, David Porubsky, Marc Jan Bonder, Arvis Sulovari, Jana Ebler, Weichen Zhou, Rebecca Serra Mari, Feyza Yilmaz, Xuefang Zhao, PingHsun Hsieh, Joyce Lee, Sushant Kumar, Jiadong Lin, Tobias Rausch, Yu Chen, Jingwen Ren, Martin Santamarina, Wolfram Höps, Hufsa Ashraf, Nelson T. Chuang, Xiaofei Yang, Katherine M. Munson, Alexandra P. Lewis, Susan Fairley, Luke J. Tallon, Wayne E. Clarke, Anna O. Basile, Marta Byrska-Bishop, André Corvelo, Uday S. Evani, Tsung-Yu Lu, Mark J. P. Chaisson, Junjie Chen, Chong Li, Harrison Brand, Aaron M. Wenger, Maryam Ghareghani, William T. Harvey, Benjamin Raeder, Patrick Hasenfeld, Allison A. Regier, Haley J. Abel, Ira M. Hall, Paul Flicek, Oliver Stegle, Mark B. Gerstein, Jose M. C. Tubio, Zepeng Mu, Yang I. Li, Xinghua Shi, Alex R. Hastie, Kai Ye, Zechen Chong, Ashley D. Sanders, Michael C. Zody, Michael E. Talkowski, Ryan E. Mills, Scott E. Devine, Charles Lee††, Jan O. Korbel††, Tobias Marshall††, Evan E. Eichler††

INTRODUCTION: The characterization of the full spectrum of genetic variation is critical to understanding human health and disease. Recent technological advances have made it possible to survey genetic variants on the level of fully reconstructed haplotypes, leading to substantially improved sensitivity in detecting and characterizing large structural variants (SVs), including complex classes.

RATIONALE: We focused on comprehensive genetic variant discovery from a human diversity panel representing 25 human populations. We

leveraged a recently developed computational pipeline that combines long-read technology and single-cell template strand sequencing (Strand-seq) to generate fully phased diploid genome assemblies without guidance of a reference genome or use of parent-child trio information. Variant discovery from high-quality haplotype assemblies increases sensitivity and yields variants that are not only sequence resolved but also embedded in their genomic context, substantially improving genotyping in short-read sequenced cohorts and providing an assessment of their potential functional relevance.



Discovery and analysis of global human genetic diversity. Starting from a global panel of human diversity (top), we discovered structural variation from fully phased diploid genome assemblies (middle), resulting in a comprehensive catalog of sequence- and context-resolved variants. This facilitates integrative analysis and identification of new associations between variants and molecular phenotypes (bottom). SAS, South Asian; AMR, Admixed American; AFR, African; EUR, European; EAS, East Asian; INV, inversion; INS, insertion; DEL, deletion; MEI, mobile element insertion.

RESULTS: We generated fully phased genome assemblies for 35 individuals (32 unrelated and three children from parent-child trios). Genomes are highly contiguous [average minimum contig length needed to cover 50% of the genome: 26 million base pairs (Mbp)], accurate at the base-pair level (quality value > 40), correctly phased (average switch error rate 0.18%), and nearly complete compared with GRCh38 (median aligned contig coverage >95%). From the set of 64 unrelated haplotype assemblies, we identified 15.8 million single-nucleotide variants (SNVs), 2.3 million insertions/deletions (indels; 1 to 49 bp in length), 107,590 SVs (≥ 50 bp), 316 inversions, and 9453 nonreference mobile elements. The large fraction of African individuals in our study (11 of 35) enhances the discovery of previously unidentified variation (approximately twofold increase in discovery rate compared with non-Africans). Overall, ~42% of SVs are previously unidentified compared with recent long-read-based studies. Using orthogonal technologies, we validated most events and discovered ~35 structurally divergent regions per human genome (>50 kbp) not yet fully resolved with long-read genome assembly. We found that homology-mediated mechanisms of SV formation are twice as common as expected from previous reports that used short-read sequencing. We constructed a phylogeny of active L1 source elements and observed a correlation between evolutionary age and features such as the activity level, suggesting that younger elements contribute disproportionately to disease-causing variation. Transduction tracing allowed the identification of 54 active SVA retrotransposon source elements, which mobilize nonrepetitive sequences at their 5' and 3' ends. We genotyped up to 50,340 SVs into Illumina short-read data from the 1000 Genomes Project and identified variants associated with changes in gene expression, such as a 1069-bp SV near the gene *LIPI*, a locus that is associated with cardiac failure. We further identified 117 loci that show evidence for population stratification. These are candidates for local adaptation, such as a 4.0-kbp deletion of regulatory DNA *LCT* (lactase gene) among Europeans.

CONCLUSION: Fully reconstructed haplotype assemblies triple SV discovery when compared with short-read data and improve genotyping, leading to insights into SV mechanism of origin, evolutionary history, and disease association. ■

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡Corresponding author. Email: eee@gs.washington.edu

(E.E.); tobias.marshall@hhu.de (T.M.); jan.korbel@embl.org (J.O.K.); charles.lee@jax.org (C.L.)

Cite this article as P. Ebert, *Science* 372, eabf7117 (2021).

DOI: 10.1126/science.abf7117

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abf7117>

RESEARCH ARTICLE SUMMARY

CANCER

Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling

Evan W. Weber, Kevin R. Parker*, Elena Sotillo*, Rachel C. Lynn, Hima Anbunathan, John Lattin, Zinaida Good, Julia A. Belk, Bence Daniel, Dorota Klysz, Meena Malipatlolla, Peng Xu, Malek Bashti, Sabine Heitzeneder, Louai Labanieh, Panayiotis Vandriss, Robbie G. Majzner, Yanyan Qi, Katalin Sandor, Ling-Chun Chen, Snehit Prabhu, Andrew J. Gentles, Thomas J. Wandle, Ansuman T. Satpathy, Howard Y. Chang, Crystal L. Mackall†

INTRODUCTION: More than 50% of patients treated with chimeric antigen receptor (CAR)-T cells for B cell malignancies develop progressive disease after CAR therapy, and these agents have not demonstrated consistent activity against solid tumors. CAR-T cell efficacy is often limited by T cell exhaustion, wherein global transcriptional and epigenetic alterations drive overexpression of immune inhibitory proteins and diminish function. Current therapeutic approaches for targeting T cell exhaustion, including immune checkpoint inhibitors, do not remodel the exhaustion-associated epigenome, which has led some to conclude that T cell exhaustion is an epigenetically fixed state with limited potential for reversal.

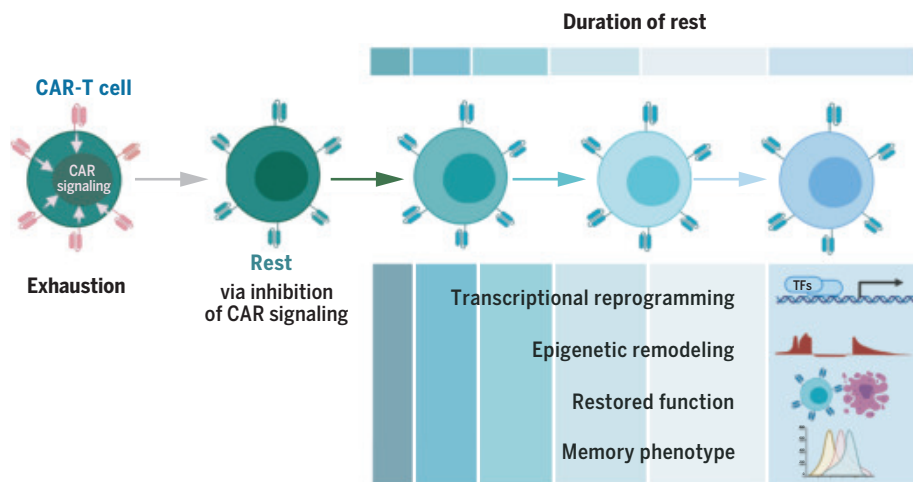
RATIONALE: We had previously demonstrated that human T cell exhaustion can be induced and maintained by tonic CAR signaling, and we therefore hypothesized that the inhibition of CAR signaling, or rest, could prevent and potentially reverse exhaustion in CAR-T cell populations. Rest was induced using a drug-regulatable degon system, whereby CAR expression and tonic CAR signaling were controlled

by the presence (ON) or absence (OFF) of a small molecule. In an alternative approach, CAR-T cells were rested using dasatinib, a clinically available tyrosine kinase inhibitor which potently and reversibly inhibits essential proximal CAR signaling kinases. Rest was induced either just before the manifestation of the exhaustion or after cells had acquired full phenotypic, functional, transcriptomic, and epigenetic hallmarks of exhaustion. Multiomics analyses and functional studies in vitro and in xenograft models were conducted to examine the effects of rest on tonically signaling CAR-T cells and nontonically signaling CAR-T cells exposed to high tumor burdens.

RESULTS: Tonically signaling CAR-T cells expanded ex vivo in the ON state manifested phenotypic, transcriptional, and epigenetic hallmarks of exhaustion, whereas CAR-T cells expanded exclusively in the OFF state or in the presence of dasatinib exhibited diminished tonic CAR signaling, which resulted in a memory-like phenotype and superior antitumor activity both in vitro and following adoptive transfer into

xenograft-bearing mice. The induction of rest in tonically signaling, pre-exhausted CAR-T cells redirected their cell fate away from exhaustion and toward a memory-like state. In CAR-T cells that had already acquired hallmark features of exhaustion, the induction of rest for as few as 4 days reversed the exhaustion phenotype and induced transcriptional reprogramming and global epigenetic remodeling to resemble healthy, nonexhausted controls. Further, exhausted CAR-T cells subjected to rest, including those previously exposed to 6 weeks of tonic CAR signaling, demonstrated restored antitumor functionality. The degree of functional reinvigoration was correlated with the duration of rest and was associated with decreased expression of the exhaustion-associated transcription factor TOX and increased expression of memory-associated transcription factors LEF1 and TCF1. These findings were not attributable to the outgrowth of a small subset of exhaustion-resistant cells because they were not associated with marked changes in proliferation, apoptosis, or clonal restriction of the T cell receptor (TCR) repertoire. Rather, functional reinvigoration was dependent on the activity of the histone methyltransferase EZH2, consistent with epigenetic remodeling in response to rest. Using dasatinib-insensitive liquid and solid-tumor xenograft models, CAR-T cells subjected to intermittent rest through the oscillation of CAR expression or pulsed dasatinib in vivo exhibited superior tumor control and enhanced survival compared with control mice. Single-cell analyses demonstrated that a single dasatinib pulse was sufficient to induce a memory-like phenotype and enhance functionality in exhausted tumor-infiltrating CAR-T cells

CONCLUSION: Inhibition of CAR signaling can enhance CAR-T cell fitness by preventing exhaustion, and it may have utility in clinical CAR-T cell manufacturing settings to enhance therapeutic efficacy. Moreover, in T cells that have acquired hallmark features of exhaustion, transient inhibition of CAR signaling, or rest, restores functionality and leads to global epigenetic remodeling, thereby challenging the notion that exhaustion is an epigenetically fixed state. These results predict that regulatable CAR-T cell regimens that are designed to incorporate periods of rest may exhibit superior efficacy compared with constitutive platforms, and they raise the prospect that targeting proximal CAR or TCR signaling kinases may represent an immunotherapeutic strategy for mitigating T cell exhaustion. ■



Giving exhausted CAR-T cells a break. Exhaustion in CAR-T cells is promoted and maintained by excessive CAR signaling (left). Transient inhibition of CAR signaling, or rest, reverses phenotypic and transcriptional hallmarks of exhaustion, remodels the exhaustion-associated epigenome, and restores antitumor functionality (right). The duration of rest correlates with the degree to which functionality is restored. TFs, transcription factors.

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: cmackall@stanford.edu

Cite this article as E. W. Weber *et al.*, *Science* 372, eaba1786 (2021). DOI: 10.1126/science.aba1786

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.aba1786>

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Gamma rhythm communication between entorhinal cortex and dentate gyrus neuronal assemblies

Antonio Fernández-Ruiz, Azahara Oliva, Marisol Soula, Florbela Rocha-Almeida, Gergo A. Nagy, Gonzalo Martín-Vazquez, György Buzsáki*

INTRODUCTION: Learning induces a dynamic reorganization of brain circuits but the neuronal mechanisms underlying this process are not well understood. Interregional gamma-frequency oscillations (~30 to 150 Hz) have been postulated as a mechanism to precisely coordinate upstream and downstream neuronal ensembles, for example, in the hippocampal system. The lateral (LEC) and medial (MEC) entorhinal cortex receive inputs from two distinct streams of cortical hierarchy (the “what” and the “where” pathways) and convey these neuronal messages to the hippocampus. However, the mechanisms by which such messages are packaged and integrated or segregated by hippocampal circuits had yet to be explored.

RATIONALE: Neuronal assemblies firing within gamma time frames in an upstream region can most effectively discharge their downstream partners. This gamma-time-scale organization appears essential for physiological functions because manipulations that impair precision timing of spikes in the hippocampus often affect behavior. How-

ever, direct support for distinct gamma-frequency communication in appropriate behavioral situations is missing. To bring physiological operations closer to behavior, we designed “spatial” and “object” learning tasks and examined the selective engagement of gamma-frequency communication between the MEC and LEC inputs and their target neuronal assemblies in the hippocampal dentate gyrus. We combined these correlational observations with optogenetic perturbation of gamma oscillations in LEC and MEC, respectively, to test their roles in pathway-specific neuronal communication and learning.

RESULTS: During spatial learning, fast gamma (100 to 150 Hz) oscillations synchronized MEC and dentate gyrus and entrained predominantly granule cells. During object learning, slow gamma (30 to 50 Hz) oscillations synchronized LEC and dentate gyrus and preferentially recruited mossy cells and CA3 pyramidal neurons, suggesting task-specific routing of MEC and LEC messages in the form of gamma-cycle-spike packets of selected

cell types. The low- and high-frequency gamma sub-bands were dominant in the outer and middle third of the dentate molecular layer, respectively, and their amplitude maxima were locked to different phases of theta oscillations.

Gamma frequency optogenetic perturbation of MEC and LEC led to learning impairments in a spatial and object learning task, respectively. In the same animals, the dentate layer-specific low- and high-frequency gamma sub-bands and spike-gamma LFP coupling were selectively reduced, coupled with deterioration of spatial and object-related firing of dentate neurons.

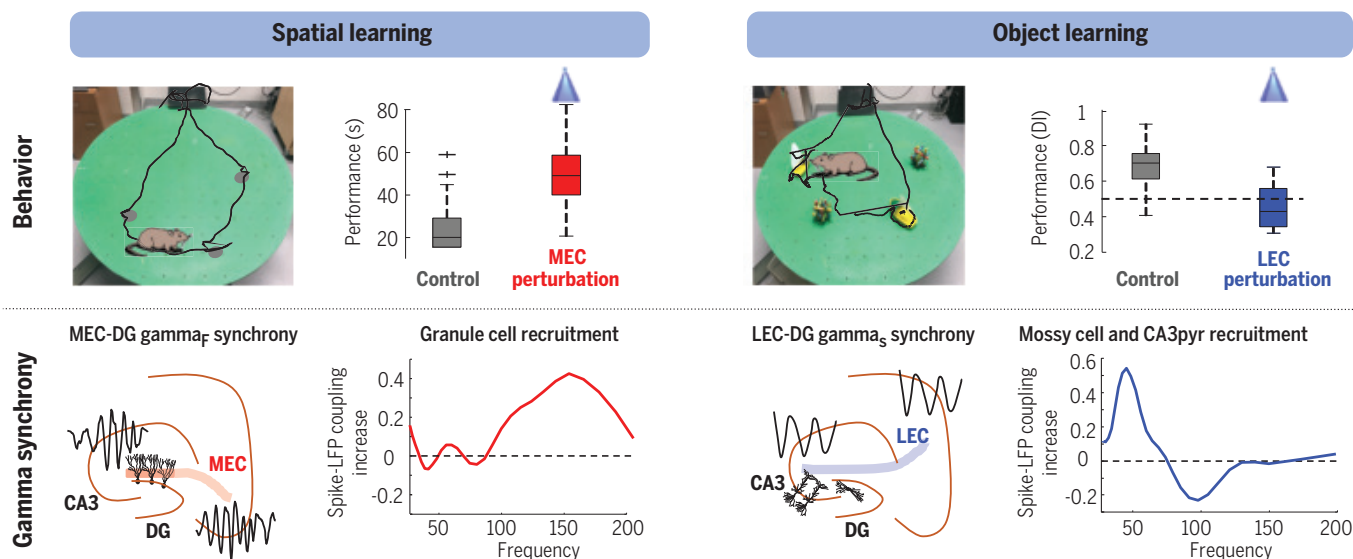
CONCLUSION: These findings demonstrate that distinct gamma-frequency-specific communication between MEC and LEC and hippocampal cell assemblies are critical for routing task-relevant information, and our selective gamma-band perturbation experiments suggest that they support specific aspects of learning. We hypothesize that sending neuronal messages by segregated gamma-frequency carriers allows a target “reader” area to disambiguate convergent inputs. In general, these results demonstrate that specific projected gamma patterns dynamically engage functionally related cell assemblies across brain regions in a task-specific manner. ■

The list of author affiliations is available in the full article online.

*Corresponding author. Email: gyorgy.buzsaki@nyumc.org (G.B.); af77@cornell.edu (A.F.-R.)

Cite this article as: A. Fernández-Ruiz et al., *Science* 372, eabf3119 (2021). DOI: 10.1126/science.abf3119

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abf3119>



Task-specific engagement and gamma-frequency coupling of distinct neuronal populations. First row: Impairment of spatial (left) and object (right) learning during gamma-frequency perturbation of MEC (left) and LEC (right). Second row: MEC and LEC project high-frequency (gamma_{HF}) and low-frequency (gamma_{LF}) gamma oscillations to DG, respectively, and entrain granule cells, mossy cells, and CA3 pyramidal neurons in a task-specific manner.

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Striatal dopamine mediates hallucination-like perception in mice

K. Schmack*, M. Bosc, T. Ott, J. F. Sturgill, A. Kepecs*

INTRODUCTION: Psychotic disorders such as schizophrenia impose enormous human, social, and economic burdens. The prognosis of psychotic disorders has not substantially improved over the past decades because our understanding of the underlying neurobiology has remained stagnant. Indeed, the subjective nature of hallucinations, a defining symptom of psychosis, presents an enduring challenge for their rigorous study in humans and translation to preclinical animal models. Here, we developed a cross-species computational psychiatry approach to directly relate human and rodent behavior and used this approach to study the neural basis of hallucination-like perception in mice.

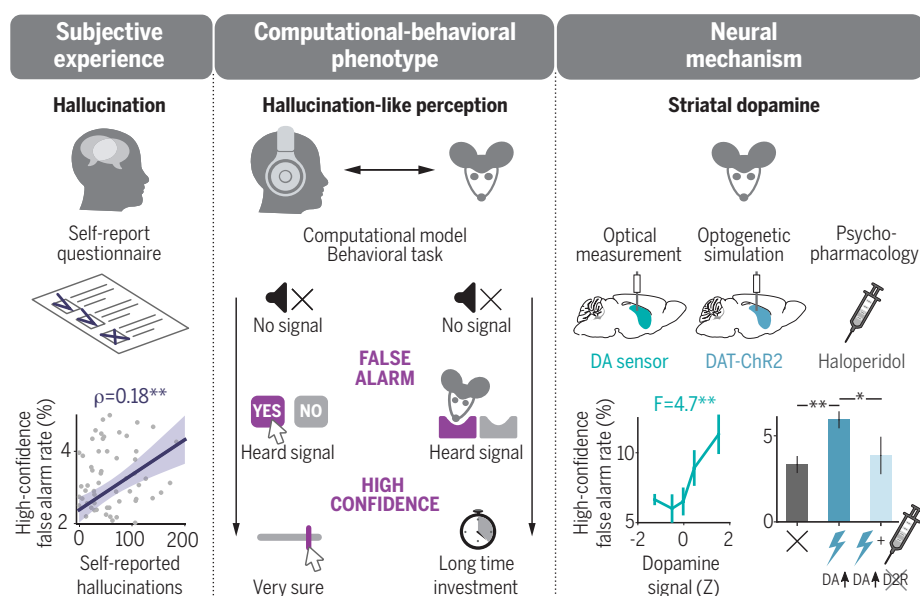
RATIONALE: Hallucinations are false percepts that are experienced with the same subjective confidence as “true” percepts. Similar false percepts can be quantitatively evaluated using a sensory detection task in which individuals report whether they heard a signal embedded in a background noise and indicate how confident

they are about their answer. Thus, we defined hallucination-like percepts as confident false alarms—that is, incorrect reports that a signal was present, which are reported with high confidence. We reasoned that such experimentally controlled hallucination-like percepts engage neural mechanisms shared with spontaneously experienced hallucinations in psychosis and can therefore serve as a translational model of psychotic symptoms. Because psychotic symptoms are thought to involve increased dopamine transmission in the striatum, we hypothesized that hallucination-like perception is mediated by increased striatal dopamine.

RESULTS: We set up analogous auditory detection tasks for humans and mice. Both humans and mice were presented with an auditory stimulus in which a tone signal was embedded in a noisy background on half of the trials. Humans pressed one of two buttons to report whether or not they heard a signal, whereas mice poked into one of two choice ports. Humans indicated how confident they were in

their report by positioning a cursor on a slider; mice expressed their confidence by investing variable time durations to earn a reward. In humans, hallucination-like percepts—high-confidence false alarms—were correlated with the tendency to experience spontaneous hallucinations, as quantified by a self-report questionnaire. In mice, hallucination-like percepts increased with two manipulations known to induce hallucinations in humans: administration of ketamine and the heightened expectation of hearing a signal. We then used genetically encoded dopamine sensors with fiber photometry to monitor dopamine dynamics in the striatum. We found that elevations in dopamine levels before stimulus onset predicted hallucination-like perception in both the ventral striatum and the tail of the striatum. We devised a computational model that explains the emergence of hallucination-like percepts as a consequence of faulty perceptual inference when prior expectations outweigh sensory evidence. Our model clarified how hallucination-like percepts can arise from fluctuations in two distinct types of expectations: reward expectations and perceptual expectations. In mice, dopamine fluctuations in the ventral striatum reflected reward expectations, whereas in the tail of the striatum they resembled perceptual expectations. We optogenetically boosted dopamine in the tail of the striatum and observed that increasing dopamine induced hallucination-like perception. This effect was rescued by the administration of haloperidol, an antipsychotic drug that blocks D2 dopamine receptors.

CONCLUSION: We established hallucination-like perception as a quantitative behavior in mice for modeling the subjective experience of a cardinal symptom of psychosis. We found that hallucination-like perception is mediated by dopamine elevations in the striatum and that this can be explained by encoding different kinds of expectations in distinct striatal subregions. These findings support the idea that hallucinations arise as faulty perceptual inferences due to elevated dopamine producing a bias in favor of prior expectations against current sensory evidence. Our results also yield circuit-level insights into the long-standing dopamine hypothesis of psychosis and provide a rigorous framework for dissecting the neural circuit mechanisms involved in hallucinations. We propose that this approach can guide the development of novel treatments for schizophrenia and other psychotic disorders. ■



Hallucination-like perception framework and striatal dopamine. In humans and mice, a computational-behavioral task models hallucinations as high-confidence false percepts. In humans, such hallucination-like percepts are correlated with self-reported hallucinations. In mice, hallucination-like percepts are mediated by striatal dopamine. Data are means \pm SEM. * $P < 0.05$, ** $P < 0.01$.

The list of author affiliations is available in the full article online.

*Corresponding author. Email: schmack@cshl.edu (K.S.);

akepecs@wustl.edu (A.K.)

Cite this article as K. Schmack et al., *Science* 372, eabf4740 (2021). DOI: 10.1126/science.abf4740

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abf4740>

RESEARCH ARTICLES

TRANSCRIPTION

Structure of the human Mediator-bound transcription preinitiation complex

R. Abdella^{1,2*}, A. Talyzina^{1,2*}, S. Chen^{1,2}, C. J. Inouye^{3,4,5,6}, R. Tjian^{3,4,5,6†}, Y. He^{1,2,7,8†}

Eukaryotic transcription requires the assembly of a multisubunit preinitiation complex (PIC) composed of RNA polymerase II (Pol II) and the general transcription factors. The coactivator Mediator is recruited by transcription factors, facilitates the assembly of the PIC, and stimulates phosphorylation of the Pol II C-terminal domain (CTD) by the TFIIF subunit CDK7. Here, we present the cryo-electron microscopy structure of the human Mediator-bound PIC at a resolution below 4 angstroms. Transcription factor binding sites within Mediator are primarily flexibly tethered to the tail module. CDK7 is stabilized by multiple contacts with Mediator. Two binding sites exist for the Pol II CTD, one between the head and middle modules of Mediator and the other in the active site of CDK7, providing structural evidence for Pol II CTD phosphorylation within the Mediator-bound PIC.

Transcription of all mRNA in eukaryotes is carried out by RNA polymerase II (Pol II) (1). Pol II cannot by itself locate the transcription start site, open a transcription bubble to expose the template strand, and transition to an elongation state. Instead, DNA-bound transcription factors position the coactivator complex Mediator to facilitate the assembly of the preinitiation complex (PIC), which consists of Pol II and the general transcription factors (GTFs) TFIIA, TFIIB, TFIID (TBP), TFIIE, TFIIF, and TFIIF (2). These GTFs help position Pol II to initiate transcription at the correct genomic locus, and TFIIF feeds DNA into the active site of Pol II, generating force against the TBP-TFIIB-TFIIA lobe to unwind the DNA and expose the template strand (3). The entire Mediator-bound PIC (Med-PIC) is 2.7 MDa in size, contains 56 polypeptides, and represents a particular challenge for structural characterization because of difficulties in obtaining and assembling these complexes, as well as to their inherent flexibility (4–9).

The largest subunit of Pol II, RPB1, contains a long, repetitive C-terminal domain (CTD) connected by a flexible linker region to the rest of the subunit (10). The CTD consists of 26 repeats in yeast and 52 repeats in humans of the consensus YSPTSPS heptamer sequence. Phos-

phorylation of the serine at position 5 (Ser⁵) by cyclin-dependent kinase 7 (CDK7) during transcription initiation leads to the recruitment of the 5' capping enzymes that are indispensable for growth (11). Neither the CTD nor the 80-residue linker is visible in structures of the PIC because of their mobility (12). CDK7 is part of the cyclin-activated kinase (CAK) module of TFIIF, together with cyclin-H and Mat1. Mounting evidence indicates that Mediator recruits TFIIF to the PIC and stimulates the CAK module's ability to phosphorylate Ser⁵ (13, 14).

Atomic models of Med-PICs are currently limited to yeast (yMed-PIC) (6–9). Mediator is divided into four modules: The head (MedHead) interacts with Pol II, the middle (MedMiddle) primarily serves a structural role, the tail (MedTail) serves as a hub for the binding of transcription factors, and the dissociable kinase module (MedKinase) is found at enhancers and prevents interaction with the PIC (15, 16). The existing high-resolution

models of yMed-PIC highlight differences in how MedHead interacts with Pol II, suggesting that the interface between Mediator and Pol II is not rigid (6, 7). MedHead is capable of binding the CTD, as shown in a cocrystal structure (17). Based on the location of the CTD modeled into the full yeast Mediator complex, the CTD also likely serves to stabilize the interface between MedHead and MedMiddle (8, 9). Functional and structural studies have identified a minimal core Mediator (cMed), devoid of both MedKinase and MedTail (18, 19). Structures of Mediator have poorly defined density for MedTail, leaving open the question of where most transcription factors bind (5, 6, 8). A second highly flexible part of Med-PIC is the CAK module of TFIIF. It has been localized to the end of the hook domain of Mediator, directly above core TFIIF (cTFIIF), to which it remains flexibly tethered (7, 8). However, the orientation of the individual subunits within this density has not yet been determined.

Here, we present the structure of the human Med-PIC assembled on a closed promoter DNA construct, with the TBP subunit of TFIID replacing the full TFIID complex. Human Mediator is held together by a central scaffold subunit, Med14, which forms two contact sites with MedTail. The precise orientation of the CAK module within Med-PIC is revealed, with clear density for the Pol II CTD in the active site. A second CTD binding site between MedHead and MedMiddle shows how Mediator positions the rest of the CTD for phosphorylation by CDK7. Many regions of Mediator that interact with transcription factors are flexibly tethered, facilitating its assembly. The structure also provides key insights into the conformational landscape of Mediator relative to the PIC.

Structural characterization of the human Med-PIC

The Med-PIC complex was assembled by extending our previous protocol for assembling

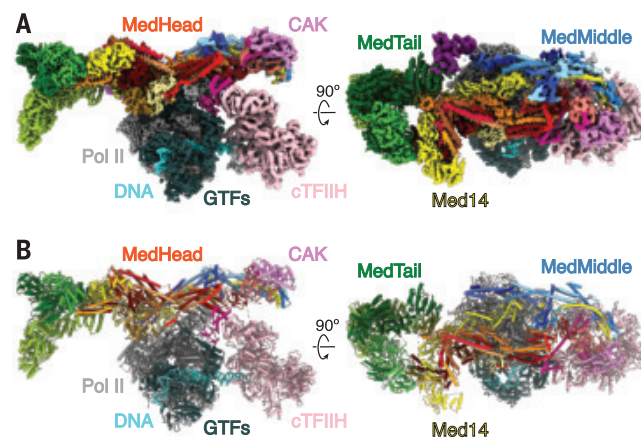


Fig. 1. Structure of the human Mediator-bound PIC. (A) Composite density map for Med-PIC built from the focused refinement maps for cPIC, cTFIIF, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1. (B) Model of the human Mediator-bound PIC. Gray, Pol II; dark gray, GTFs; pink, TFIIF core; salmon, CDK7; violet, cyclin H; medium violet red, Mat1; cyan, DNA; red shades, MedHead; blue shades, MedMiddle; yellow, Med14; green shades, MedTail.

¹Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA. ²Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL, USA.

³Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. ⁴Li Ka Shing Center for Biomedical and Health Sciences, University of California, Berkeley, Berkeley, CA, USA. ⁵CIRM Center of Excellence, University of California, Berkeley, Berkeley, CA, USA. ⁶Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA. ⁷Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, USA. ⁸Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Northwestern University, Chicago, IL, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: yuanhe@northwestern.edu (Y.H.); jmlim@berkeley.edu (R.T.)

the closed complex PIC from purified factors to accommodate the addition of Mediator (fig. S1A) (12). In contrast to previous protocols where factors were added in a stepwise manner, three subcomplexes—DNA-TBP-TFIIB-TFIIA, Pol II-TFIIF, and TFIIE-TFIIF-Mediator—were first assembled and were next incubated together. Negative-stain electron microscopy (EM) of assembled complexes indicated that a subset of particles contained all components of Med-PIC and that substantial conformational heterogeneity exists (fig. S1B).

A cryo-EM data set was collected, and two-dimensional (2D) classification in Relion-3 showed many classes representing the full complex (fig. S1C and table S1) (20). A subset of 156,383 particles refined to a resolution of 4.8 Å, but because of the high intrinsic flexibility of Med-PIC, distal regions—including MedMiddle, MedTail, and TFIIF—are mostly averaged out in the postprocessed map. Focused refinements on subcomplexes were used to improve the resolution of all portions of the density compared with the full complex (figs. S2 and S3). These regions were chosen either because the subcomplex behaves like a rigid body within the full complex, as is the case for the core PIC (cPIC), cTFIIF, MedHead, MedTail, and MedMiddle-CAK, or to center a region within the box to improve its resolution, as in the case of Med1 and Med14C. These refinements improved the resolution of the vast majority of MedTail, Med14, MedHead, and Pol II to 3.5 Å or better (fig. S3) and that of flexible regions, including Med1, MedMiddle-CAK, and cTFIIF, to 5.8, 6.5, and 7.1 Å, re-

spectively. These improvements allowed the building, refining, or flexible fitting of atomic models for nearly the entire complex (Fig. 1, A and B; figs. S4 to S7; table S2; and movie S1). Overall, the structure of Med-PIC is highly similar to previous human PIC complexes and yeast Med-PIC complexes (fig. S8) (7, 8, 12). The presence of Mediator does not cause substantial changes in the structures of Pol II or the GTFs, including TFIIB, TBP, TFIIA, TFIIE, and TFIIF. Med-PIC is compatible with the incorporation of TFIID because no clashes are observed upon superimposition of the structure of TFIID-TFIIA-DNA (fig. S8) (21).

Structure of human Mediator

The human Mediator complex within Med-PIC is divided into three modules that are held together by the central Med14 scaffold subunit (Fig. 2). MedMiddle closely resembles the structure of its yeast counterpart (7, 8). Homology models for the human MedMiddle subunits Med4, Med7, Med9, Med10, Med19, Med21, and Med31, based on the *Saccharomyces cerevisiae* ortholog structures, were built using the MedMiddle-CAK map (Fig. 2A and fig. S4). The N-terminal 200 residues of Med14 were modeled similarly. Additional density near the connector domain of MedMiddle could be assigned to Med26, a metazoan-specific subunit that has been shown to localize in this part of Mediator and interact with Med4, Med7, and Med19 (Fig. 2A) (15). The C terminus of Med26 is sufficient to interact with Mediator, strongly suggesting that the C terminus of Med26 is what can be seen, leaving the N

terminus flexibly attached. The N terminus has been shown to interact with the super elongation complex (SEC), which is responsible for the release of paused Pol II through phosphorylation of the Pol II CTD and SPT5 by CDK9 (22). Additional unmodeled density attributed to the N terminus of Med1 is located between the plank domain (Med4 and Med9) and MedTail subunit Med24 (Fig. 2A and fig. S9A). This is consistent with the location of Med1 shown in both yeast and humans previously (15). Density for the plank, Med1, and the N terminus of Med24 is considerably worse than surrounding areas, indicating that this portion of Mediator moves independently of MedMiddle and MedTail. Previous structures of yeast Med-PICs show interactions between Med9 and the foot domain of Pol II (fig. S10, A and B) (6, 7). In *Schizosaccharomyces pombe*, Med4 and Med9 also interact with Med1, but there is no change in the overall structure compared with *S. cerevisiae*, where Med1 was not included during complex assembly. The contact between Med9 and the foot domain of Pol II is broken in the human Med-PIC. Instead, Med9 is very close to RPB8, and the interactions between Med4 and Med9 with Med1 are retained (fig. S10C). These differences are likely driven by the presence of the larger MedTail in the human Med-PIC, which positions Med1 further away from the plank through interactions with Med24.

MedHead adopts a very similar structure to the yeast model except for the presence of the additional subunits Med27, Med28, Med29, and Med30 (Fig. 2B and fig. S5).

Fig. 2. Models and observed structural interactions for human Mediator. (A to C) Model and observed structural interaction diagram for MedMiddle and the CAK module of TFIIF (A), MedHead (B), and MedTail (C). The N terminus of the scaffold subunit Med14

extends the length of MedMiddle. Putative densities for Med1 and Med26 are shown and colored purple and dark blue, respectively. The C terminus of Med14 forms extensive interactions with MedHead. MedTail also interacts with the C terminus of Med14 but on the opposite face. Portions for the models that were built are shown in color; unmodeled sections are shown in gray. Known domains are shown with a light-to-dark (top-to-bottom) gradient. Everything else is shown with a dark-to-light gradient. Models are colored as in Fig. 1.

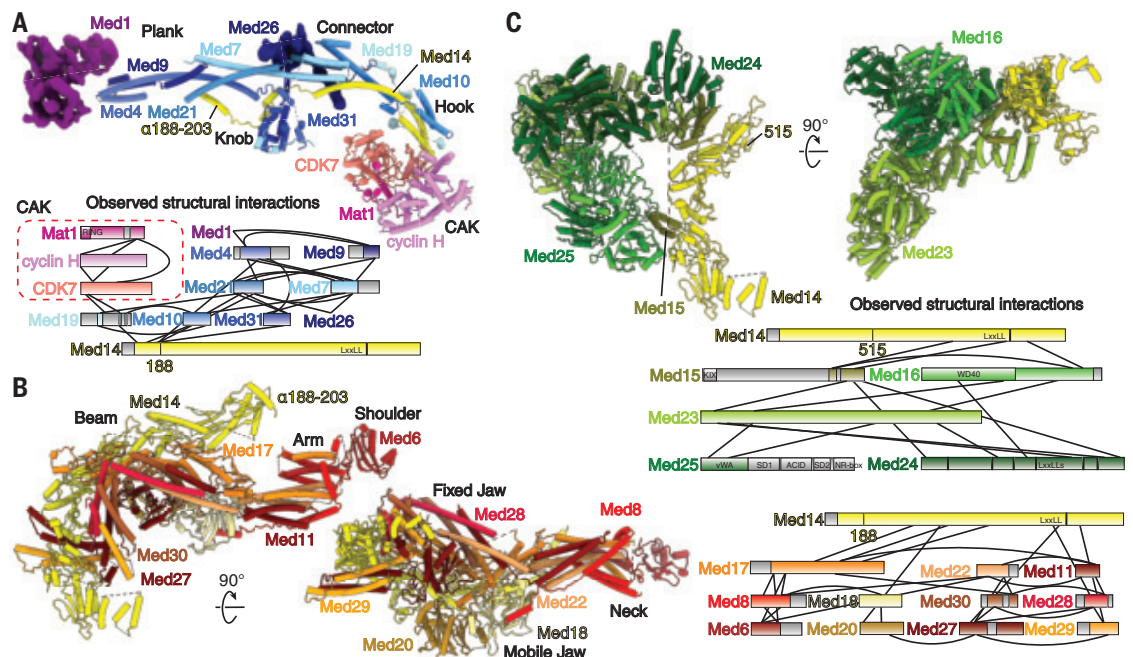
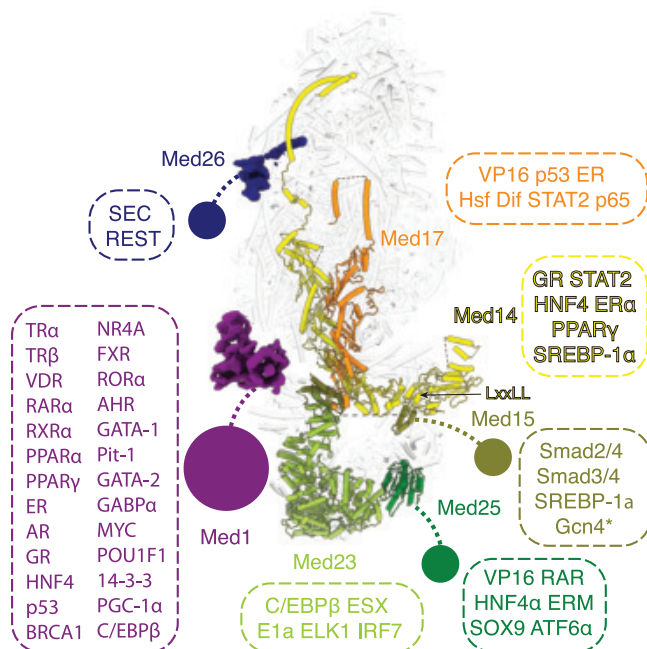


Fig. 3. Location of Mediator domains and subunits that interact with transcriptional activators or elongation factors. Flexible tethered domains are indicated by solid circles connected by dashed lines. All interactions shown are between human factors except for Gcn4, which is from yeast and indicated by an asterisk.



first interface with MedHead described above (figs. S7 and S9B). The N terminus of Med24 interacts with Med1 and is much more flexible than the rest of MedTail. We could only identify a single domain of Med25, the von Willebrand factor type A (vWA) domain, wedged in a pocket formed by Med16 and Med23 (fig. S9E).

Almost all domains that are bound by transcription factors in Mediator; including the N terminus of Med15, the N terminus of Med25, and the C terminus of Med1, are flexibly attached to the main body and not visible in the density map (Fig. 3). The first visible portion of Med15 is located underneath MedTail, near the upstream DNA, allowing its N terminus to easily engage with DNA-bound transcription factors. The C terminus of Med1 contains the NR-boxes important for nuclear receptor (NR) binding (24). Many NRs also bind to a C-terminal fragment (residues 1147 to 1454) of Med14 (25–27). These two binding interfaces for NRs are quite far from each other (Fig. 3). The NR AF-1 and AF-2 domains that mediate these interactions are at opposite ends of NR sequences, suggesting that NRs might have to stably associate with the full complex to bridge these two interfaces.

The VP16 activation domain (AD) used to purify Mediator for this study binds to the ACID domain located at the N terminus of Med25 (28, 29). The VP16 AD appears to stay bound to Mediator during complex assembly. Because of the absence of density of the ACID domain bound to the VP16 AD in this structure, we can conclude that the ACID domain remains flexibly tethered upon activator binding. It has been hypothesized that conformational changes after activator binding to Mediator could lead to the activation of Med-PIC (30–32). Given that so many of the activator-binding domains within Mediator are flexibly tethered to the main body, it is unlikely that this is a universal mechanism for activating Med-PIC for transcription.

The overall architecture of Mediator appears highly conserved based on recent structures of mouse and *Chaetomium thermophilum* Mediator (33, 34). The putative locations of Med1 and Med26 described earlier are in agreement with the mouse structure. The loss of subunits present in other species, including Med27 to Med30, Med23, and Med24, appears to increase the flexibility of MedTail in the *C. thermophilum* structure.

Mediator stabilizes the CAK module of TFIIF

Although previous structural studies of Med-PICs established that the CAK module of TFIIF occupies a position between the shoulder and hook domains of Mediator, the position and orientation of each CAK module subunit could not be determined (6–9). Rigid body docking of the human CAK module structure into our density led to an unambiguous orientation

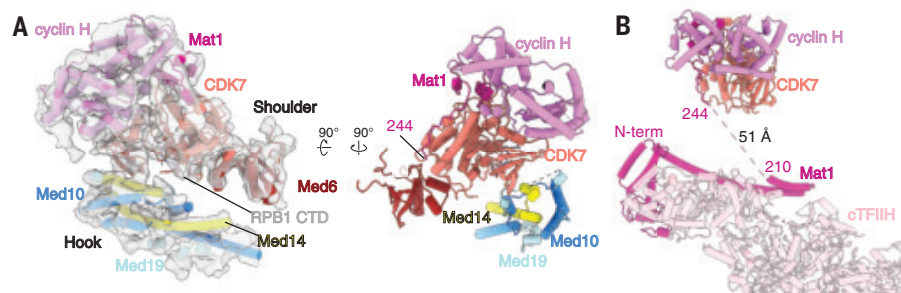


Fig. 4. Structure of TFIIF within Med-PIC. (A) Docking of the CAK module (CDK7, cyclin-H, and Mat1) within the MedMiddle-CAK density. The CAK module of TFIIF is stabilized in the Med-PIC by interactions between CDK7 and Med6, the N terminus of Med14, and a small fragment of Med19. (B) The model of the complete human TFIIF complex places the two modeled segments of Mat1 (residues 1 to 210 and 244 to 308) close to each other. The missing 34 residues can easily span the 51-Å distance between the termini. Models are colored as in Fig. 1.

These subunits, which were assigned previously to either MedHead or MedTail (15, 23), exhibit extensive interactions with the fixed jaw of MedHead and were therefore assigned to MedHead. The C terminus of the scaffold subunit Med14 extends the RM1 and RM2 repeats, which are visible in yeast structures, and wraps around MedHead, serving as a clear divider between MedHead and MedTail. Med17, a scaffold subunit within MedHead, stabilizes the fixed jaw on one face and interacts with the RM1 and RM2 repeats of Med14 on the other (Fig. 2B and fig. S9B).

MedTail connects to the rest of Mediator through two relatively small interfaces with MedHead and Med14. Two C-terminal domains of Med15 are crucial for forming both interfaces. The first contact site is located near the

C terminus of Med14. Two helices each from Med27 and Med29 project underneath Med14, with two helices of Med15 (residues 617 to 649) wedged between them (fig. S9C). A concave surface on Med16 makes contact with both this site and Med14. The second site is formed by a C-terminal extension between β^{18} and β^{19} (residues 596 to 620) of Med17 that interacts with the Ring-WD40-DEAD (RWD) domain of Med15 (residues 674 to 692) (figs. S5 and S9D). The RWD domain of Med15 is wedged in a large cavity between Med23 and Med24.

The rest of MedTail is formed by subunits Med16 and Med23 to Med25. Med16 is divided into N-terminal and C-terminal domains, with the N terminus forming a seven-blade WD40 domain and the C terminus forming a mostly helical domain that constitutes much of the

of the CAK module with the active site of CDK7 facing the hook domain of MedMiddle (Fig. 4A) (35). Mediator stabilizes the CAK module through interactions involving Med6, the N terminus of Med14, and a small fragment of Med19 (residues ~133 to 148) with CDK7 (Fig. 4A). This orientation of the CAK module positions the C terminus of Mat1 ~50 Å from the N terminus bound to cTFIIH, a distance easily spanned by the small fragment of Mat1 (residues 211 to 243) missing in the structure (Fig. 4B).

CDK7 adopts the active conformation seen in the human CAK module structure, with the T-loop projecting toward Mat1 and away from the active site (Fig. 5, A and B) (35). Clear electron density in the CDK7 active site closely matches the location of the substrate peptide in the CDK2–cyclin-A–substrate peptide complex (36) (Fig. 5, A to C). This peptide shares the identical serine-proline sequence that is found in the RPB1 CTD targeted by the kinase. Therefore, we built a model for the RPB1 CTD in the active site that we designate as $_{CDK}$ CTD.

S. cerevisiae MedHead (scMedHead) was cocrystallized with a short peptide of the RPB1 CTD, which shows slightly more than three full repeats engaged with scMedHead at the shoulder and neck domains (17). We observed additional electron density in this same location and used the *S. cerevisiae* structure to build a model for this portion of the CTD that we will refer to as $_{MED}$ CTD (Fig. 5, D and E). $_{MED}$ CTD is 16 residues long, is slightly more than two full repeats, and adopts a somewhat different path than the yeast structure, likely because of the presence of Med31, which interacts with the other side of $_{MED}$ CTD. In scMedHead, the elongated structure of the N-terminal portion of $_{MED}$ CTD forms extensive interactions with Med17. By contrast, we see clear density for $_{MED}$ CTD starting to wrap around Med31. The C-terminal end of $_{MED}$ CTD also does not form as extensive of an interface with Mediator as it does in scMedHead, because of a clash with the Med7 N terminus. Experiments in *S. pombe* show that the CTD is necessary for interaction between MedHead and Pol II in vitro, suggesting that $_{MED}$ CTD is critical for this interaction (8). $_{MED}$ CTD binding to Mediator would likely be disrupted after phosphorylation of Ser⁵ because of close interactions between Ser⁵ and the end of Med31 helix α^2 (Fig. 5D).

The directionality of $_{MED}$ CTD and $_{CDK}$ CTD is the same, with the N-terminal end of $_{MED}$ CTD pointing toward Pol II and the C-terminal end of $_{CDK}$ CTD leading away from Med-PIC (Fig. 5F). This observation strongly suggests that $_{MED}$ CTD is N-terminal to $_{CDK}$ CTD within the full CTD sequence. The distance between the termini of those two CTD fragments is 48 Å. In an elongated state, one repeat of the CTD can span about 25 Å (37), so although

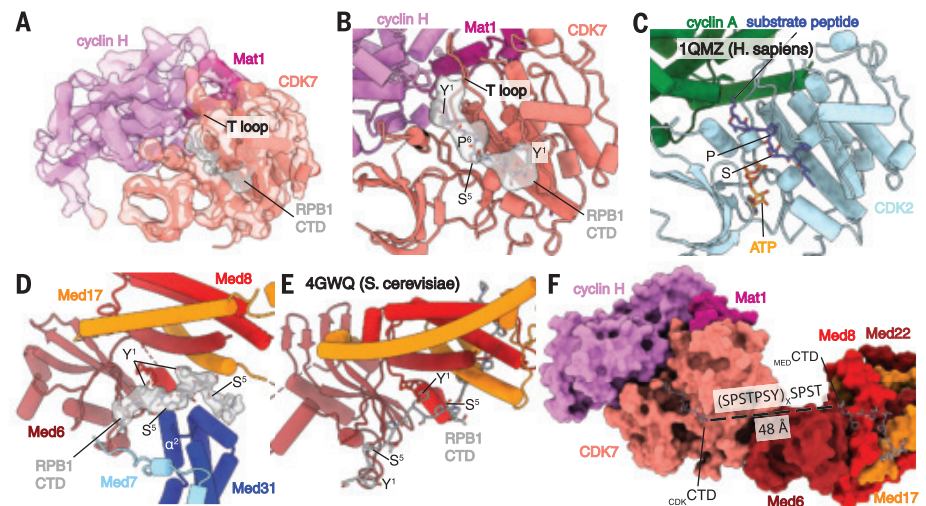
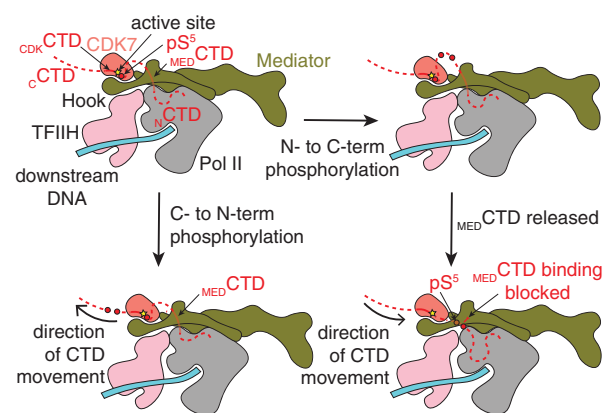


Fig. 5. Location of RPB1 CTD binding in Med-PIC. (A) Structure of the TFIIF CAK module. Segmented map of MedMiddle-CAK shows clear density representing an active conformation of the T-loop of CDK7 and density for Pol II CTD in the active site of CDK7. (B) Model of the CAK module with density observed for the $_{CDK}$ CTD in the active site. A consensus sequence of the Pol II CTD is modeled because of limited resolution. The T-loop is in the extended, active conformation. Y¹, Tyr¹; P⁶, Pro⁶; S⁵, Ser⁵. (C) Model of the CDK2–cyclin A–substrate peptide structure shows high similarity to the CAK module structure with the conserved SP motif that is common to substrates of both enzymes. ATP, adenosine triphosphate. *H. sapiens*, *Homo sapiens*. (D) Model and density of $_{MED}$ CTD with interacting subunits of MedHead and MedMiddle. Ser⁵ makes close contacts with α^2 of Med31, preventing binding of phosphorylated repeats in this location. (E) Model of $_{MED}$ CTD in the yeast MedHead crystal structure shows a more extensive interface between $_{MED}$ CTD and MedHead than in the Med-PIC, likely because of the presence of MedMiddle in the Med-PIC. (F) View of $_{CDK}$ CTD and $_{MED}$ CTD within the human Med-PIC structure. Based on the directionality of the CTD, $_{CDK}$ CTD is C-terminal to $_{MED}$ CTD, and the gap between them would require at least two repeats of the CTD. MedMiddle is hidden for easier visibility. Models are colored as in Fig. 1. Annotated domains of Mediator are labeled in black. The black dashed line represents the missing residues (SPSTPSY)_xSPST.

Fig. 6. Model for phosphorylation of the Pol II CTD by CDK7.

$_{MED}$ CTD binding positions the rest of the CTD in the CDK7 active site. After phosphorylation, which is indicated by a red circle, translocation of the CTD toward the N terminus (bottom) would place phosphorylated repeats further from the nascent RNA emerging from Pol II. Separation of Mediator and Pol II would be difficult without separation of the CAK module and Mediator. Translocation of the CTD toward the C terminus would position phosphorylated repeats to block binding of the CTD at $_{MED}$ CTD, a possible way to favor disassembly of Med-PIC. Phosphorylated repeats would also be substantially closer to the RNA exit tunnel of Pol II to recruit the capping complex properly. CTD, C-terminal domain of RPB1; pS⁵, phosphorylated Ser⁵ residue (red circle).



two repeats of the CTD may be sufficient to bridge that gap, we would likely see better-defined density for the CTD in that case. Therefore, we suspect that three or more repeats are likely looped out between $_{MED}$ CTD and $_{CDK}$ CTD.

The role of $_{MED}$ CTD binding is likely to capture the CTD and position it in the correct direction and close to the active site of CDK7 to facilitate phosphorylated Ser⁵ (pSer⁵) formation. Mass spectrometry experiments with both yeast and human complexes show that pSer⁵

can be found within any repeat of the CTD except the final repeat (38, 39). However, the phosphorylation patterns of individual CTD peptides and the direction in which sequential phosphorylation can occur remain unknown. Two possibilities exist for the direction of sequential phosphorylation that generate different outcomes (Fig. 6). If the CTD is phosphorylated in a C- to N-terminal direction, binding at _{MED}CTD precedes phosphorylation, and it is not clear how Pol II would dissociate from Mediator given that the CTD is threaded through a hole in Mediator formed by the hook, knob, and shoulder domains and the CAK module of TFIIF. Phosphorylated repeats would also be located far from the nascent RNA that needs to be capped.

If the CTD is phosphorylated in an N- to C-terminal direction, C-terminal phosphorylated repeats would not be able to bind at _{MED}CTD because of steric clashes that would arise with the added phosphates. Given that the CTD is important for Pol II-Mediator interaction and that phosphorylation of the CTD leads to dissociation of Pol II and Mediator, we find this mechanism more likely (8, 40). Separation of MedHead and Pol II would place the phosphorylated CTD close to the nascent RNA for capping to occur.

Given the large movements of MedMiddle and the CAK module of TFIIF relative to the PIC, we speculate that these conformational changes play an important role in the sequential phosphorylation of the CTD. The intrinsic flexibility of Mediator has been linked to the opening and closing of the _{MED}CTD binding site on Mediator (5, 6), and if this movement is tied to binding and release of the CTD at _{MED}CTD, it could also facilitate the progression of CDK7 along the CTD.

REFERENCES AND NOTES

- M. C. Thomas, C.-M. Chiang, *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178 (2006).
- J. Soutourina, *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018).
- E. Nogales, R. K. Louder, Y. He, *Annu. Rev. Biophys.* **46**, 59–83 (2017).
- C. Bernecky, P. Grob, C. C. Ebmeier, E. Nogales, D. J. Taatjes, *PLoS Biol.* **9**, e1000603 (2011).
- L. El Khattabi et al., *Cell* **178**, 1145–1158.e20 (2019).
- K.-L. Tsai et al., *Nature* **544**, 196–201 (2017).
- S. Schilbach et al., *Nature* **551**, 204–209 (2017).
- P. J. Robinson et al., *Cell* **166**, 1411–1422.e16 (2016).
- C. Plaschka et al., *Nature* **518**, 376–380 (2015).
- A. Meinhardt, T. Kamenski, S. Hoepfner, S. Baumbli, P. Cramer, *Genes Dev.* **19**, 1401–1415 (2005).
- A. Ghosh, C. D. Lima, *Wiley Interdiscip. Rev. RNA* **1**, 152–172 (2010).
- Y. He et al., *Nature* **533**, 359–365 (2016).
- D. Nair, Y. Kim, L. C. Myers, *J. Biol. Chem.* **280**, 33739–33748 (2005).
- C. Esnault et al., *Mol. Cell* **31**, 337–346 (2008).
- K.-L. Tsai et al., *Cell* **157**, 1430–1444 (2014).
- N. Petrenko, Y. Jin, K. H. Wong, K. Struhl, *Mol. Cell* **64**, 443–454 (2016).
- P. J. Robinson, D. A. Bushnell, M. J. Trnka, A. L. Burlingame, R. D. Kornberg, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17931–17935 (2012).
- C. Plaschka, K. Nozawa, P. Cramer, *J. Mol. Biol.* **428**, 2569–2574 (2016).
- M. A. Cevher et al., *Nat. Struct. Mol. Biol.* **21**, 1028–1034 (2014).
- T. Nakane, D. Kirmanius, E. Lindahl, S. H. W. Scheres, *eLife* **7**, e36861 (2018).

- R. K. Louder et al., *Nature* **531**, 604–609 (2016).
- H. Takahashi et al., *Cell* **146**, 92–104 (2011).
- C. Jeronimo, F. Robert, *Trends Cell Biol.* **27**, 765–783 (2017).
- A. Y. Belorousova et al., *Nucleic Acids Res.* **48**, 11199–11213 (2020).
- A. B. Hittelman, D. Burakov, J. A. Iñiguez-Lluhi, L. P. Freedman, M. J. Garabedian, *EMBO J.* **18**, 5380–5388 (1999).
- S. Malik, A. E. Wallberg, Y. K. Kang, R. G. Roeder, *Mol. Cell. Biol.* **22**, 5626–5637 (2002).
- L. Grøntved, M. S. Madsen, M. Boergesen, R. G. Roeder, S. Mandrup, *Mol. Cell. Biol.* **30**, 2155–2169 (2010).
- E. Vojnic et al., *Nat. Struct. Mol. Biol.* **18**, 404–409 (2011).
- A. G. Milbradt et al., *Nat. Struct. Mol. Biol.* **18**, 410–415 (2011).
- K. D. Meyer, S. C. Lin, C. Bernecky, Y. Gao, D. J. Taatjes, *Nat. Struct. Mol. Biol.* **17**, 753–760 (2010).
- D. J. Taatjes, A. M. Näär, F. Andel III, E. Nogales, R. Tjian, *Science* **295**, 1058–1062 (2002).
- C. Bernecky, D. J. Taatjes, *J. Mol. Biol.* **417**, 387–394 (2012).
- H. Zhao et al., *Nat. Commun.* **12**, 1355 (2021).
- H. Zhang et al., *Mol. Cell* **81**, 2765(21)00042:3 (2021).
- B. J. Greber et al., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 22849–22857 (2020).
- N. R. Brown, M. E. Noble, J. A. Endicott, L. N. Johnson, *Nat. Cell Biol.* **1**, 438–443 (1999).
- O. Jasnovidova, R. Steff, *Wiley Interdiscip. Rev. RNA* **4**, 1–16 (2013).
- H. Suh et al., *Mol. Cell* **61**, 297–304 (2016).
- R. Schüller et al., *Mol. Cell* **61**, 305–314 (2016).
- T. Max, M. Sogaard, J. Q. Svejstrup, *J. Biol. Chem.* **282**, 14113–14120 (2007).

ACKNOWLEDGMENTS

We thank past and present lab members for advice, assistance, and comments on the manuscript. We thank J. Pattie for computer support. We thank J. Meyers, R. M. Haynes, and H. Scott at the Pacific Northwest Center for Cryo-EM (PNCC) for data collection support. We are grateful to A. Rosenzweig, I. Radhakrishnan, and J. Brickner for helpful discussion and comments on the manuscript. We thank the staff at the Structural Biology Facility (SBF) of Northwestern University for technical support. **Funding:** This work was supported by a Corneal Innovation Award from the Chemistry of Life Processes Institute at Northwestern University (to Y.H.), a Catalyst Award by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (to Y.H.), an Institutional Research Grant from the American Cancer Society (IRG-15-173-21 to Y.H.), an H Foundation Core Facility Pilot Project Award (to Y.H.), and a Pilot Project Award under U54-CA193419 (to Y.H.). Y.H. is supported by R01-GM135651

and P01-CA092584 from the NIH. R.A. and A.T. are supported by the Molecular Biophysics Training Program from the National Institute of General Medical Sciences-NIH (T32-GM008382). A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at Oregon Health & Science University and accessed through the Environmental Molecular Sciences Laboratory (grid.436923.9), a Department of Energy, Office of Science User Facility sponsored by the Office of Biological and Environmental Research. R.T. is a Howard Hughes Medical Institute investigator. This work used the Sapphire imager from the Northwestern University Keck Biophysics Facility funded by NIH grant 1S100D026963-01, as well as the resources of the Northwestern University SBF, which is generously supported by National Cancer Institute Cancer Center Support Grant P30 CA060553 awarded to the Robert H. Lurie Comprehensive Cancer Center. **Author contributions:** Y.H. and R.T. conceived the project. Y.H., C.J.I., and R.A. purified proteins. A.T. assembled Med-PIC complexes and prepared cryo-EM samples. S.C. and Y.H. processed cryo-EM data. R.A. and A.T. built atomic models. R.A. and A.T. made the figures. R.A. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Electron density maps and coordinates for the Med-PIC have been deposited in the Electron Microscopy Data Bank (EMDB) with ID code EMD-23255 and the Protein Data Bank (PDB) with ID code 7LBM, respectively. Electron density maps for the focused refinements on cPIC, cTFIIF, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1 have been deposited in the EMDB with ID codes EMD-23256, EMD-23257, EMD-23258, EMD-23259, EMD-23260, EMD-23261, and EMD-23262, respectively.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/52/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (41–68)
MDAR Reproducibility Checklist
Movies S1 to S4

23 December 2020; accepted 3 March 2021
Published online 11 March 2021
10.1126/science.abg3074

MARS

Long-term drying of Mars by sequestration of ocean-scale volumes of water in the crust

E. L. Scheller^{1*}, B. L. Ehlmann^{1,2}, Renyu Hu², D. J. Adams¹, Y. L. Yung^{1,2}

Geological evidence shows that ancient Mars had large volumes of liquid water. Models of past hydrogen escape to space, calibrated with observations of the current escape rate, cannot explain the present-day deuterium-to-hydrogen isotope ratio (D/H). We simulated volcanic degassing, atmospheric escape, and crustal hydration on Mars, incorporating observational constraints from spacecraft, rovers, and meteorites. We found that ancient water volumes equivalent to a 100 to 1500 meter global layer are simultaneously compatible with the geological evidence, loss rate estimates, and D/H measurements. In our model, the volume of water participating in the hydrological cycle decreased by 40 to 95% over the Noachian period (~3.7 billion to 4.1 billion years ago), reaching present-day values by ~3.0 billion years ago. Between 30 and 99% of martian water was sequestered through crustal hydration, demonstrating that irreversible chemical weathering can increase the aridity of terrestrial planets.

There is abundant geomorphological evidence for large volumes of surface liquid water early in martian history (1), with estimated volumes equivalent to a ~100 to 1500 m global equivalent layer (GEL) (1–4). Liquid water on Mars decreased

over geological time; presently, most water is stored in the polar ice caps or as subsurface ice. Estimates for the total modern water inventory, in the atmosphere and as ice, total a 20 to 40 m GEL (5–8). The availability of water to participate in the hydrologic cycles of terrestrial

planets is expected to influence their climate and habitability. However, the processes that caused the decline of available water on Mars are poorly constrained.

Previous studies have suggested that Mars experienced substantial water loss from atmospheric escape, which is supported by the current atmospheric deuterium-to-hydrogen isotope ratio (D/H) of 5 to 10 × SMOW (standard mean ocean water on Earth; D/H at 1 SMOW is 155.76×10^{-6}) (5, 9–11). The D/H value at ~4 billion years ago was 2 to 4 × SMOW, inferred from martian meteorites (fig. S1) (12, 13). Existing models used these observations, combined with assumed atmospheric escape fractionation factors (α_{escape}) of 0.016 to 0.32 during loss, to estimate integrated atmospheric escape of at least 10 to 200 m GEL (fig. S1) (4, 5, 11, 14, 15). These estimates imply an initial 50 to 240 m GEL of water on ancient Mars, which is consistent only with the lower range of geological estimates (100 to 1500 m GEL) (1–4). This has been interpreted as implying a large, unknown reservoir of water on present-day Mars (4).

For present-day Mars, the rate of atmospheric water loss is measured from the H escape flux because water vapor dissociates in the atmosphere and its hydrogen escapes. Spacecraft measurements of the current H escape flux, 10^{26} to 10^{27} H atoms s^{-1} , are equivalent to the escape of 3 to 25 m GEL water across 4.5 billion years (16, 17) and cannot explain all the water loss. Another potential water loss mechanism is crustal hydration through irreversible chemical weathering, in which water and/or hydroxyl are incorporated into minerals. Orbital and in situ data show that widespread chemical weathering has produced a substantial reservoir of hydrous minerals on Mars, potentially totaling hundreds of meters of GEL in the crust (5, 18). We hypothesized that crustal hydration during the first 1 billion to 2 billion years decreased the volume of the hydrologically available water reservoir, followed by subsequent atmospheric loss that fractionated the martian atmosphere to its current observed D/H. We simulated water loss through geological time to constrain Mars' water history and to compare the simulations to D/H data from the Curiosity rover (5) and laboratory analyses of martian meteorites (fig. S1) (12, 13, 19–21).

A hydrogen isotope water reservoir model

We developed a water budget and D/H model that integrates water sinks and sources, including crustal hydration, volcanic degassing, and atmospheric escape (Fig. 1) (5). Most

previous models included only atmospheric escape (4, 11, 14); one model (15) also included volcanic degassing. We treat liquid water, ice, and atmospheric vapor as a single exchangeable reservoir, an isotopic modeling technique that was originally developed for carbon reservoir models (22). We assume that liquid and solid phases, not vapor, dominate the exchangeable reservoir and that fractionation between them is negligible [the fractionation factor is $\alpha_{\text{ice-liquid}} = 1.02$ (23)]. Our simulations are constrained so that the exchangeable reservoir can never be negative and must reproduce 20 to 40 m GEL water today. The initial exchangeable reservoir size ($X_{\text{ex},0}$)—the ancient hydrologically available water inventory—is a free parameter except during sensitivity analyses. We determined permitted ranges of source and sink fluxes for crustal hydration (F_{crust}), volcanic degassing (F_{volcanic}), and atmospheric escape (F_{esc}) during the Noachian (~4.0 billion to 3.7 billion years ago), Hesperian (~3.7 billion to 3.0 billion years ago), and Amazonian (~3.0 billion years ago to present) periods of martian geological history following observational and previous model constraints (Fig. 1 and table S1) (5). Models were evaluated by their ability to reproduce the D/H of the present-

day exchangeable reservoir ($R_{\text{ex, end}}$) of 5 to 10 × SMOW. We also compared our simulation results with a compilation of Curiosity rover Sample Analysis at Mars (SAM) data sets that recorded a D/H composition range of 3 to 5 × SMOW for gas released from Hesperian samples during high-temperature (>374°C) combustion experiments (5).

We calculated a permitted range of F_{crust} from measurements of water wt % in Mars surface materials and global remote sensing observations of hydrated minerals. The mass fraction of crustal water is based on rover measurements from Gale crater, orbital global infrared and neutron spectrometer data, and measurements of the NWA 7034 martian meteorite (0.5 to 3 wt % water) (5). The volume of the crustal reservoir is based on orbital measurements of clay exposure depths in the Valles Marineris canyon and craters 5 to 10 km in depth (5, 18). We adopted permitted ranges of 100 to 900 m GEL of water in Noachian-aged crust and 10 to 100 m GEL of water in Hesperian-aged crust on the basis of this analysis (table S1) (5, 18). Although F_{crust} is based on observations of hydrated minerals, we considered crustal water as a single reservoir representing any combination of ice, liquid, and

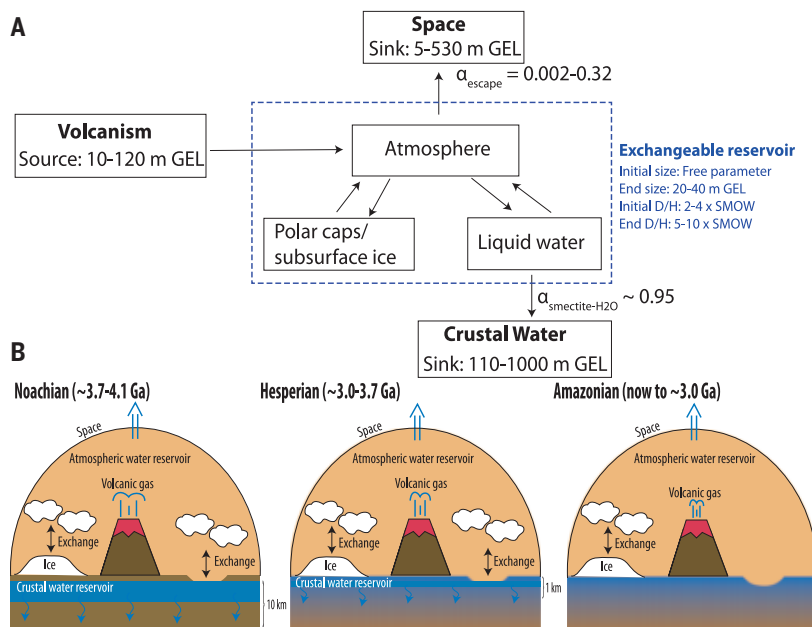


Fig. 1. Schematic illustration of water sink and source fluxes considered in our simulations. (A) Box model representation with ranges of integrated water sinks, sources, reservoir sizes, and fractionation factors adopted in our simulations. The crustal water reservoir is based on rover and remote sensing observations and represents all unexchangeable subsurface ice, liquid water, and structural water in minerals (5). The integrated amount of H escape to space is based on measurements of the current flux and KINETICS calculations of fluxes (figs. S2 and S3). The integrated volcanic degassing is based on thermochemical models (5, 24). The blue box indicates the exchangeable reservoir, with its properties in blue text. (B) Schematic representation of our assumptions for the Noachian, Hesperian, and Amazonian periods. During the Noachian, the fluxes associated with crustal hydration and volcanic degassing are high. These all reduce during the Hesperian. During the Amazonian, volcanic degassing falls further, and there is negligible crustal hydration because the water is predominantly solid ice. Ga, billion years ago.

¹Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. ²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA.

*Corresponding author. Email: eschelle@caltech.edu

structural water, formerly participating in the hydrologic cycle, that now no longer exchange isotopes with the exchangeable reservoir. We determined F_{volcanic} using previous thermochemical models of the martian mantle (24). Different parameterizations of those models (24) predict outgassing of a 10 to 120 m GEL of water from volcanic processes since 4.1 billion years ago (5, 24). We considered Noachian and Hesperian F_{esc} values between 10^{25} and 10^{30} H atoms s^{-1} and adopted the measured current escape rate of 5×10^{26} H atoms s^{-1} for the Amazonian (table S1) (5). We compared these escape fluxes with simulations using the one-dimensional (1D) photochemical model KINETICS (25, 26) with adopted past solar extreme ultraviolet flux, variable atmospheric pressures, and mesospheric and surface temperatures (table S2) (5).

Controls on D/H and water loss

In our model, stepwise mixing between the exchangeable reservoir and the depleted volcanically outgassed water vapor (0.8 to $2 \times \text{SMOW}$) (fig. S1 and table S1) (5, 19, 27) causes the D/H of the exchangeable reservoir to decrease (5). We do not include fractionation associated with degassing or its redox sensitivity because these are negligible compared with the large range of potential D/H compositions of the volcanic gas inferred from meteorites (5). Atmospheric escape causes D/H of the exchangeable reservoir to fractionate toward heavier values, which we modeled through stepwise Rayleigh distillation, a common isotopic reservoir modeling technique, at each 10-million-year time step with an α_{escape} of 0.002 to 0.32 (28–30). The fractionation factor between smectite, the most common hydrated mineral found on Mars, and water [$\alpha_{\text{smectite-H}_2\text{O}} = 0.95$ (5)] is used in the stepwise Rayleigh distillation model as a first-order approximation of fractionation through crustal hydration (table S3) (5); we found that this fractionation is minor compared with that caused by atmospheric escape.

The D/H of the exchangeable reservoir increases during the Noachian in all our simulations, and through the Hesperian in most of them, because of a combination of crustal hydration and atmospheric escape (Figs. 2 and 3). Higher $F_{\text{esc,N}}$ and $F_{\text{esc,H}}$ increase D/H fractionations of the exchangeable reservoirs (Fig. 2, A and B). We found that the Noachian and Hesperian H escape flux ranges that satisfy the model constraints (fig. S2) have a wide allowable range, ~ 0.1 to 1000 times the current 5×10^{26} H atoms s^{-1} escape flux. Independently, our KINETICS photochemical simulations (5) produced the same range ($\sim 10^{25}$ to 5×10^{29} H atoms s^{-1}) (fig. S3). We considered multiple scenarios, including (i) a range of standard ancient Mars conditions, (ii) high-altitude water injection [60 parts per million (ppm)

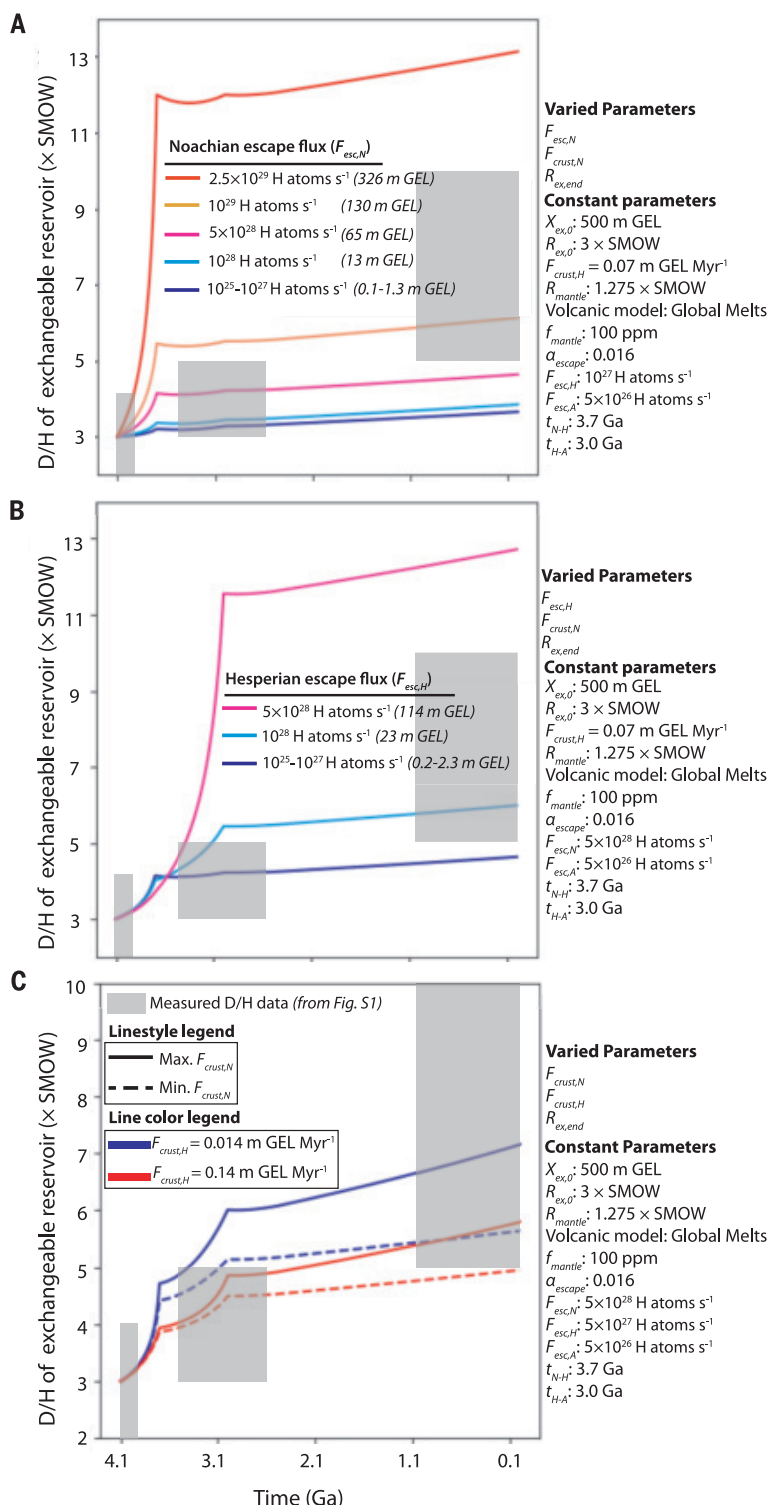


Fig. 2. Simulated D/H evolution for different assumptions of crustal hydration and atmospheric escape rates. (A to C) The evolution of the D/H of the exchangeable reservoir in our simulation. Most parameters, including $X_{\text{ex,0}}$, are fixed; $R_{\text{ex, end}}$ is a free parameter to visualize the model sensitivity. The colored lines show results for different assumptions of the flux rates. The large range of D/H measurements from meteorite, rover, and telescope observations are indicated with gray rectangles (fig. S1). (A) Effects of increasing the Noachian escape flux ($F_{\text{esc,N}}$). (B) Effects of increasing the Hesperian escape flux ($F_{\text{esc,H}}$). (C) Effects of increasing the Noachian ($F_{\text{crust,N}}$) and Hesperian ($F_{\text{crust,H}}$) crustal hydration fluxes. When $F_{\text{crust,N}}$ increases, the exchangeable reservoir becomes smaller, inducing larger fractionations during the Noachian. When $F_{\text{crust,H}}$ increases, the allowed values of $F_{\text{crust,N}}$ decrease, causing less fractionation during the Noachian.

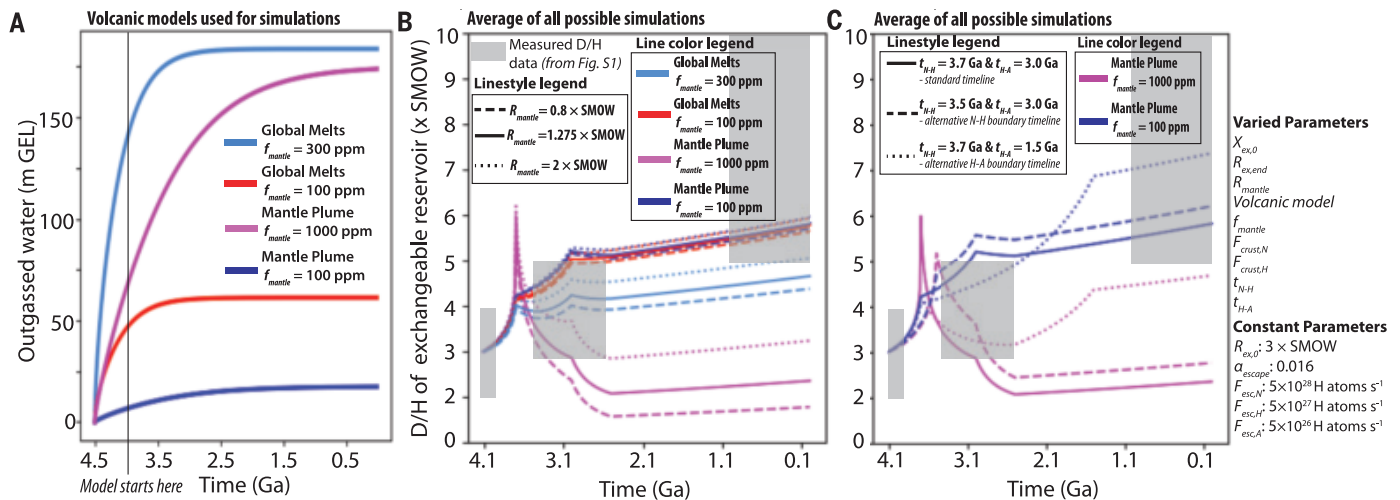


Fig. 3. Simulated D/H evolution for different assumptions of the volcanic outgassing as a function of time. (A) Adopted volcanic models (5, 24).

The Mantle Plume model (24) assumes an initial mantle water content (f_{mantle}) of 100 ppm (dark blue) or 1000 ppm (purple). The alternative Global Melts model (24) assumes f_{mantle} is 100 ppm (red) or 300 ppm (light blue). **(B)** The evolution of the D/H ratio in the exchangeable reservoir from an average of simulations with each assumed volcanic model. Line colors are the same as in (A), and gray boxes are the same as in Fig. 2. Line styles refer to assumed D/H composition of volcanic gas [dashed, $0.8 \times \text{SMOW}$ (27); solid, $1.275 \times \text{SMOW}$

(47); and dotted, $2 \times \text{SMOW}$ (19)]. **(C)** Evolution of the D/H in the exchangeable reservoir for average of simulations with different assumptions of volcanic model and age of the Noachian-Hesperian boundary ($t_{\text{N-H}}$) and the Hesperian-Amazonian boundary ($t_{\text{H-A}}$) (5). These transition ages control when F_{esc} and F_{crust} values change under our assumptions for the Noachian, Hesperian, and Amazonian periods (5). Line colors are the same as in (A). Line styles refer to the assumed timing of $t_{\text{N-H}}$ and $t_{\text{H-A}}$ (solid, standard boundary ages where $t_{\text{N-H}}$ is 3.7 Ga and $t_{\text{H-A}}$ is 3.0 Ga; dashed, $t_{\text{N-H}}$ is moved to 3.5 Ga; dotted, $t_{\text{H-A}}$ is moved to 1.5 Ga). In these simulations, $R_{\text{ex},\text{end}}$ is allowed to vary.

at 100 km], and (iii) fixing a surface H_2 mixing ratio of 10^{-3} , which is higher than present-day levels of 10^{-5} (26). The maximum KINETICS-permitted escape flux ($\sim 5 \times 10^{29}$ H atoms s^{-1}) and our D/H model maximum permitted flux (4×10^{29} H atoms s^{-1}) match the diffusion-limited escape of 5×10^{29} H atoms s^{-1} that we calculated using equations from (31). The injection of high-altitude water and increased surface H_2 concentrations both increase the production of high-altitude H_2 ; one or both would be required for loss fluxes 100 to 1000 times higher than that of the present (fig. S3).

Crustal hydration during early Mars history also increases D/H fractionation of the exchangeable reservoirs, with the permitted range of $F_{\text{crust},N}$ depending on the assumed $F_{\text{crust},H}$ (Fig. 2C). This is primarily because higher $F_{\text{crust},N}$ decreases the exchangeable reservoir size, not because of the fractionation [$\alpha_{\text{smectite-H}_2\text{O}} = 0.95$ (5)] associated with clay formation. Because the exchangeable reservoir is reduced through crustal hydration, less atmospheric escape is needed to produce the modern D/H of the atmosphere. During the Noachian, decreasing exchangeable reservoir size and increasing D/H are a feature of all of our simulations. Changes to the assumed timing of the boundary between the Noachian and Hesperian ($t_{\text{N-H}}$) and balance of $F_{\text{crust},N}$ to $F_{\text{crust},H}$ only slightly affect the Noachian D/H fractionation (Figs. 2C and 3C). During the Amazonian, the exchangeable reservoir size is low, and its D/H increases slightly in all

our simulations because of the lack of crustal hydration, low H escape flux (assumed equal to the present rate), and a low volcanic degassing flux (Figs. 2 and 3). By contrast, the D/H evolution during the Hesperian is less well constrained because models with low total volcanic outgassing (10 to 20 m GEL) result in D/H increases, whereas models with high outgassing (60 to 120 m GEL) result in D/H decreasing or staying approximately constant (Fig. 3, A and B). The amount of volcanic degassing controls the required sizes of F_{crust} and F_{esc} for different $X_{\text{ex},0}$ to reproduce the present-day D/H ($R_{\text{ex},\text{end}}$) (figs. S4 to S6). Evolution of Hesperian D/H is also sensitive to the absolute timing of the debated (5) boundary between the Hesperian and Amazonian periods ($t_{\text{H-A}}$) because in our model, that boundary sets the hydration and volcanic flux magnitudes (Fig. 3C).

Crustal hydration as a water sink

Considering the simulations over our whole parameter space, we found that the amounts of water lost through crustal hydration and atmospheric escape vary in ratios ranging from 3:8 to 99:1 (Fig. 4 and figs. S4 to S6), which is equivalent to ~ 30 to 99% of initial water being lost through crustal hydration (5). The maximum proportional contribution of atmospheric escape occurs when the volume of the crustal water reservoir is minimum and vice versa. Any larger proportional escape would produce D/H heavier than the present-

day observed value ($>10 \times \text{SMOW}$). However, the absolute allowed volumes of integrated crustal hydration and atmospheric escape are dependent on the size of the initial exchangeable reservoir (figs. S4 to S6). For some of our model solutions, no difference in the average atmospheric escape flux relative to the present-day flux is required to account for the observed increase in D/H and decrease in the exchangeable water reservoir (Fig. 4 and figs. S3 and S4). Both the maximum and minimum escape-to-space cases (Fig. 4 and figs. S4 to S6) occur with intermediate assumed initial exchangeable reservoir volumes (~ 500 m GEL).

Accounting for water loss through both crustal hydration and atmospheric escape (figs. S4 to S6) resolves the apparent contradiction between the estimates of integrated H escape, the D/H of present-day Mars, and geological estimates of a large and ancient exchangeable reservoir (1, 4). These can be reconciled because the amount of atmospheric escape needed for the atmosphere to reach the present-day D/H is reduced by the removal of large initial water volumes through crustal hydration. Our models require larger Noachian exchangeable reservoirs (100 to 1500 m GEL) than those of previous work (50 to 240 m GEL) because we include crustal hydration (Fig. 4F). The whole parameter space allows for initial exchangeable water reservoirs of 100 to 1500 m GEL at 4.1 billion years ago, 20 to 300 m GEL at the Noachian-Hesperian boundary, and a near-constant 20 to 40 m

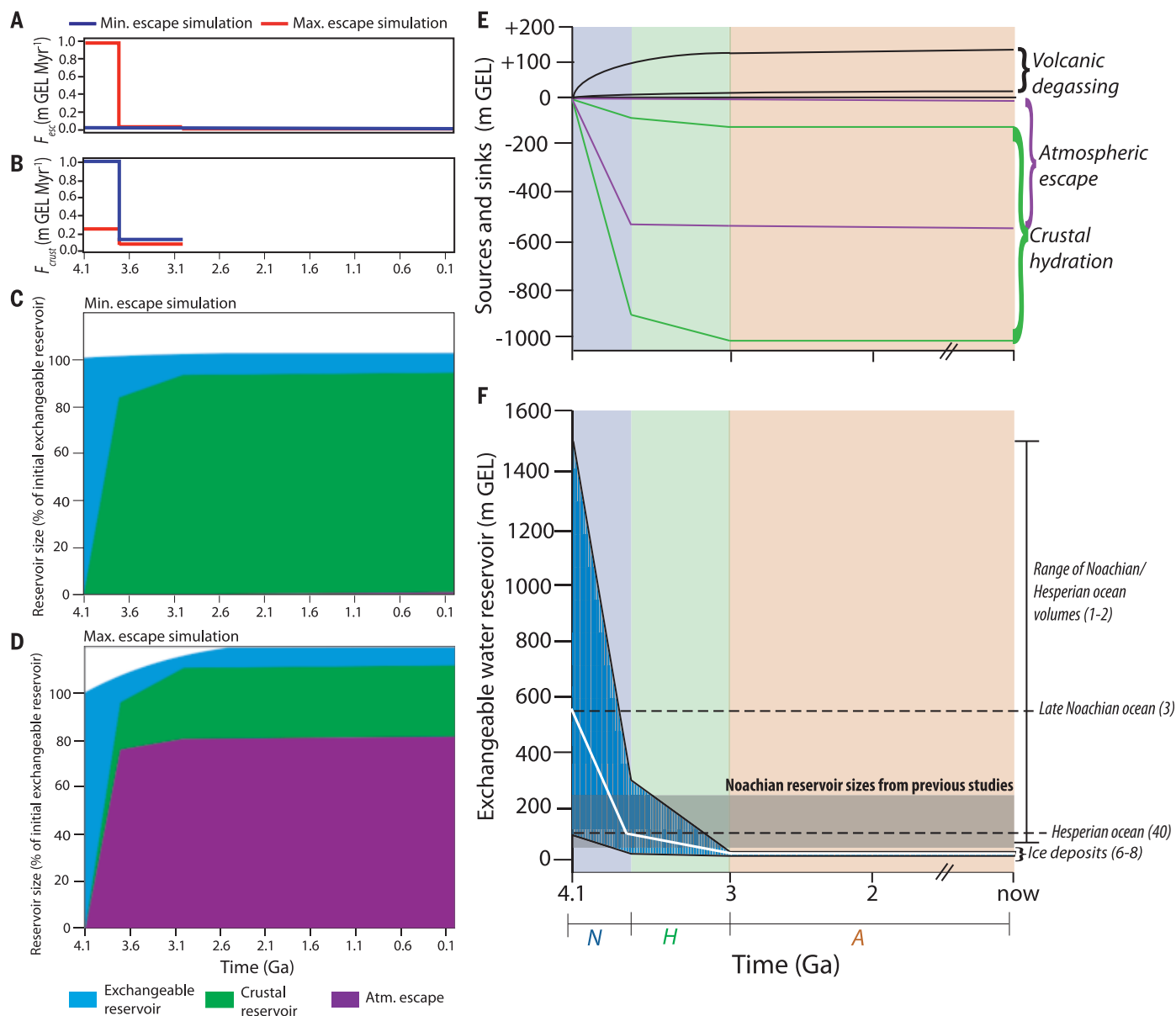


Fig. 4. Compilation of relative reservoir sizes through time from all our simulations. (A to D) Model simulations with minimum and maximum possible atmospheric escape fluxes (F_{esc}) and crustal hydration fluxes (F_{crust}) within allowed parameter space and simulation constraints, where the exchangeable reservoir D/H of 5 to 10 \times SMOW must be reproduced. (A) Evolution of minimum (blue line) and maximum (red line) F_{esc} within the constrained simulation space through geological time. (B) Evolution of minimum (red line) and maximum (blue line) F_{crust} within the constrained simulation space through geological time. (C) and (D) Size evolution of three simulated reservoirs through geological time shown as a cumulative percentage. Colored areas indicate the time evolution within the exchangeable reservoir (blue), crustal reservoir (green), and water

escaped to the atmosphere (purple). (E) The scenario in which F_{esc} is minimized and F_{crust} is maximized. (F) The scenario in which F_{esc} is maximized and F_{crust} is minimized. (G) Upper and lower bounds on sources and sinks from Fig. 1 through time derived from our simulations (black, volcanic degassing source; green, crustal hydration sink; purple, atmospheric escape sink) (5). (H) The range of exchangeable reservoir sizes (teal) permitted by our simulations. For comparison, we show the reservoirs derived by previous studies (gray rectangle) (4, 11, 14, 15) and ocean sizes based on geomorphological evidence (dashed lines) (1–3, 40). Our preferred simulation scenario is shown as a solid white line. Noachian (N), Hesperian (H), and Amazonian (A) time intervals used in model are shaded in blue, green, and red, respectively.

GEL throughout the Amazonian (Fig. 4F). We chose a preferred solution on the basis of observational constraints on the parameter space (Table 1 and Fig. 4F). In this preferred simulation, the Noachian and Hesperian H escape fluxes are twice that of today: $F_{esc,N} = F_{esc,H} \sim 10^{27}$ H atoms s⁻¹. The KINETICS simu-

lations indicate that the most probable long-term H escape flux was similar to that of today, although there may have been enhancements of shorter duration, such as during dust storms or surface fluxes of H₂ from geologic processes (figs. S2 and S3) (5). In the preferred model, crustal hydration removes 500 m

GEL and 50 m GEL during the Noachian and Hesperian, respectively, corresponding to roughly 3 wt % H₂O in Noachian crust of 5 km thickness and 1 wt % H₂O in Hesperian crust of 1 km thickness (18). This is compatible with the range of present-day water contents and crustal reservoir depths measured from orbit

Table 1. Summary of parameters assumed or calculated in our preferred scenario. We list the assumed parameter values for our preferred simulation (Fig. 4F) and our reasoning for each choice. This preferred simulation reproduces a D/H composition of $\sim 5.3 \times \text{SMOW}$ for the present-day atmosphere and an initial exchangeable reservoir size of ~ 570 m GEL. Myr, million years.

Variable	Meaning	Value	Units	Reasoning
Calculated				
$R_{\text{ex, end}}$	D/H of present-day exchangeable reservoir	$\sim 5.3 \times \text{SMOW}$	N/A	Calculated result of our preferred model
$X_{\text{ex, 0}}$	Initial size of exchangeable reservoir	~ 570	m GEL	Calculated result of our preferred model
Assumed				
$R_{\text{ex, 0}}$	Initial D/H of exchangeable reservoir	$4 \times \text{SMOW}$	N/A	D/H measurements of ALH84001 (13)
R_{mantle}	D/H of mantle	$1.275 \times \text{SMOW}$	N/A	D/H measurements of meteorites (47)
$\alpha_{\text{smectite-H}_2\text{O}}$	D/H fractionation factor between smectite and water	0.95	N/A	Literary review of geochemical experiments (table S2) (5)
α_{escape}	D/H fractionation factor of atmospheric escape	0.16	N/A	Photochemical model result (29)
$X_{\text{ex, end}}$	Present-day size of exchangeable reservoir	20 to 40	m GEL	A range of remote sensing evidence (5)
$F_{\text{crust, N}}$	Rate of water drawdown by crustal hydration during the Noachian	1.25	m GEL Myr ⁻¹	Intermediate value based on remote sensing evidence (5, 18)
$F_{\text{crust, H}}$	Rate of water drawdown by clay formation during the Hesperian	0.07	m GEL Myr ⁻¹	Intermediate value based on remote sensing evidence (5, 18)
f_{mantle}	Water content of mantle	100	ppm	Most commonly adopted meteorite measurements (5, 24)
F_{volcanic}	Rate of volcanic degassing of H ₂ O	Time-dependent fluxes	m GEL Myr ⁻¹	Compiled from two thermal evolution models (24)
$F_{\text{volcanic, A}}$	Rate of volcanic production after 2.5 Ga	2×10^{-4}	m GEL Myr ⁻¹	Geological remote sensing evidence (5)
$F_{\text{esc, A}}$	Present-day H escape flux	5×10^{26}	H atoms s ⁻¹	Spacecraft measurements (5, 16)
$F_{\text{esc, N}}$	H escape flux during the Noachian	10^{27}	H atoms s ⁻¹	Modeled in this study (figs. S2 and S3) (5)
$F_{\text{esc, H}}$	H escape flux during the Hesperian	10^{27}	H atoms s ⁻¹	Modeled in this study (figs. S2 and S3) (5)
$t_{\text{N-A}}$	End of deep, Noachian crustal alteration	3.7	Ga	Most commonly adopted age (5)
$t_{\text{H-A}}$	End of shallow, Hesperian crustal alteration	3.0	Ga	Most commonly adopted age (5)

and rovers (5). F_{volcanic} is assumed on the basis of volcanic degassing simulations (24), which themselves assumed $f_{\text{mantle}} = 100$ ppm on the basis of meteorite measurements (5). This is compatible with observational constraints on crustal production rates and water contents of martian meteorites (5). Our preferred simulation is therefore similar to the minimum escape case shown in Fig. 4C. These simulations adopt $R_{\text{ex, 0}} = 4 \times \text{SMOW}$ on the basis of meteorite measurements (5) and produce a present-day D/H of $\sim 5.3 \times \text{SMOW}$.

Consequences for Mars evolution

If the planet accreted with 0.1 to 0.2 wt % water (32), the large Noachian exchangeable reservoirs predicted by the model are consistent with Mars primordial water volumes. A martian primordial volume of >1100 m GEL (potentially thousands of meters of GEL) could have been produced by catastrophic outgassing of the mantle (~ 500 to 6000 m GEL) (33, 34), delivery of water through impacts (600 to 2700 m GEL) (35), and/or capture

of gases from the protoplanetary disc (36). However, the high hydrogen loss rates indicated by the D/H at 4.1 billion years ago recorded within meteorites (4, 11) and possible evidence for hydrodynamic escape in xenon isotopes (37) suggest that a large part of the primordial atmosphere and water were lost during the pre-Noachian period. Our proposed volumes of a 100 to 1500 m GEL during the early Noachian are within the lower end of these predicted primordial volumes and would therefore be compatible with the loss of a large part of the primordial atmosphere. After loss of the primordial atmosphere, isotope measurements of carbon and argon suggest that loss of a large fraction of these elements from the remaining martian atmosphere and the reservoirs that exchange with the atmosphere would have occurred after 4.1 billion years ago (22, 37–39). This matches our proposed trajectory of water loss within the exchangeable reservoir, which is reduced by 80 to 99% after 4.1 billion years ago within our model simulations.

Our modeled initial reservoirs are also consistent with geological estimates of Noachian and Hesperian surface water volumes. A 100 to 150 -m GEL ocean during the Hesperian (1, 40) has been suggested from geomorphological observations and is compatible with our preferred simulation. A larger 550 m GEL ocean that has been suggested at the Noachian-Hesperian boundary (3) is possible in simulations in which F_{crust} and F_{esc} are both maximized in the Noachian and Hesperian, requiring the initial exchangeable water reservoir at 4.1 billion years ago to be a ~ 1500 m GEL (Fig. 4F). Even larger oceans of 1000 to 1500 m GEL have been proposed on the basis of geomorphology (1, 2); these would be permitted only in certain simulation scenarios during the early Noachian and not later epochs (Fig. 4F).

Our models are compatible with the major observed trajectories of the martian climate. A high-volume Noachian exchangeable reservoir is consistent with geomorphological evidence for large volumes of Noachian surface waters and observed widespread hydrated mineral

formation. Aqueous alteration of the crust could have produced periods of warmer and wetter climates (supplementary text) (41–43) through accumulation of H₂ in the atmosphere (figs. S4 to S6). In cases in which atmospheric escape dominates water loss over the crustal hydration sink, H loss could be balanced by atmospheric oxygen escape (18 to 58 m GEL) and crustal oxidation (~30 to 380 m GEL) (supplementary text). However, in cases in which crustal hydration dominates water loss, short-term accumulation of H₂ could have occurred (supplementary text). In our KINETICS simulations, the accumulation of H₂ in the atmosphere results in increased H escape flux (fig. S3) (5).

The permitted parameter space of our D/H model allows either (i) a Hesperian exchangeable reservoir that was initially large but smaller than the Noachian reservoir (≤300-m GEL) and decreased or (ii) a Hesperian reservoir that was similar to present-day levels of a 20 to 40 m GEL (Fig. 4F). In case (i), the Hesperian may have had sustained periods of warm and wet climate, which could have caused chemical weathering on a global scale and potentially formed an ocean (1, 40). In case (ii), the Hesperian climate was likely similar to the Amazonian climate, with the exception of few local and short-lived instances of surface liquid water reservoirs (44). During the Amazonian period, the low H escape flux and low volcanic degassing flux counter each other, producing low model water availability within the exchangeable reservoir that is consistent with geomorphological and mineralogical evidence of an arid climate (Fig. 4F) (31, 45).

Crustal hydration would produce a buried water reservoir with a composition reflecting that of the Noachian exchangeable reservoir of ~2 to 4 × SMOW. Martian meteorites that are 1.6 billion to 0.1 billion years old have D/H values of ~2 to 3 × SMOW (20, 21). Previously proposed explanations include a distinct subsurface fluid reservoir, mixing between low-D/H igneous and high-D/H present-day atmospheric material, or terrestrial contamination (20, 21). We suggest that exchange between younger igneous rocks and fluids derived from hydrated Noachian (~2 to 4 × SMOW) crust could account for the intermediate D/H in these meteorites.

Comparative planetary evolution

We conclude that the increasing aridity of Mars over its history was caused by the sink of chemical weathering of the crust (Fig. 4), which was recorded in the widespread Noachian hydrated minerals on the planet's surface (18). On Earth, crustal hydration also occurs, but plate tectonics enables recycling of crustal water that is eventually outgassed to the atmosphere through volcanism (46). This has facilitated sustained participation of water in the hydro-

logic cycle throughout geological history on Earth (46). The ancient age of most hydrated minerals (45) indicates that any such recycling did not persist on Mars. Irreversible chemical weathering therefore plays a role in regulating the habitability of terrestrial planets by controlling the time scales of sustained participation of water in the hydrologic cycle.

Our model makes testable predictions for D/H measurements of the rock and ice record (Figs. 2 and 3): a substantial long-term secular increase in D/H over the Noachian and potentially Hesperian, with little change over the Amazonian. Under a variable climate, our model also indicates that the geological record might contain evidence of short-term D/H cyclicity: Transient warm periods with greater atmospheric H₂O (42) would periodically increase crustal hydration and escape flux, rapidly increasing D/H, whereas during cold periods, the D/H would decrease or increase slowly, depending on the balance between volcanic degassing and atmospheric escape.

REFERENCES AND NOTES

- M. H. Carr, J. W. Head, *J. Geophys. Res. Planets* **108**, 5042 (2003).
- S. M. Clifford, T. J. Parker, *Icarus* **154**, 40–79 (2001).
- G. Di Achille, B. M. Hynke, *Nat. Geosci.* **3**, 459–463 (2010).
- H. Kurokawa et al., *Earth Planet. Sci. Lett.* **394**, 179–185 (2014).
- Materials and methods are available as supplementary materials.
- M. T. Zuber et al., *Science* **282**, 2053–2060 (1998).
- J. J. Plaut et al., *Science* **316**, 92–95 (2007).
- P. R. Christensen, *Elements* **2**, 151–155 (2006).
- T. M. Donahue, *Nature* **374**, 432–434 (1995).
- C. R. Webster et al., *Science* **341**, 260–263 (2013).
- G. L. Villanueva et al., *Science* **348**, 218–221 (2015).
- N. Z. Bector, C. M. O. D. Alexander, J. Wang, E. Hauri, *Geochim. Cosmochim. Acta* **67**, 3971–3989 (2003).
- J. P. Greenwood, S. Itoh, N. Sakamoto, E. P. Vicenzi, H. Yurimoto, *Geophys. Res. Lett.* **35**, L05203 (2008).
- H. Lammer et al., *Int. J. Astrobiol.* **2**, 195–202 (2003).
- N. R. Alsaed, B. M. Jakosky, *J. Geophys. Res. Planets* **124**, 3344–3353 (2019).
- B. M. Jakosky et al., *Icarus* **315**, 146–157 (2018).
- M. A. Chaffin et al., *Geophys. Res. Lett.* **41**, 314–320 (2014).
- J. F. Mustard, Sequestration of volatiles in the Martian crust through hydrated minerals: A significant planetary reservoir of water, in *Volatiles in the Martian Crust* (Elsevier, ed. 2, 2019), pp. 247–264.
- L. A. Leshin, *Geophys. Res. Lett.* **27**, 2017–2020 (2000).
- T. Usui, C. M. O. D. Alexander, J. Wang, J. I. Simon, J. H. Jones, *Earth Planet. Sci. Lett.* **410**, 140–151 (2015).
- Y. Liu et al., *Earth Planet. Sci. Lett.* **490**, 206–215 (2018).
- R. Hu, D. M. Kass, B. L. Ehlmann, Y. L. Yung, *Nat. Commun.* **6**, 10003 (2015).
- T. Chacko, D. R. Cole, J. Horita, *Rev. Mineral.* **43**, 1–81 (2001).
- M. Grott, A. Morschhauser, D. Breuer, E. Hauber, *Earth Planet. Sci. Lett.* **308**, 391–400 (2011).
- M. Allen, Y. L. Yung, J. W. Waters, *J. Geophys. Res. Space Phys.* **86**, 3617–3627 (1981).
- H. Nair, M. Allen, A. D. Anbar, Y. L. Yung, R. T. Clancy, *Icarus* **111**, 124–150 (1994).
- P. Gillet et al., *Earth Planet. Sci. Lett.* **203**, 431–444 (2002).
- E. Cangi, M. S. Chaffin, J. Deighan, *J. Geophys. Res. Planets* **125**, (2020).
- V. Krasnopolsky, *Icarus* **148**, 597–602 (2000).
- Y. L. Yung et al., *Icarus* **76**, 146–159 (1988).
- J. F. Kasting, J. B. Pollack, *Icarus* **53**, 479–508 (1983).
- R. Brasser, *Space Sci. Rev.* **174**, 11–25 (2013).

- L. T. Elkins-Tanton, *Earth Planet. Sci. Lett.* **271**, 181–191 (2008).
- N. V. Erkaev et al., *Planet. Space Sci.* **98**, 106–119 (2014).
- J. Lunine, J. Chambers, A. Morbidelli, L. A. Leshin, *Icarus* **165**, 1–8 (2003).
- H. Lammer et al., *Space Sci. Rev.* **174**, 113–154 (2013).
- B. M. Jakosky, J. H. Jones, *Rev. Geophys.* **35**, 1–16 (1997).
- S. K. Atreya et al., *Geophys. Res. Lett.* **40**, 5605–5609 (2013).
- B. M. Jakosky et al., *Science* **355**, 1408–1410 (2017).
- M. H. Carr, J. Head, *Icarus* **319**, 433–443 (2019).
- R. D. Wordsworth, *Annu. Rev. Earth Planet. Sci.* **44**, 381–408 (2016).
- R. D. Wordsworth, L. Kerber, R. T. Pierrehumbert, F. Forget, J. W. Head, *J. Geophys. Res. Planets* **120**, 1201–1219 (2015).
- N. J. Tosca, I. A. M. Ahmed, B. M. Tutolo, A. Ashpitel, J. A. Hurwitz, *Nat. Geosci.* **11**, 635–639 (2018).
- J. P. Grotzinger et al., *Science* **350**, aac7575 (2015).
- B. L. Ehlmann et al., *Nature* **479**, 53–60 (2011).
- H. Lammer et al., *Astron. Astrophys. Rev.* **17**, 181–249 (2009).
- T. Usui, C. M. O. D. Alexander, J. Wang, J. I. Simon, J. H. Jones, *Earth Planet. Sci. Lett.* **357**, 358, 119–129 (2012).
- E. L. Scheller, Mars D/H model and KINETICS data files (version 1.0). CaltechDATA (2021); doi:10.22002/D1.1879.

ACKNOWLEDGMENTS

We thank A. Hoffmann, P. Mahaffey, C. Webster, H. Franz, J. Stern, D. Breuer, J. Dickson, J. Eiler, J. Grotzinger, Y. Liu, E. Stolper, and the Sample Analysis at Mars (SAM) Science Team for discussion. We thank B. Jakosky and two anonymous referees for suggestions that strengthened the manuscript. **Funding:** R.H., B.L.E., and Y.L.Y. were supported by a NASA Habitable Worlds grant (NNN13D466T, later changed to 80NM0018F0612). Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (grant 80NM0018D0004). E.L.S. was supported by a NASA Earth and Space Science Fellowship (NESSF) (grant 80NSSC18K1255). D.J.A. was supported by a Future Investigator in NASA Earth and Space Science and Technology (FINESST) fellowship (grant 80NSSC19K1548). **Author contributions:** E.L.S. drafted the manuscript, developed the code, and performed the simulations for the D/H model. B.L.E. devised the original idea of an integrated approach to a water budget and D/H model. E.L.S., B.L.E., and R.H. developed the water budget and D/H model concept. E.L.S., B.L.E., R.H., D.J.A., and Y.L.Y. adapted the KINETICS model input parameters for this study. D.J.A. implemented the KINETICS adaption, and D.J.A. and Y.L.Y. analyzed the KINETICS output. All authors participated in the writing and editing of the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** The equations used for the D/H model and our adopted parameter ranges are given in the supplementary materials. The Mars D/H model code, input and output files, the KINETICS data files used for fig. S3, and visualization scripts are all available in the CaltechDATA repository at (48). The KINETICS software was developed by a combination of authors (D.J.A. and Y.L.Y.) and a large number of nonauthors (25, 26), so we do not have permission to distribute the source code. An executable version with adjustable input parameters, to reproduce all simulation scenarios in this paper, is available at the same DOI, 10.22002/D1.1879. The SAM data were taken from the Planetary Data System at https://pds-geosciences.wustl.edu/msl/msl-m-sam-2-rdr-10-v1/mslsam_bxxo/data; we used level 2 data for samples eid25094, eid25123, eid25173, eid25413, eid25484, eid25515, and eid25538 (5).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/56/suppl/DC1
Materials and Methods
Supplementary Text
Tables S1 to S3
Figs. S1 to S6
References (49–122)

11 June 2020; accepted 4 March 2021
Published online 16 March 2021
10.1126/science.abc7717

PALEOECOLOGY

Extinction at the end-Cretaceous and the origin of modern Neotropical rainforests

Mónica R. Carvalho^{1,2*}, Carlos Jaramillo^{1,3,4*}†, Felipe de la Parra⁵, Dayenari Caballero-Rodríguez¹, Fabiany Herrera^{1,6}, Scott Wing⁷, Benjamin L. Turner^{1,8}, Carlos D'Apolito^{1,9}, Millerlandy Romero-Báez^{1,10}, Paula Narváez^{1,11}, Camila Martínez¹, Mauricio Gutiérrez^{1,12}, Conrad Labandeira^{7,13,14}, German Bayona¹⁵, Milton Rueda¹⁶, Manuel Paez-Reyes^{1,17}, Dairon Cárdenas¹⁸, Álvaro Duque¹⁹, James L. Crowley²⁰, Carlos Santos²¹, Daniele Silvestro^{22,23}

The end-Cretaceous event was catastrophic for terrestrial communities worldwide, yet its long-lasting effect on tropical forests remains largely unknown. We quantified plant extinction and ecological change in tropical forests resulting from the end-Cretaceous event using fossil pollen (>50,000 occurrences) and leaves (>6000 specimens) from localities in Colombia. Late Cretaceous (Maastrichtian) rainforests were characterized by an open canopy and diverse plant–insect interactions. Plant diversity declined by 45% at the Cretaceous–Paleogene boundary and did not recover for ~6 million years. Paleocene forests resembled modern Neotropical rainforests, with a closed canopy and multistratal structure dominated by angiosperms. The end-Cretaceous event triggered a long interval of low plant diversity in the Neotropics and the evolutionary assembly of today's most diverse terrestrial ecosystem.

Paleontological evidence indicates that the bolide impact at Chicxulub, 66.02 million years ago (Ma) (1), had immediate catastrophic effects on plant communities and reshaped terrestrial ecosystems worldwide (2–4). Despite the extent of this ecological disruption, the long-term extinction and recovery patterns were geographically heterogeneous (5). As much as 90% of pre-extinction palynomorphs reappeared during the Danian (66 to 61.6 Ma) in Patagonia and New Zealand (6, 7), and species-rich Danian megafloral assemblages with diverse types of insect damage indicate rapid recovery of diversity in Patagonia

(8, 9). By contrast, palynofloral extinction was up to 30% in the Northern Great Plains of North America (10), and floral and insect-damage diversity may not have reached pre-extinction levels until the latest Paleocene or early Eocene [(11, 12); but see (3)].

Phylogenies of several plant lineages suggest that the Cretaceous–Paleogene (K/Pg) event marking the end of the Cretaceous played a role in shaping modern tropical lowland rainforests (13–15), but the fate of tropical forests following the K/Pg boundary is not well under-

stood. Assessing plant extinction and recovery requires a thoroughly sampled fossil record, yet aside from an impact-related fern-spore spike in deep-water strata from Gorgonilla, Colombia (16), the plant fossil record across the K/Pg boundary in the lowland Neotropics is sparse (17). Here, we quantify changes in the diversity, structure and composition of forests across the K/Pg boundary in tropical South America using a palynological dataset spanning the Maastrichtian–Paleocene interval, including 39 stratigraphic sections from outcrops and wells, 637 samples, 1048 taxa, and 53,029 occurrences (Fig. 1 and table S1) (18). As fossil pollen assemblages typically integrate information at large spatial scale (i.e., tens of square kilometers), we also examined the composition and diversity of autochthonous assemblages of leaf fossils, which instead reflect local forest communities. These included 2053 fossils from the Maastrichtian Guaduas Formation and 4898 fossils from the middle-late Paleocene Bogotá and Cerrejón formations (19). Situated near the paleo-equator, this then-coastal region of northern South America was wet and megathermal throughout the globally warm Maastrichtian and Paleocene. As a result, the effect of the end-Cretaceous event on the fossil record is not confounded by major changes in climate.

Extinction and turnover of tropical vegetation

We estimated diversity using the corrected sampled-in-bin diversity (20), the shareholder quorum subsampling (SQS) (21), origination and

¹Smithsonian Tropical Research Institute, Panama. ²Grupo de Investigación Paleontología Neotropical Tradicional y Molecular (PaleoNeo), Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario, Bogotá, Colombia. ³ISEM, U. Montpellier, CNRS, EPHE, IRD, Montpellier, France. ⁴Department of Geology, Faculty of Sciences, University of Salamanca, Salamanca, Spain. ⁵Instituto Colombiano del Petróleo, Bucaramanga, Colombia. ⁶Negaunee Institute for Plant Conservation, Chicago Botanic Garden, Chicago, IL, USA. ⁷Department of Paleobiology, National Museum of Natural History, Washington, DC, USA. ⁸Soil and Water Science Department, University of Florida, Gainesville, FL, USA. ⁹Faculdade de Geociências, Universidade Federal de Mato Grosso, Cuiabá, Brazil. ¹⁰ExxonMobil Corporation, Spring, TX, USA. ¹¹Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales, CCT-CONICET, Mendoza, Argentina. ¹²Departamento de Geología, Universidad de Chile, Santiago, Chile. ¹³Department of Entomology, University of Maryland, College Park, MD, USA. ¹⁴College of Life Sciences, Capital Normal University, Beijing, China. ¹⁵Corporación Geológica Ares, Bogotá, Colombia. ¹⁶Paleoflora Ltda, Zapotoca, Colombia. ¹⁷Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA. ¹⁸Instituto Amazónico de Investigaciones Científicas SINCHI, Leticia, Colombia. ¹⁹Departamento de Ciencias Forestales, Universidad Nacional de Colombia, Medellín, Colombia. ²⁰Department of Geosciences, Boise State University, Boise, ID, USA. ²¹BP Exploration Operating Company Limited, Chertsey Road, Sunbury-on-Thames, Middlesex, UK. ²²Department of Biology, University of Fribourg, Fribourg, Switzerland. ²³Department of Biological and Environmental Sciences, University of Gothenburg and Gothenburg Global Biodiversity Centre, Gothenburg, Sweden. *These authors contributed equally to this work. †Corresponding author. Email: jaramilloc@si.edu

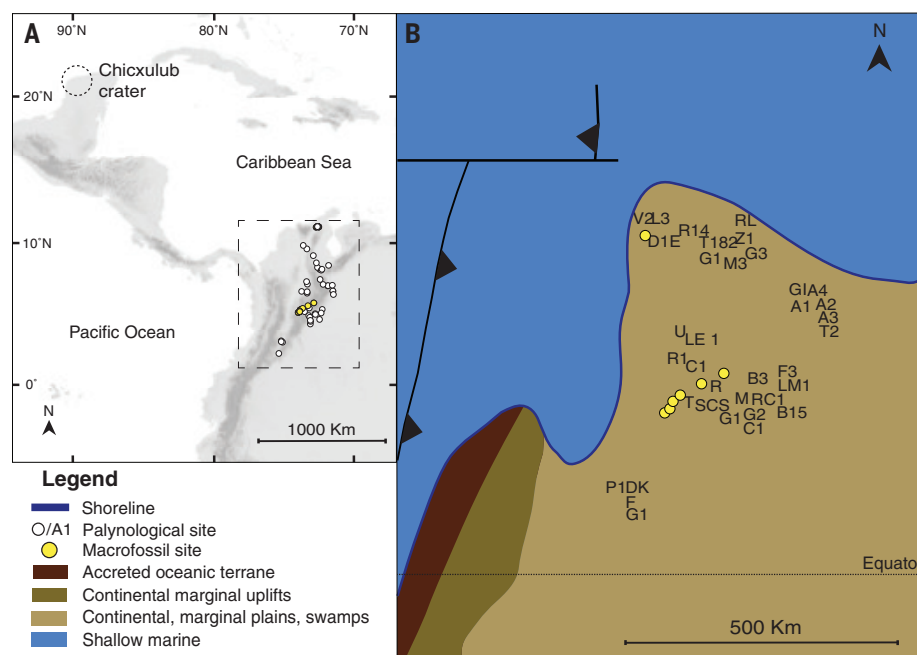


Fig. 1. Location of stratigraphic sections and macrofossil localities in northern South America. (A) Map showing modern-day distance to Chicxulub crater. (B) Paleogeographic reconstruction of northern South America [area delimited by dotted rectangle in (A)] during the late Maastrichtian, based on (64).

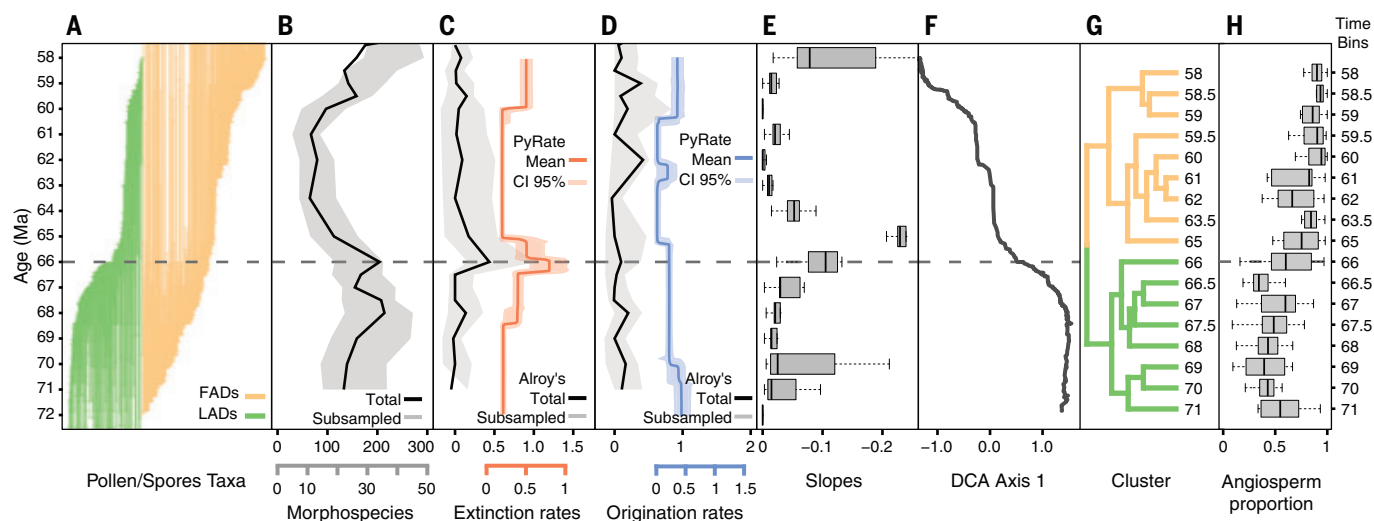


Fig. 2. Changes in diversity and composition of Maastrichtian-Paleocene palynofloras in northern South America. (A) Stratigraphic ranges of taxa across the Maastrichtian-Paleocene interval. Shown in green are the taxa that became extinct and in orange, the taxa that originated during this time period. (B) Corrected sampled-in-bin diversity. (C) PyRate (23) extinction rate mean and 95% credible interval (orange shadow) and Alroy's second-for-third (22) extinction rate. (D) PyRate origination rate mean and 95% credible interval (blue shadow), and

Alroy's second-for-third origination rates with 0.95 confidence interval (SQS = 0.95; gray shadow). (E) Boxplot of slopes from the survivorship analysis performed on 1-million-year bin cohorts. (F) Change in floral composition shown by scores of samples on DCA axis 1 plotted against time. (G) Sørensen Cluster showing two distinct clusters, Maastrichtian (green) and Paleocene (orange); see fig. S1 for individual samples cluster. (H) Boxplot of the proportion per bin of angiosperm grains versus total flora; see fig. S2 for proportion of individual samples.

extinction rates using the second-for-third method (22), and PyRate (23) [see materials and methods (18)]. Palynofloral diversity was higher in the Maastrichtian (72 to 66 Ma) in tropical South America than in the early and middle Paleocene (66 to 60 Ma) (Fig. 2B; mean of Maastrichtian bins 172.3 versus Paleocene bins 84.1, t test, $df = 10.9$, $P < 0.001$; table S3), regardless of differences in sampling size (Fig. 2B, SQS estimates 27 versus 12.7, t test, $df = 3872.7$, $P < 0.001$) or depositional environments (table S4). This marked decrease in diversity coincides with a peak in extinction rates at 66 Ma (66 to 66.5 age bin; log Bayes factors >6 with a 95% credible interval between 66.4 and 65.7) that diminishes palynomorph diversity by 45% and significantly exceeds Maastrichtian or Paleocene background extinction (Fig. 2C, extinction rate of 0.44 versus a mean of 0.04 for all other bins, SQS estimates 0.53 versus 0.03, t test, $df = 309.08$, $P < 0.001$; PyRate extinction rate 0.75; credible interval (CI): 0.45 to 1 versus a median rate of 0.07; CI: 0.04 to 0.09 in the Maastrichtian and 0.05; CI 0.03 to 0.07 in the early Paleocene). As a result, most Maastrichtian cohorts (groups of palynomorphs that coexist at a given time) decline in the first bin of the Paleocene (65 Ma bin), well above the mean cohort reduction observed throughout the Paleocene (Fig. 2E, mean slope of all cohorts at 65 bin 0.24 versus mean slope of all other cohorts, 0.05, t test, $df = 13$, $P < 0.001$).

Following the K/Pg boundary, palynomorph diversity did not recover to pre-extinction

levels until after 60 Ma (Fig. 2B) and further increased beyond pre-extinction levels throughout the Paleocene–Eocene Thermal Maximum and early Eocene (24, 25). The second-for-third estimates identify a peak in origination during the 59- to 59.5-Ma interval (Fig. 2D; mean origination rate 0.38, mean at all other intervals 0.09; SQS estimates 0.95, 0.28 versus 0.08, t test, $df = 332.54$, $P < 0.001$), whereas PyRate found support (log Bayes factors >6) for a drop in origination in the earliest Paleocene (from 0.23; CI: 0.2 to 0.27 to 0.04; CI: 0.01 to 0.08) and a strong increase between 60.7 and 60.2 Ma (rate 0.37; CI: 0.28 to 0.47). A reanalysis of the data allowing the PyRate algorithm to search for rate shifts at a higher temporal resolution resulted in similar patterns of origination and extinction overall (fig. S3). However, the analysis detected an additional brief but strong peak in origination rates between 59.6 and 59.2 Ma, when the origination rates increased from 0.26 (CI: 0.18 to 0.38) to 1.30 (CI: 0.90 to 1.72).

We used detrended correspondence analysis (DCA) and cluster analysis to evaluate changes in palynofloral composition across the K/Pg boundary. Rapid change through time in the first axis scores of samples (Fig. 2F, first axis explains 57% of variation) and a distinct clustering of Maastrichtian and Paleocene plant communities (Fig. 2G and fig. S1) reflect a major and permanent change in floristic composition. Although the Maastrichtian contained roughly equal proportions of angiosperm (47.9%) and spore grains (49.5%), angiosperm grains dominated in the Paleocene (mean abun-

dance 84% versus 16% of non-angiosperms, Wilcoxon test, W -statistic = 14,552, $P < 0.001$; fig. S2). Gymnosperms (mostly Araucariaceae) are 2.5% of Maastrichtian grains but only 0.4% of Paleocene grains (Mann-Whitney test, U -statistic = 17,509, $P < 0.01$). Gymnosperms also occur in 75% of Maastrichtian samples but only in 24% of Paleocene samples having >100 grains. Sediments from the Maastrichtian Umir Formation (central Colombia) are rich in gymnosperm lipid biomarkers (26), supporting the abundance of gymnosperms prior to the K/Pg extinction. Living species of Araucariaceae occur as large trees and are often underrepresented in the palynological soil record and do not disperse long distances (27), such that their low abundance in Maastrichtian deposits is likely to be an underestimation of their true abundance.

Leaf physiognomy and forest types

We recognize 41 angiosperm and 4 fern morphotypes in the Maastrichtian Guaduas macroflora. In the Paleocene, we found 46 angiosperms and 2 ferns in the Bogotá flora and 58 angiosperms, 5 ferns, and 1 conifer leaf morphotype in the Cerrejón flora. The foliar physiognomy of nonmonocot angiosperm leaves (ANA-grade angiosperms: Amborellales, Nymphales, and Austrobaileyales; magnoliids; and eudicots) in both the Maastrichtian and Paleocene assemblages resembles that of modern tropical rainforests, characterized by leaves of large size, untoothed margins, and elongated drip tips (Fig. 3). Of the 36 species of non-

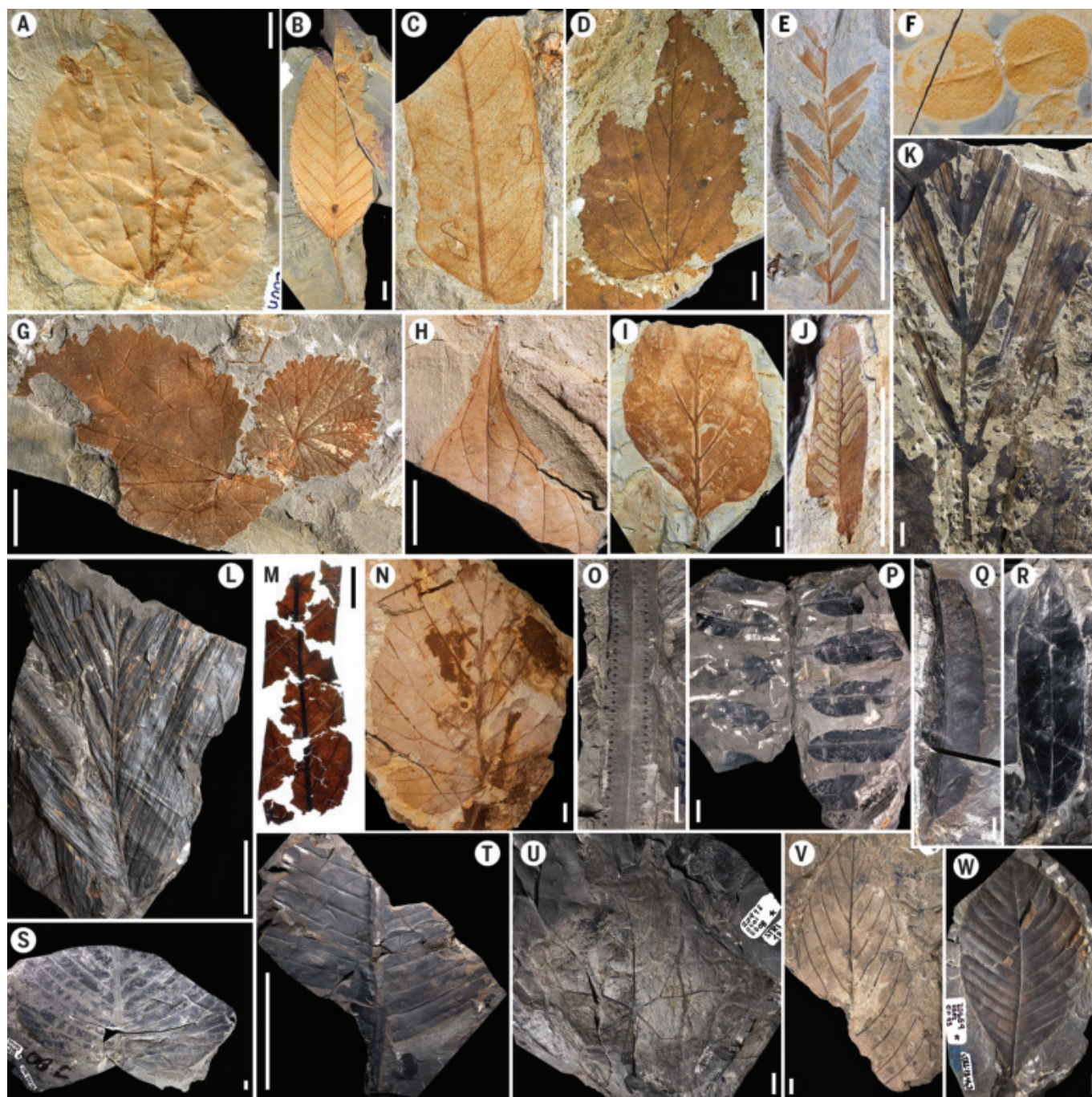


Fig. 3. Representative leaf taxa. (A to K) Taxa from Paleocene Bogotá and (L to W) Maastrichtian Guaduas floras. (A) Menispermaceae (BF6). (B) Salicaceae (BF5) with midrib gall. (C) Fabaceae leaflet (BF38) with surface feeding damage. (D) Euphorbiaceae (BF37) with hole and margin feeding. (E) Fabaceae, Caesalpinioideae (BF21). (F) Water fern, *Salvinia bogotensis*, Salviniaceae (BF22). (G) *Malvaciphyllum* sp. Malvaceae (BF4). (H) Example of drip tip in Salicaceae (BF23). (I) aff. Eleocarpaceae (BF13). (J) Fabaceae leaflet

(BF21, 5 mm) with hole feeding damage. (K) Arecaceae (BF27). (L) Arecaceae (GD47, 10 cm). (M) aff. Lauraceae (GD54). (N) aff. Hamamelidaceae (GD56). (O and P) Fertile and sterile fragments of Polypodiaceae (GD22). (Q) aff. Salicaceae (GD6). (R) Lauraceae (GD7) with drip tip. (S) aff. Urticaceae (GD52). (T) Zingiberales (GD46, 5 cm). (U) aff. Cucurbitaceae (GD8). (V) *Bernhamniophyllum* sp. Rhamnaceae (GD1). (W) aff. Dilleniaceae (GD3). Scale bars: 1 cm except where noted in parentheses after taxon.

monocots in the Guaduas flora, 89% have leaves larger than 45 cm² (mesophylls), 81% have untoothed margins, and 11 of the 25 species with preserved apices have drip tips (44%). In the Paleocene assemblages, 63 and 76% of nonmonocot species have untoothed margins

(Bogotá and Cerrejón, respectively), and 30 to 35% have elongated drip tips. Estimates of mean annual rainfall based on Leaf Area Analysis (18, 28, 29) indicate annual precipitation of 234 to 293 cm year⁻¹ for the Guaduas flora, 182 to 184 cm year⁻¹ for the Bogotá flora, and

240 to 308 cm year⁻¹ for the Cerrejón flora (Table 1 and table S8).

Leaf mass per area (LMA) values, estimated on the basis of the scaling relationship between leaf mass and petiole diameter observed in living plants (30), were consistent with modern

Table 1. Leaf physiognomy and precipitation of the Maastrichtian-Paleocene floras. Numbers in parentheses indicate numbers of quarries (Total specimens), number of census localities (Census), and number of morphotypes with preserved apices (drip tips). MAP, mean annual precipitation.									
Formation	Age	Total specimens	Census numbers	Leaf taxa	Nonmonocot taxa	Non-monocots with entire margins	Nonmonocots with drip tips	Leaves mesophylls or larger	MAP (cm year ⁻¹)
Guaduas	Maastrichtian	2053 (12)	1650 (2)	45	36	29 (81%)	11 (25)	32 (89%)	234–293
Bogotá	Paleocene	2416 (19)	1370 (1)	48	40	25 (63%)	6 (20)	25 (63%)	182–184
Cerrejón*	Paleocene	2482 (18)	1190 (2)	65	46	35 (76%)	12 (34)	44 (68%)	240–304

*Data reported by (19).

evergreen rainforest environments across all three floras (Guaduas: 36 to 206 g m⁻²; Bogotá: 52 to 206 g m⁻²; Cerrejón: 44 to 126 g m⁻²), yet LMAs of the Guaduas and Bogotá floras are lower than those of Cerrejón (*t* test, *P* < 0.001, tables S9 and S10) (18). Evergreen trees tend to have higher LMAs when living under drier climates (31), which is consistent with the relatively lower precipitation of the Bogotá flora compared with Cerrejón. The Guaduas and Cerrejón floras had similar precipitation (>200 cm year⁻¹), so it is possible that the higher Guaduas LMA may reflect a higher irradiance related to canopy structure (see below) or poorer soils (31).

A notable feature of the Paleocene Cerrejón flora is its resemblance to modern Neotropical rainforests in terms of family-level composition of angiosperms (19). To examine this, we compared the natural affinities of leaf taxa in the Guaduas with those at Bogotá, Cerrejón, and modern Neotropical rainforests. Some Maastrichtian angiosperms have confirmed or tentative affinities to families that are widely distributed in (but not necessarily restricted to) the lowland tropics, including Lauraceae (two or three morphotypes), Araceae (two morphotypes), Theaceae (one or two morphotypes), Arecaceae, Rhamnaceae (32), Piperaceae (33), Salicaceae, Canellaceae, Dilleniaceae, Urticaceae, and Monimiaceae (one morphotype each), among others (Fig. 4A, table S6, and data S4). The flowering plants of the Paleocene Bogotá flora closely resemble those of the Cerrejón flora (19) and include the dominant tree families in modern Neotropical rainforests. The Bogotá flora has two leaflet types of Fabaceae, one of these representing the earliest record of Caesalpinioideae (with abundant legume pods) (34), Euphorbiaceae, Lauraceae, Salicaceae, Violaceae (two morphotypes each), Malvaceae, Melastomataceae (35), Rhamnaceae, Arecaceae, Eleoocarpaceae, and Araceae (one morphotype each; table S7 and data S5). Fossil seeds of Annonaceae, Icacinaceae, Menispermaceae, and Passifloraceae are also present in the Bogotá flora. Because nearly autochthonous leaf assemblages reflect tree biomass as a combination of stem abundance

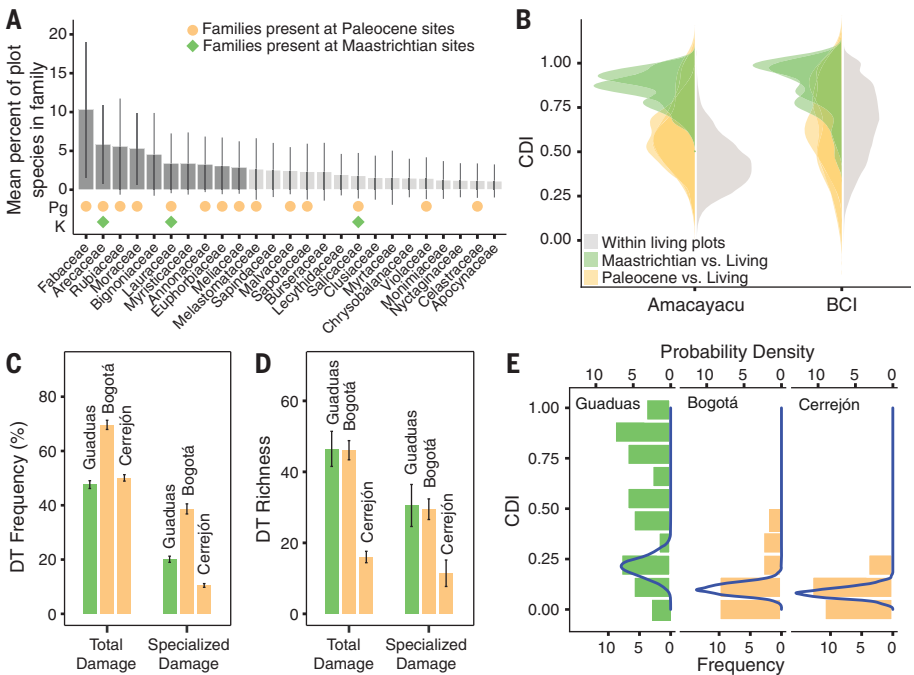


Fig. 4. Forest composition and insect-feeding damage of fossil floras. (A) Percentage of tree species in 72 extant Neotropical forest plots that belong to the 25 plant families that together account for 75% of diversity. Half of the stems belong to the 10 families shown as dark bars. Orange circles indicate families present in the Bogotá or Cerrejón floras (Paleocene), and green diamonds indicate families present in the Maastrichtian Guaduas flora. (B) Density plot of dissimilarity in family composition between fossil assemblages and samples of living Neotropical forests (see materials and methods for details). Chao-Sørensen dissimilarity (CDI) was calculated between randomly selected subregions of the 50-ha plot at Barro Colorado Island (Panama) and the 25-ha plot at Amacayacu (Colombia). Gray areas depict the distribution of dissimilarities of the randomly selected subregions within each site. (C) Average frequency of damaged leaves in 400 randomly selected leaves from each fossil flora. (D) Richness of total and specialized insect-mediated damage types, rarefied to 95 and 90% sample coverage, respectively. Gray lines indicate 95% confidence intervals. (E) Histogram of leaf damage beta-diversity across host plant species with more than 20 leaves at the Guaduas, Bogotá, and Cerrejón floras. Pairwise beta-diversity was quantified using CDI and is depicted in solid bars. The blue curve indicates the probability density for the null expectation that the observed DTs are randomly distributed across host plant species.

and stem diameter (36), we compared the family-level composition of five unbiased census sites (two from Guaduas, one from Bogotá, and two from Cerrejón) with permanent plots in two living Neotropical rainforests: Barro Colorado Island (BCI), Panama (37), and Amacayacu, Colombia (38). In living tropical

rainforests, samples of leaf litter that are analogous to single fossil quarry sites can represent most of the standing vegetation (90% biomass) in a 12.5-m radius (36). The Paleocene census sites are more similar in family composition to the living forest at BCI (Fig. 4B; Wilcoxon test, *W* = 46882, *P* < 0.001) and Amacayacu

(Wilcoxon test, $W = 7806$, $P < 0.001$) than they are to the Maastrichtian census sites (fig. S5).

Canopy structure is reflected in the distribution of leaf vein length per area (VLA) and stable carbon isotope ratios ($\delta^{13}\text{C}$) within individual taxa (39, 40). Most nonmonocots from Guaduas have relatively high VLA values (39, 41), yet the unimodal distribution of VLA within single taxa in the Guaduas flora (39) and the low range of $\delta^{13}\text{C}$ measured in leaf cuticles (40) suggest that these forests did not have the range of light environments seen in modern multistratal rainforests. By contrast, leaves of the Paleocene Cerrejón flora show the same bimodal distribution of single-taxon VLA and the wide range of cuticle $\delta^{13}\text{C}$ observed in modern closed canopy, multistratal forests (39, 40). Maastrichtian wet tropical forests, therefore, likely had an open canopy that promoted mixing of respired and atmospheric CO_2 and a small light gradient between the understory and the canopy compared to modern Neotropical forests. These open canopy forests may have recycled less rainfall through transpiration than their multistratal Paleocene equivalents, potentially influencing regional and global climate (42).

Diversity of plant-insect interactions

The diversity of insect-feeding damage on leaves reflects the richness of insect herbivores (43). We quantified insect damage in the Guaduas and Bogotá floras following a standard damage type (DT) system and compared it with damage from the Paleocene Cerrejón flora (18). Over 50% of leaves in all three floras show insect herbivory (Fig. 4C), indicating intense biotic interactions in both Maastrichtian and Paleocene forests. The richness of insect DTs in the Guaduas flora is comparable to that at Bogotá and greater than at Cerrejón, both for total DT richness resampled at 95% coverage (Guaduas versus Bogotá: 46.5 versus 47.8 DTs, t test one-tailed, $t = 0.325$, $\text{df} = 579$, $P = 0.745$; Guaduas versus Cerrejón: 46.5 versus 16.0 DTs, t test one-tailed, $t = 6.23$, $\text{df} = 483$, $P < 0.001$) and for specialized damage only (at 90% coverage: Guaduas versus Bogotá: 30.53 versus 29.48 DTs, t test, $t = 0.20$, $\text{df} = 802$, $P = 0.841$; Guaduas versus Cerrejón: 30.53 versus 11.43 DTs, t test, $t = 5.09$, $\text{df} = 1121$, $P < 0.001$; Fig. 4D).

Because insect-feeding damage reflects inflicting herbivores, DT beta-diversity across host species provides evidence of host specificity among insect herbivore communities. Leaf damage beta-diversity across host taxa in the Maastrichtian Guaduas flora is higher than expected by chance (Wilcoxon test, $W = 41615$, $P < 0.001$) and higher than that observed at either Bogotá or Cerrejón (Guaduas versus Bogotá: Wilcoxon test, $W = 1322$, $P < 0.001$; Guaduas versus Cerrejón: Wilcoxon test, $W = 1410$, $P < 0.001$) (Fig. 4E). This distribution of DTs in the Guaduas flora suggests greater her-

bivore community specificity than at either Paleocene site.

The end-Cretaceous shaped modern Neotropical rainforests

Prior to the end-Cretaceous, Neotropical rainforests had relatively open canopies; contained a mixture of angiosperms, ferns, and conifers (mostly Araucariaceae); and suffered intense and host-specific insect herbivory. Paleocene forests, by contrast, were more similar to modern Neotropical rainforests in having closed, multistratal canopies, biomass dominated by angiosperms, and a similar plant family composition. Yet, Paleocene rainforests were less diverse than Maastrichtian, Eocene, or modern rainforests (19), and the low plant diversity seen throughout the Paleocene shows a long lag in the recovery of diversity following the P/Kg event.

The differences between Maastrichtian and Paleocene forests in floral composition and canopy structure, but similar leaf physiognomy, denote two fundamentally distinct ecosystems that developed under the same wet, tropical climate. Because of their open canopies, lower angiosperm abundance, and a constant, albeit minor, presence of conifers, Maastrichtian rainforests may have been accompanied by slower rates of carbon fixation, transpiration, and nutrient cycling when compared to Paleocene rainforests. In addition, the development of closed canopy rainforests in the Paleocene would have created stronger vertical gradients in light and water use, providing opportunities for new plant habit and growth forms and leading to the vertical complexity seen in modern rainforests.

These notable differences raise two questions: (i) Why did Maastrichtian rainforests lack a closed canopy? By the Late Cretaceous, angiosperms were taxonomically and ecologically diverse (44, 45) and had evolved a wide range of growth habits, ranging from aquatic plants to large trees (45, 46), making it unlikely that they were inherently unable to form a closed canopy. (ii) Why did Paleocene rainforests establish a different plant community composition and structure instead of returning to the Maastrichtian-like rainforests? This is particularly perplexing given the similarity in Paleocene and Maastrichtian climates.

We offer three, non-mutually exclusive explanations for the observed pattern. One is disturbance by large herbivores. Sustained trampling and extensive feeding by large herbivores, mostly dinosaurs (47), could have maintained an open canopy by reducing competition for light among neighboring plants through continuous habitat disturbance and gap generation. Such pervasive disturbance could explain the abundance of ferns in Maastrichtian palynofloras, as they typically thrive in successional vegetation (48). The extinction of large

herbivores at the end-Cretaceous would have reduced gap formation, triggering a “race for light” among tropical plants, and creating more shaded habitats in which a wider variety of light and growth strategies could succeed (49). A second explanation involves soil nutrients. Extensive and stable lowlands developed in northern South America during the Maastrichtian (50), with a persistent humid climate over millions of years. Maastrichtian forests therefore must have grown on strongly weathered soils characterized by extreme infertility (51) with nutrient limitation of growth exacerbated by the high CO_2 concentrations and associated high water-use efficiency that reduces nutrient uptake by mass flow (52, 53). These low-nutrient conditions would have promoted an open canopy structure by favoring the conifers, which in modern tropical forests are typically associated with infertile soils (54). Ashfall from the Chicxulub impact added weatherable phosphorus minerals to terrestrial ecosystems worldwide (55), instantly resetting fertility to the high-phosphorus, low-nitrogen period that characterizes young stages of ecosystem development (51). This set the stage for the diversification of nitrogen-fixing taxa in the Fabaceae, whose rise in the Paleocene (34) would have increased soil fertility, stimulated forest productivity (56), and enhanced the relative advantage of high-growth-rate angiosperms over conifers and ferns (57, 58). These proposed changes in nutrient cycling could be tested by analyzing paleosol composition and isotopic signatures across the Maastrichtian-Paleocene interval. A third explanation of the observed pattern concerns selective extinction. Although the Araucariaceae were not diverse, they could have been important in structuring the Late Cretaceous canopy environment (59). Lineages with narrow ecological ranges and tree growth forms such as Araucariaceae are particularly susceptible to mass extinction events (60). By contrast, high ecological diversity within Maastrichtian angiosperm lineages (44, 45) may have made them more resistant to extinction (60), as might their higher capacity for whole-genome duplication (61–63). The near disappearance of conifer trees from tropical rainforest canopies at the end of the Cretaceous may have released resources upon which the modern angiosperm canopy-forming lineages diversified during the Paleocene. This scenario could be tested by assessing shifts in diversification rates across the K/Pg of Neotropical canopy trees, epiphytes, and lianas.

Although there is still much to be learned about the Cretaceous and Paleocene tropical forests, the changes described here show that the end-Cretaceous event had profound consequences for tropical vegetation, ultimately enabling the assembly of modern Neotropical rainforests. It is notable that a single historical

the interfacial symmetry, we designed a vdW interface with an in-plane electronic polarization. We observed the emergence of a controllable spontaneous photovoltaic effect (SPE)—photo-induced spontaneous current in noncentrosymmetric crystals (9) without a semiconductor p-n junction or bias voltage. In our system, the SPE appeared along the polar direction but was absent in the direction perpendicular to the polarization.

The SPE has attracted increasing interest not only as a new principle of photovoltaic devices but also for fundamental studies of its inherent nature and intrinsic mechanism associated with the energy band topology or geometry such as Berry curvature (dipole) or Berry connection (10–15). Investigations of the SPE in bulk polar crystals, including oxide materials (16–20), organic polar crystals (21), and halide compounds (22–24), and also in individual vdW bulk crystals and flakes (25–28) suggest that its emergence is closely related to symmetry reduction that creates polar symmetry. Characteristics of the SPE that we observed were also well described and understood by the polarity-induced geometrical shift current mechanism at the vdW heterointerface without commensurability.

We chose WSe₂ and black phosphorus (BP) as the building blocks of the interface because each compound has distinct rotational and mirror symmetries. WSe₂ has threefold rotational symmetry and mirror planes exist along the armchair direction, and BP has twofold rotational symmetry and sets of mirror planes (the rotational axes and mirror planes of each crystal are shown by the circled dots and green lines, respectively, in Fig. 1A). However, the heterointerface of WSe₂ and BP has no rotational symmetry because the threefold and twofold rotational symmetries are not compatible, although mirror symmetry can still remain if the mirror planes of both WSe₂ and BP are parallel. For an interface with only one mirror plane, electronic polarization would appear along the direction parallel to the mirror plane, and the resulting photocurrent generation would be expected along the in-plane polar direction.

In the WSe₂/BP interface (Fig. 1B), a stripe moiré pattern appears both along the polarization direction (parallel to the mirror plane) and vertical to the polarization direction (perpendicular to the mirror plane), unlike the hexagonal moiré patterns in twisted graphene (7). This stripe pattern results from the lattice mismatch between WSe₂ and BP and reflects both the trigonal symmetry of WSe₂ and the anisotropic potential of BP that induced the in-plane polarity at this interface. Monolayer WSe₂ transferred on BP with a thickness of around 40 to 50 nm was illuminated with a laser from the monolayer WSe₂ side (Fig. 1C; see supplementary materials for details). The room-temperature current-voltage (*I*-*V*) characteristic of the WSe₂/BP interface under dark conditions (black) or under linearly polarized light (green) is shown in Fig. 1D. A typical short circuit current under zero bias, or spontaneous photocurrent, was observed after laser illumination with a wavelength of 532 nm and intensity of 1.44 mW but not under dark conditions.

In Figure 2, we systematically studied the spatial distribution of spontaneous photocurrent in monolayer WSe₂ (Fig. 2, A, D, and G), BP (Fig. 2, B, E, and H), and the WSe₂/BP interface (Fig. 2, C, F, and I). The photocurrent mapping images (Fig. 2, D to F) were obtained by scanning the laser spot across the devices shown in Fig. 2, A to C. The SPE in the monolayer WSe₂ (Fig. 2D) and BP device (Fig. 2E) was absent when the laser illuminated the center of the devices. The generated photocurrent appeared only around the electrodes in monolayer WSe₂ and BP, which showed an

antisymmetric spatial profile and could be attributed to the effect of the Schottky barriers, photothermal effects, or both (10, 25, 26).

In the WSe₂/BP stacking device, a photocurrent was observed even when the laser spots were far from the electrodes (Fig. 2F). This observation of the SPE only in the WSe₂/BP device indicated that it originated from the changes in the symmetry at the interface. The SPE appeared at the interface in several measured devices (devices 1 through 4; see supplementary materials). The SPE was also observed around the electrodes with antisymmetric spatial distribution, as shown in the position dependence of the photocurrent along the lines in Fig. 2C (Fig. 2I).

To clarify the intrinsic nature of the SPE at the WSe₂/BP interface, we measured the directional dependence of the photocurrent response. We fabricated the WSe₂/BP interface so that the mirror plane of WSe₂ was parallel to that of BP (device 2; Fig. 3A). The electrodes were patterned to be either parallel (E1 and E2) or perpendicular (E3 and E4) to the expected polar direction, and spatial maps of the spontaneous photocurrent are shown in Fig. 3, B and C, respectively. The SPE was observed only for the measurement geometry with the E1 and E2 electrodes; it was negligibly small within the background noise level for the E3 and E4 electrodes. These results indicated that the electronic polarization at the interface was almost parallel to the E1 and E2 electrodes (expected polar direction) and also excluded the possibility of the SPE originating from an extrinsic mechanism such as a randomly

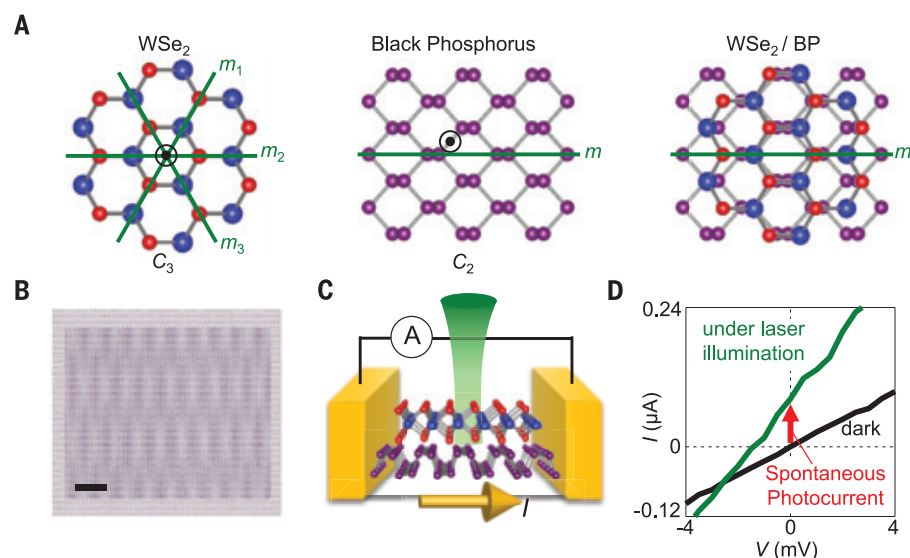


Fig. 1. Symmetry and moiré pattern of the vdW interface WSe₂/BP. (A) Schematic illustrations of monolayer WSe₂ (left), BP (middle), and a heterointerface of WSe₂/BP (right). Green lines and circled dots represent mirror planes and rotational axes, respectively. (B) Moiré patterns of the WSe₂/BP heterointerface for the case where the mirror planes of WSe₂ and BP are parallel. The scale bar (black line) represents 5 nm. (C and D) A schematic of the experiment (C) and *I*-*V* characteristic of the WSe₂/BP device (D). The spontaneous photocurrent is defined in Fig. 1D and has been measured in Figs. 2 and 3; more measurements are in the supplementary materials. The circled "A" in (C) represents the electrical system for photocurrent measurement.

¹Quantum-Phase Electronics Center (QPEC) and Department of Applied Physics, The University of Tokyo, Tokyo 113-8656, Japan. ²College of Engineering and Applied Sciences and National Laboratory of Solid-State Microstructures, Nanjing University, Nanjing, China. ³Department of Physics and Astronomy, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada. ⁴Quantum Matter Institute, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. ⁵Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. ⁶International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. ⁷Department of Physics, Case Western Reserve University, Cleveland, OH, USA. ⁸RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan.

*These authors equally contributed to this work.

†Corresponding author. Email: ideue@ap.t.u-tokyo.ac.jp

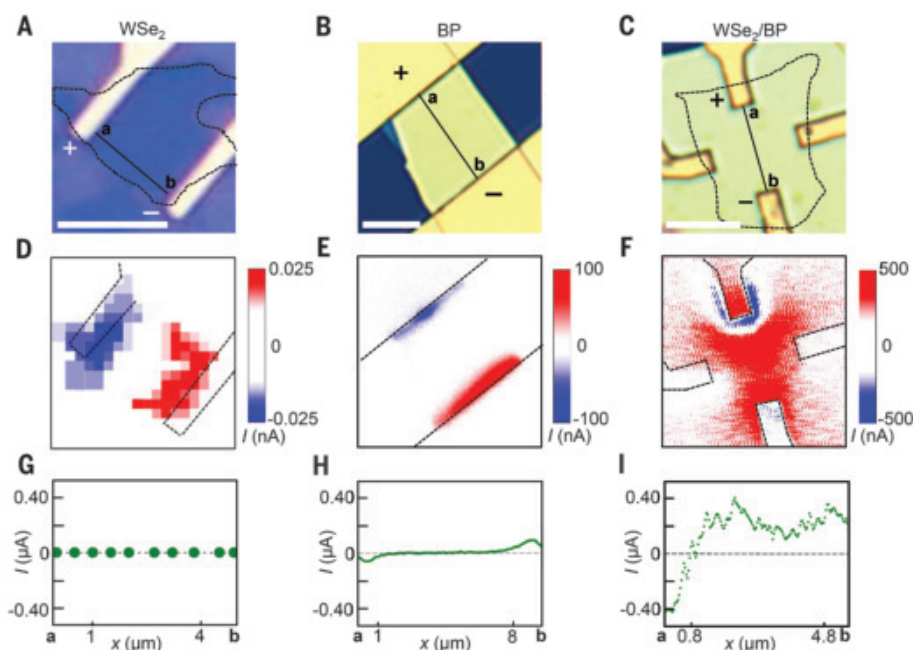


Fig. 2. Photovoltaic response in vdW crystals and interfaces with different symmetry. (A to C) Device pictures of monolayer WSe₂ (A), BP (B), and heterointerfaces of WSe₂/BP (device 1) (C). The scale bars (white lines) represent 5 μ m. Dashed lines in (A) and (C) indicate the edge of monolayer WSe₂ flake, and “+” and “–” indicate the definition of the sign of the generated photocurrent. The solid black lines ab in (A) to (C) indicate the positions corresponding to the positions in (G) to (I). (D to F) Photocurrent mapping for monolayer WSe₂ (D), BP (E), and WSe₂/BP (F) devices. Linearly polarized light of 532-nm wavelength was used. The SPE emerges at the center of the WSe₂/BP interface. (G to I) Position dependence of the photocurrent along the lines in (A) to (C).

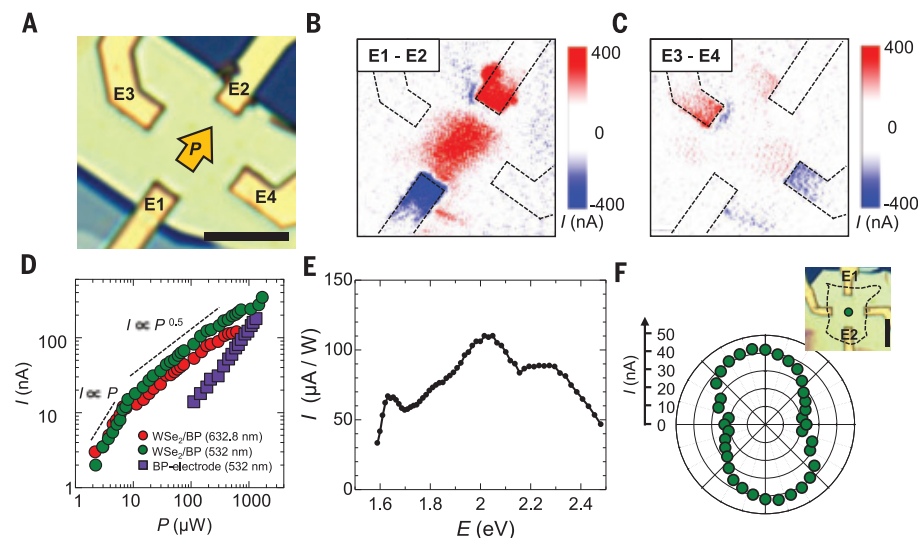


Fig. 3. Characteristics of the SPE in the WSe₂/BP interface. (A) Optical microscope image of the WSe₂/BP device (device 2). The scale bar (black line) represents 5 μ m. (B and C) Photocurrent mapping for the E1 and E2 electrodes (B) and the E3 and E4 electrodes (C). The SPE was observed along the E1 and E2 electrode direction, whereas it is absent along the E3 and E4 electrode direction. (D) Laser power (P) dependence of the photocurrent (I) for two different wavelengths of 632.8 nm (red circles) and 532 nm (green circles). In the low-power region, I is proportional to P , whereas the plot shows $I \propto P^{0.5}$ power dependence in the high-power limit. On the other hand, photocurrent originating from the Schottky barrier at the electrodes of the BP device (purple squares) shows the linear power dependence. (E) Photon energy dependence of the photocurrent (device 2). (F) Polar plot of I represented by green dots as a function of the linear polarization angle along the E1 and E2 electrodes for the WSe₂/BP device (device 1). The inset shows the optical microscope image of the measured device, and the green dot within it represents the position of the laser spot.

formed p-n like potential [such as the Dember effect (29)] or the anisotropy of BP. In the supplementary materials, we also present measurements of photocurrent in twisted interfaces and clarify that the direction or magnitude of the photocurrent can be tuned by stacking angle.

In Fig. 3D, we show the laser-power dependence of the spontaneous photocurrent for two different wavelengths. Both measurements show the characteristic crossover from linear to square-root dependence, an effect that could not be explained by the photovoltaic effect with Schottky barrier near the electrodes, which normally shows the linear power dependence (purple squares in Fig. 3D). Among other possible mechanisms for the SPE, the shift current mechanism, a photovoltaic effect in inversion-broken crystals that arises from the shift of the electron wave packet upon photoexcitation, is a plausible origin because it predicts such a power dependence crossover caused by a saturation of the carrier excitations (14).

To further understand the microscopic origin of the SPE, we studied the dependence of the bulk photovoltaic effect in the WSe₂/BP interface (device 2; Fig. 3E) with photon energy E with incident linear polarization parallel to the armchair direction. The photocurrent had peaks around E of 1.65 and 2.05 eV that seemed to correspond to the exciton peaks of WSe₂. This result indicated that exciton resonance could enhance the bulk photovoltaic response as theoretically predicted (30, 31). Although the effect of excitons on the shift-current mechanism requires further investigation, these results implied that two-dimensional materials, in which excitons are stable and have a long lifetime, could enable a large photovoltaic response. In a higher-energy region ($E > 2.1$ eV), nonmonotonic changes were also observed (Fig. 3E).

The polar diagram of the photocurrent as a function of the linear polarization angle (Fig. 3F) showed no sign change when we rotated the polarization direction and was anisotropic along the armchair direction, which is the polar direction of the interface. This result indicated that the photocurrent was finite even if the incident light was unpolarized. The photovoltaic response is, in principle, allowed along the specific direction for the linearly polarized light in noncentrosymmetric trigonal WSe₂ (9), but distinct polarization angle dependence, including the sign change, was expected and photocurrent should vanish under unpolarized light in trigonal crystals. Thus, the present results cannot be explained by the bulk photovoltaic effect in trigonal crystals (in this case, the WSe₂ monolayer).

We could explain the observed photon-energy dependence in the high-energy region and polarization angle dependence with a shift current model. We used an effective tight-binding model

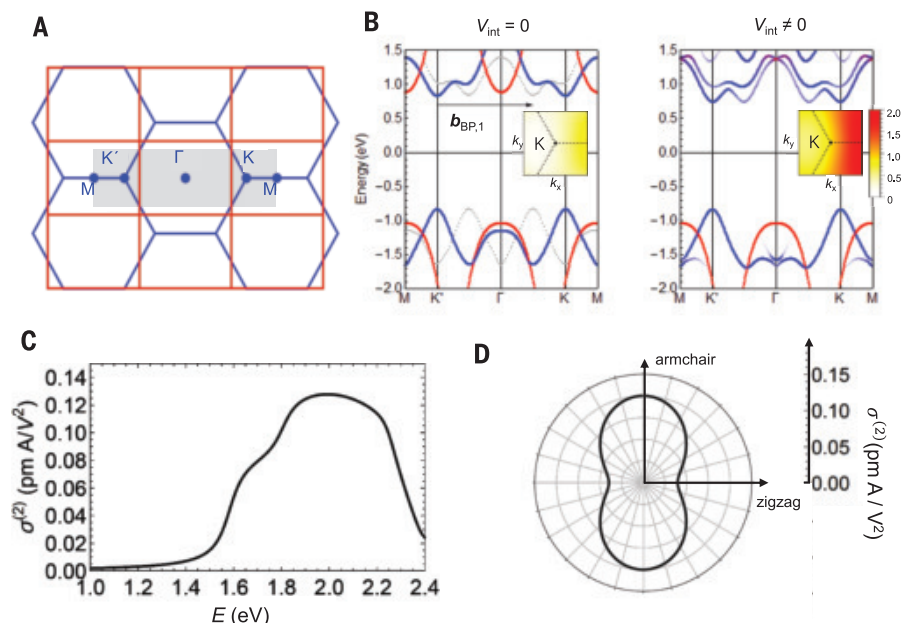


Fig. 4. Tight-binding model of the WSe₂/BP interface and its shift current. (A) BZs of WSe₂ (blue) and BP (red). (B) Energy dispersion of the tight-binding model of the WSe₂/BP interface. Interlayer coupling hybridizes the energy bands from WSe₂ and BP. The weight of the bands is color coded, with blue and red for WSe₂ and BP, respectively. The inset shows a color plot of the magnitude of the shift vector around the K point. (C) Nonlinear conductivity σ_{yyy} and σ_{yxx} from the shift current mechanism. (D) The calculated shift current along the armchair direction as a function of the linear polarization angle ($E = 2.15$ eV) showed no sign change and anisotropy along the polar direction.

of the WSe₂/BP interface to calculate electronic polarization (see supplementary materials) and shift current as follows: We consider tight-binding models of WSe₂ and BP layers and introduce effective coupling between the two layers. Specifically, the low-energy effective theory of the WSe₂/BP interface is described by the Hamiltonian

$$H(\mathbf{k}) = \begin{pmatrix} H_{\text{WSe}_2}(\mathbf{k}) & 0 & V_{\text{int}}(\mathbf{k}) \\ 0 & H_{\text{WSe}_2}(\mathbf{k} + \mathbf{b}_{\text{BP},1}) & V_{\text{int}}(\mathbf{k} + \mathbf{b}_{\text{BP},1}) \\ V_{\text{int}}^\dagger(\mathbf{k}) & V_{\text{int}}^\dagger(\mathbf{k} + \mathbf{b}_{\text{BP},1}) & H_{\text{BP}}(\mathbf{k}) \end{pmatrix}$$

where H_{WSe_2} and H_{BP} are Hamiltonians for WSe₂ and BP, \mathbf{k} is the momentum, V_{int} is the interlayer coupling, and $\mathbf{b}_{\text{BP},1} \parallel \mathbf{e}_x$ is the primitive reciprocal lattice vector of BP.

This V_{int} caused the band hybridization between WSe₂ and BP that led to the spontaneous electronic polarization and the resultant shift current, as discussed below (for details, see supplementary materials). We considered the Bloch bands of WSe₂ and BP and hybridization between them at the same momentum \mathbf{k} in the extended Brillouin zone (BZ) (Fig. 4, A and B). Although the BZ of the interface was not well defined because of the mismatch between the unit cells of WSe₂ and BP (Fig. 4A), we could still consider the Bloch bands of WSe₂ and BP and hybridization between them at the same momentum \mathbf{k} in the extended BZ.

Once the momenta \mathbf{k} for the two layers in the extended BZ picture were folded back into their respective first BZs, coupling appeared between Bloch states with different momenta because the two layers have different reciprocal vectors. Assuming that the interlayer coupling decayed quickly as a function of the relative position \mathbf{r} , it was sufficient to keep only a few Bloch states from each layer, which resulted in the above Hamiltonian. This treatment led to the band structure shown in Fig. 4B, where bands from WSe₂ and BP are hybridized with each other (the weight of the bands is color coded). Note that the momentum was no longer a good quantum number, and the band had a weight (less than 1) at a particular momentum \mathbf{k} in general.

We used this model to calculate the generation of the shift current. The polarization of photoexcited electron-hole pairs was quantified by the shift vector that was defined by the difference of Berry connection between valence and conduction bands (for details, see supplementary materials). The shift vector was substantially enhanced by the interlayer coupling between WSe₂ and BP (inset of Fig. 4B), and the shift current was enhanced by the hybridization of the WSe₂ and BP bands, which can be seen in the nonlinear conductivity calculated from the formula for the shift current (Fig. 4C). In the supplementary materials, we show the detailed derivation of the coupling Hamiltonian,

calculation of the shift current, and another simpler effective two-band model by which nonvanishing polarization and photocurrent can be intuitively understood. Theoretical calculation of the electronic polarization at this interface, which only exists along the mirror planes under finite interfacial coupling, is also displayed in the supplementary materials.

The observed photon energy dependence is well reproduced by this calculation and can be understood as follows: In the low-energy region ($E < 2.1$ eV), optical transition between bands around K (K') points (originally WSe₂ bands) mainly contributed to the shift current. With increasing photon energy, the shift current was initially enhanced because of the increase in the density of states. It was eventually suppressed in the higher-energy region because of the reduction of shift vector. The shift current rapidly diminished when optical transitions from the different valence band (originally BP band) were allowed, which contributed a shift current with the opposite sign that resulted in partial cancellation of the signal (see supplementary materials for details).

In the above argument, we considered the interface of 1L-WSe₂ and 1L-BP. Further calculations confirmed that the magnitude and qualitative behavior of the photon energy dependence of the shift current was almost independent of BP thickness. Also, we experimentally clarified that the photocurrent value was not affected by BP thickness, indicating that the SPE was mainly driven by the in-plane polarity at the interface and that the thick bulk part of BP did not contribute (see supplementary materials). The anisotropic polarization angle dependence of the photocurrent without sign change was also reproduced by these calculations (Fig. 4D). The concept of “symmetry engineering” and the polarity in a nonperiodic system have been established in this work.

REFERENCES AND NOTES

1. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
2. C.-H. Lee et al., *Nat. Nanotechnol.* **9**, 676–681 (2014).
3. L. Britnell et al., *Science* **335**, 947–950 (2012).
4. T. Song et al., *Science* **360**, 1214–1218 (2018).
5. M. Yankowitz et al., *Nat. Phys.* **8**, 382–386 (2012).
6. C. R. Dean et al., *Nature* **497**, 598–602 (2013).
7. Y. Cao et al., *Nature* **556**, 43–50 (2018).
8. Y. Liu et al., *Nat. Nanotechnol.* **13**, 828–834 (2018).
9. B. I. Sturman, V. M. Fridkin, *The Photovoltaic and Photoelectronic Effects in Noncentrosymmetric Materials* (Gordon and Breach Science Publishers, 1992).
10. Q. Ma et al., *Nat. Phys.* **13**, 842–847 (2017).
11. S.-Y. Xu et al., *Nat. Phys.* **14**, 900–906 (2018).
12. G. B. Osterhoudt et al., *Nat. Mater.* **18**, 471–475 (2019).
13. J. Ma et al., *Nat. Mater.* **18**, 476–481 (2019).
14. T. Morimoto, N. Nagaosa, *Sci. Adv.* **2**, e1501524 (2016).
15. B. M. Fregoso, T. Morimoto, J. E. Moore, *Phys. Rev. B* **96**, 075421 (2017).
16. I. Grinberg et al., *Nature* **503**, 509–512 (2013).
17. P. S. Brody, *J. Solid State Chem.* **12**, 193–200 (1975).
18. S. Y. Yang et al., *Nat. Nanotechnol.* **5**, 143–147 (2010).
19. R. Nechache et al., *Appl. Phys. Lett.* **98**, 202902 (2011).
20. M. Nakamura et al., *Phys. Rev. Lett.* **116**, 156801 (2016).
21. M. Nakamura et al., *Nat. Commun.* **8**, 281 (2017).
22. Z. Xiao et al., *Nat. Mater.* **14**, 193–198 (2015).
23. Z. Sun et al., *Angew. Chem. Int. Ed.* **55**, 6545–6550 (2016).

24. N. Ogawa, M. Sotome, Y. Kaneko, M. Ogino, Y. Tokura, *Phys. Rev. B* **96**, 241203 (2017).
25. H. Yuan *et al.*, *Nat. Nanotechnol.* **9**, 851–857 (2014).
26. Y. J. Zhang *et al.*, *Nature* **570**, 349–353 (2019).
27. A. M. Cook, B. M. Fregoso, F. de Juan, S. Coh, J. E. Moore, *Nat. Commun.* **8**, 14176 (2017).
28. T. Rangel *et al.*, *Phys. Rev. Lett.* **119**, 067402 (2017).

ACKNOWLEDGMENTS

Funding: This work was supported by JSPS Grant-in-Aid for Scientific Research (S) (JP19H05602), the A3 Foresight Program, Grant-in-Aid for Challenging Research (Exploratory) (no. JP19K21843), Grant-in-Aid for Scientific Research on Innovative Areas (JP20H05264), Grant-in-Aid for Scientific Research (B) (JP19H01819), JST PRESTO (JPMJPR19L1 and JPMJPR19L9), JST CREST (JPMJCR19T3) and Grant-in-Aid for JSPS Fellows (JP17J08941 and JP17J09152). M.O. was supported by the Advanced Leading Graduate Course for Photon Science (ALPS). Y.N. was supported by the Materials Education program for the

future leaders in Research, Industry, and Technology (MERIT) J.L. was supported by the Nakatani RIES program. H.Y. acknowledges support from the National Natural Science Foundation of China (91750101, 51861145201, 52072168, and 21733001), the National Key Basic Research Program of the Ministry of Science and Technology of China (2018YFA0306200), the Fundamental Research Funds for the Central Universities (021314380078, 021314380104, and 021314380147), the Priority Academic Program Development of Jiangsu Higher Education Institutions (021314416201), and Jiangsu Key Laboratory of Artificial Functional Materials. D.Y. and Z.Y. acknowledge support from the Natural Sciences and Engineering Research Council of Canada, Canada Foundation for Innovation, New Frontiers in Research Fund, Canada First Research Excellence Fund, and Max Planck–UBC–UTokyo Centre for Quantum Materials. Z.Y. is also supported by the Canada Research Chairs Program. **Author contributions:** T.J. and Y.I. conceived and designed the experiments. T.A. and D.Y. made and characterized devices. Y.N. developed the transfer technique. T.A., T.J., L.Z., J.H., M.O., D.Y., Y.Z., and H.Y.

performed the photocurrent mapping and spectrum measurement. S.K., M.Y., and T.M. performed the theoretical analysis. T.A., T.J., H.Y., and Y.I. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/68/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (29–39)

19 October 2019; resubmitted 14 December 2020
Accepted 18 February 2021
10.1126/science.aaz9146

TOPOLOGICAL OPTICS

Nonlinear tuning of PT symmetry and non-Hermitian topological states

Shiqi Xia^{1*}, Dimitrios Kaltsas^{2*}, Daohong Song^{1*}, Ioannis Komis², Jingjun Xu¹, Alexander Szameit³, Hrvoje Buljan^{1,4†}, Konstantinos G. Makris^{2,5†}, Zhigang Chen^{1,6†}

Topology, parity-time (PT) symmetry, and nonlinearity are at the origin of many fundamental phenomena in complex systems across the natural sciences, but their mutual interplay remains unexplored. We established a nonlinear non-Hermitian topological platform for active tuning of PT symmetry and topological states. We found that the loss in a topological defect potential in a non-Hermitian photonic lattice can be tuned solely by nonlinearity, enabling the transition between PT-symmetric and non-PT-symmetric regimes and the maneuvering of topological zero modes. The interaction between two apparently antagonistic effects is revealed: the sensitivity close to exceptional points and the robustness of non-Hermitian topological states. Our scheme using single-channel control of global PT symmetry and topology via local nonlinearity may provide opportunities for unconventional light manipulation and device applications.

Although there are numerous distinct phenomena mediated by topology, symmetry, and nonlinearity, a complex system simultaneously exhibiting all these features is hard to find. In 2008, two important concepts—the quantum Hall edge state and parity-time (PT) symmetry—were introduced to photonics (1, 2), leading to the birth of topological photonics (3) and non-Hermitian optics (4), respectively. Topological photonics, the use of topological ideas to control the behavior of light, has been realized in a variety of photonic settings (3, 5–7). PT

symmetry in optics, by contrast, is implemented using a complex refractive index and has also provided a plethora of alternative designs for controlling light, aiming toward new types of photonic devices based on non-Hermitian physics (4, 8–10). Combining the two areas is conceptually challenging, but experiments have shown that topological edge states can indeed be observed in non-Hermitian systems (11, 12). Moreover, non-Hermitian characteristics give rise to intriguing topological phenomena such as topological light steering and funneling (13, 14). One striking development relevant to technological applications is the demonstration of topological insulator lasers (15, 16), in which topology and non-Hermiticity naturally coalesce and conspire: Lasing is based on topologically protected modes, and a laser system is inherently non-Hermitian because gain and loss are present. However, so far, non-Hermitian topological photonics have mainly been restricted to the linear-optics regime, and only recently has it become clear that many intriguing phenomena arise when nonlinearity is taken into account

in nonlinear non-Hermitian (NNH) topological systems (17–21).

We demonstrate a scheme for single-channel nonlinear control of a complex system with underlying global dynamics driven by the interplay among topology, non-Hermiticity, and nonlinearity. Our experimental platform is based on photonic Su-Schrieffer-Heeger (SSH) (22, 23) lattices consisting of laser-written continuous (“gain”) and sectioned (“loss”) waveguides with an interface defect (Fig. 1A), but it applies equally to a broad spectrum of NNH systems that have intensity-dependent gain or loss. The SSH lattices represent a prototypical one-dimensional topological system with chiral symmetry (3), as is widely used for the study of topologically protected quantum states (24), nonlinearity-driven topological effects (21, 25–27), and topological lasing (28), among other things.

It is known that an active linear non-Hermitian PT-symmetric system can be directly mapped onto a system with only loss simply by introducing a global decay factor (equivalent to an offset for the imaginary part described by the gain-loss profile) (4). In such “passive” PT-symmetric systems, non-Hermitian PT phenomena have been demonstrated without using actual material gain (8, 12). Unlike previous implementations of loss, we use a weak continuous-wave (cw) laser to write nonlinear non-Hermitian SSH lattices (NNH-SSHs) in a bulk nonlinear crystal (27), as illustrated in Fig. 1A. The continuous waveguides (red) represent the “gain” ones, whereas the sectioned waveguides can be “lossy” (blue) or “neutral” (green) depending on the gap size introduced between sections. An NNH-SSH that is realized in the passive PT-symmetric regime may, under the action of self-focusing nonlinearity experienced by a probe beam at the interface, change into a non-PT “gain” system, as self-focusing reduces diffraction loss and leakage in the center waveguide. Likewise, under the action of self-defocusing nonlinearity, it may turn into a non-PT “loss” system, because in

¹MOE Key Laboratory of Weak-Light Nonlinear Photonics, TEDA Applied Physics Institute and School of Physics, Nankai University, Tianjin 300457, China. ²Department of Physics, University of Crete, Heraklion 71003, Greece.

³Institut für Physik, Universität Rostock, 18059 Rostock, Germany. ⁴Department of Physics, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia. ⁵Institute of Electronic Structure and Laser (IESL)—FORTH, Heraklion 71110, Greece. ⁶Department of Physics and Astronomy, San Francisco State University, San Francisco, CA 94132, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: hbuljan@phy.hr (H.B.); makris@physics.uoc.gr (K.G.M.); zgchen@nankai.edu.cn (Z.C.)

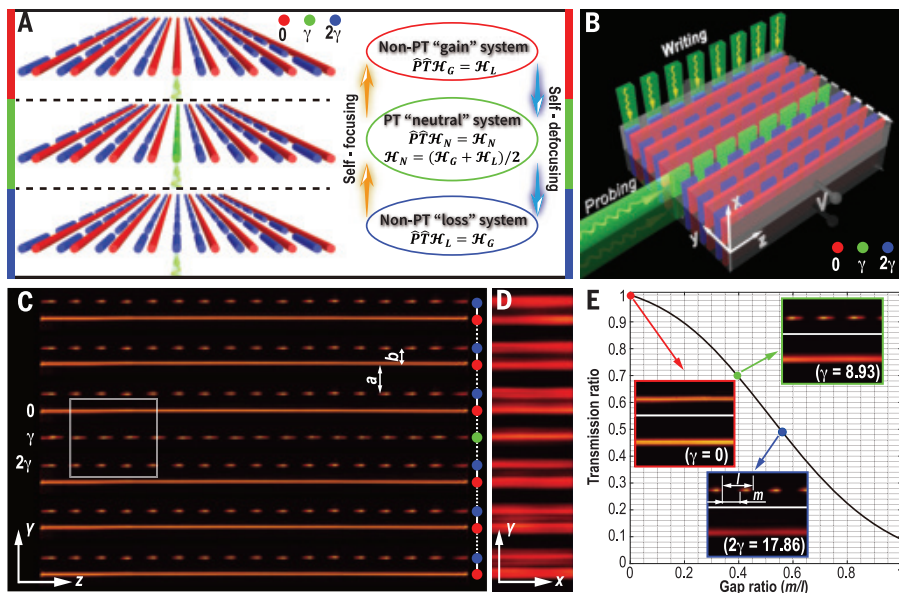


Fig. 1. Experimental realization of NNH-SSHs for nonlinear tuning of PT symmetry and topological states. (A) Illustration of a passive PT-symmetric “neutral” SSH lattice (middle) switched to a non-PT “gain” lattice (top) or a non-PT “loss” lattice (bottom) by local nonlinearity at the topological defect. The switching direction can be readily reversed. Red, green, and blue cylinders and dots represent “gain,” “neutral,” and “loss” lattice sites (γ is the loss coefficient); colored bars denote different stages of the NNH-SSHs. The Hamiltonian relations are illustrated for corresponding active non-Hermitian systems (29). (B) Schematic for cw-laser writing and probing the lattices. A dashed white arrow marks the writing sequence. (C) Side view of the writing beam pattern, where $a = 22.8 \mu\text{m}$ and $b = 15.2 \mu\text{m}$ represent waveguide spacing for the dimer lattice. (D) The written NNH-SSH examined by a broad plane-wave beam. (E) Plot of intensity transmission ratio as a function of the gap ratio in a single waveguide obtained from simulation. Insets show side views of a waveguide portion taken from the experiment in (C) at $m/l = 0$ (red), 0.40 (green), and 0.56 (blue), where in each inset a single writing beam is shown at the top and the guided output probe beam at the bottom.

this case the nonlinearity enhances the leakage and entails more loss in the waveguide. In this way, nonlinear excitation of a single channel (i.e., the interface waveguide) can locally affect the properties of the whole lattice, leading to transition from PT-symmetric to non-PT-symmetric regimes. Because all three NNH-SSHs (“gain,” “loss,” and “neutral”) can be initially implemented in experiment, our explored platform enables a convenient approach to achieving nonlinearity-induced switching between different non-Hermitian lattices. Interestingly, the Hamiltonians of these NNH-SSHs are inherently related (29), and the underlying connection directly affects corresponding complex eigenvalue spectra across the exceptional point (EP)—a special kind of degeneracy with coalesced eigenvalues and eigenstates unique to non-Hermitian Hamiltonians (30).

Our experimental method to establish the NNH-SSHs is illustrated in Fig. 1B, where the writing beam is either a uniform stripe beam (for writing the “gain” waveguides) or a periodically modulated beam (for writing the sectioned “loss” and “neutral” waveguides). The waveguides are written sideways one by one in a biased photorefractive (SBN:61) crystal (27). In the entire writing process, the bias field is $E_0 = 160 \text{ kV/m}$, the writing beam has a power

of $\sim 200 \mu\text{W}$, and the probing beam has a power of $\sim 2.5 \mu\text{W}$. [See (29) for other experimental details.] A passive PT-symmetric SSH system requires precise control of loss. To achieve this, we keep the same total number of waveguide sections (with section length l) in each channel (17 sections in the 20-mm-long crystal), but we make the spacing between adjacent sections (i.e., the gap length m) smaller in the “neutral” waveguide than in the “loss” waveguides. A superimposed writing beam pattern is shown in Fig. 1C, consisting of alternating continuous and sectioned stripes relative to the center defect channel. A typical NNH-SSH written this way is shown in Fig. 1D, for which strong and weak coupling correspond to smaller and larger channel separation, respectively. Figure 1E plots the normalized intensity transmission ratio (defined as $\gamma = I_{\text{out}}/I_0$, where I_{out} and I_0 are the output intensity of the same probe beam from a sectioned and a continuous waveguide, respectively) as a function of the gap ratio (defined as $x = m/l$, which controls the waveguide loss). As the gap length increases, the loss in the waveguide increases, and thus the transmission decreases (Fig. 1E, insets). The loss coefficient γ is determined from the intensity transmission $I_{\text{out}} = I_0 \exp(-2\gamma L)$, where $L = 20 \text{ mm}$ corresponds to the crystal

length (29). This plot serves as the basis for determining the parameters for the writing beams in the experiment. For example, the “gain” waveguide ($\gamma = 0$) corresponds to the red dot at $(x, y) = (0, 1)$, because it is continuous ($m = 0$) and lossless. The “neutral” waveguide, marked by the green dot at $(x, y) = (0.40, 0.70)$, corresponds to a gap ratio of 0.40 and a transmission ratio of 0.70, which yields $\gamma = 8.93 \text{ m}^{-1}$. From this, we can in turn find the parameters for the “loss” waveguides, marked by the blue dot at $(x, y) = (0.56, 0.49)$, and obtain the desired loss 2γ . The NNH-SSH established with such judiciously introduced losses fulfills the requirement for PT symmetry.

Before presenting the experimental results, we examine theoretically the topological states in a corresponding active PT-symmetric SSH lattice with an interface defect, as illustrated in Fig. 2A. Under the tight-binding approximation, the linear coupled mode equations are

$$-i \frac{\partial}{\partial z} \varphi_n = \beta^* \varphi_n + c_1 \varphi_{n-1} + c_2 \varphi_{n+1}, \quad n = 2, 4, \dots \text{ or } -1, -3, \dots \quad (1a)$$

$$-i \frac{\partial}{\partial z} \varphi_n = \beta \varphi_n + c_2 \varphi_{n-1} + c_1 \varphi_{n+1}, \quad n = 1, 3, \dots \text{ or } -2, -4, \dots \quad (1b)$$

$$-i \frac{\partial}{\partial z} \varphi_0 = \beta_0 \varphi_0 + c_2 \varphi_1 + c_2 \varphi_{-1}, \quad n = 0 \quad (1c)$$

(1I, 1J), where φ_n denotes the modal amplitude in the n th waveguide, $\beta = \alpha + i\gamma$ (α and γ are the real and imaginary parts of the waveguide potential), c_1 and c_2 are the strong and weak coupling coefficients, and β_0 denotes the potential of the center defect waveguide at $n = 0$. If $\gamma = 0$ for all waveguides, the non-Hermitian SSH collapses to the Hermitian model that supports topologically protected mid-gap (zero-mode) states (25). Even when the loss or gain is introduced ($\gamma \neq 0$), the non-Hermitian SSH lattice described above still supports a PT-symmetric topological state, provided that there is no gain or loss at the dimerization defect (12)—that is, $\beta_0 = \alpha$, $\gamma_0 = 0$. In our model, we assume that the lattice is terminated at the weak-coupling bond (c_2) such that no edge states are present on either side (23). The results are summarized in Fig. 2B, showing how a topological interface state is affected by non-Hermiticity and nonlinearity. The above equations can be expressed in a matrix form, and the relations between the Hamiltonians \mathcal{H}_G , \mathcal{H}_L , and \mathcal{H}_N (corresponding to the three lattices) are given in Fig. 1A (29).

Our lattices consist of 33 waveguides with $c_1 = 4$, $c_2 = 1$. The linear propagation constant for all waveguides is set as $\alpha = 0$, $\gamma = 1$, except for $n = 0$. In the linear regime, α is the same for all waveguides, and a typical PT-symmetric mid-gap interface state is represented by point A in Fig. 2B. As seen from the top left panel

of Fig. 2B, all eigenmodes have only real eigenvalues, because the lattice is in the PT symmetry-unbroken regime (12). In the nonlinear regime, the propagation constant is

intensity-dependent: $\beta(I) = \alpha(I) + i\gamma(I)$, where I is the intensity of the excitation beam. As such, the eigenvalue of the topological state can be moved away from its initial mid-gap

position by the action of the nonlinearity (25, 27). Because a probe beam excites only the center defect channel while it experiences an overall loss in a passive NNH-SSH, it is reasonable to model the system with nonlinearity present only in the defect channel: $\beta_0(I) = \alpha_0(I) + i\gamma_0(I)$. If the nonlinearity only changes the real part of the potential while keeping $\gamma_0 = 0$, the eigenvalue of the zero mode is shifted away from the center of the gap, moving upward (or downward) as a consequence of the self-focusing (or -defocusing) effect to point B (or C) in Fig. 2B. The eigenmode profiles remain as symmetric as that of the mid-gap mode because the lattice overall still preserves the PT symmetry. In contrast, if the nonlinearity also changes the imaginary part of the potential γ_0 , PT symmetry is destroyed. This scenario corresponds to results marked by points D and E, where the imaginary part of the eigenvalues is shifted away from the zero-mode position. In this non-PT regime, the eigenmode profiles become asymmetric, as more energy of the modes flows to the “loss” (D) or “gain” (E) waveguides depending on the sign of the nonlinearity. Thus, the observation of nonlinearity-induced asymmetric mode profiles in the NNH-SSHs serves as a signature for the change of the imaginary part of the defect potential, indicating whether PT symmetry is present or not.

With the NNH-SSHs implemented by cw-laser writing (Fig. 1), we experimentally demonstrated this nonlinear tuning of the zero mode by launching a probe beam into the defect channel (Fig. 3, bottom row). Starting from

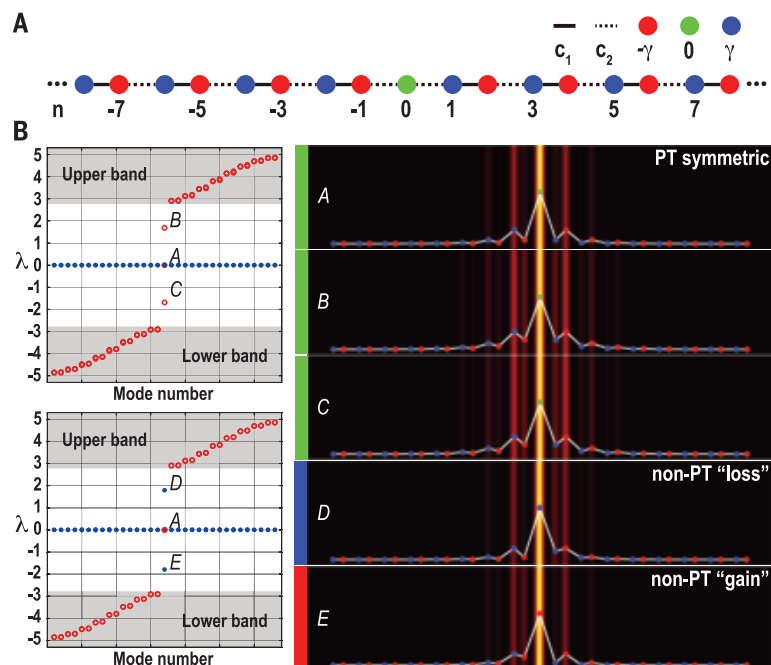


Fig. 2. Calculated non-Hermitian topological interface states tuned by nonlinearity. (A) Illustration of an active PT-symmetric SSH with an interface topological defect located at site $n = 0$. Colored dots represent different lattice sites. (B) Left: Calculated eigenvalues λ for a finite lattice with 33 sites. Red circles and blue dots denote real and imaginary parts of the eigenvalues, respectively; shaded regions illustrate the band structure of an infinite lattice. Right: The corresponding eigenmode profiles, where the eigenvalues for points A to E are obtained with propagation constants $\beta_0 = 0, 2, -2, 2i$, and $-2i$ while keeping β for all other waveguides unchanged. Color codes for different waveguides are the same as in Fig. 1.

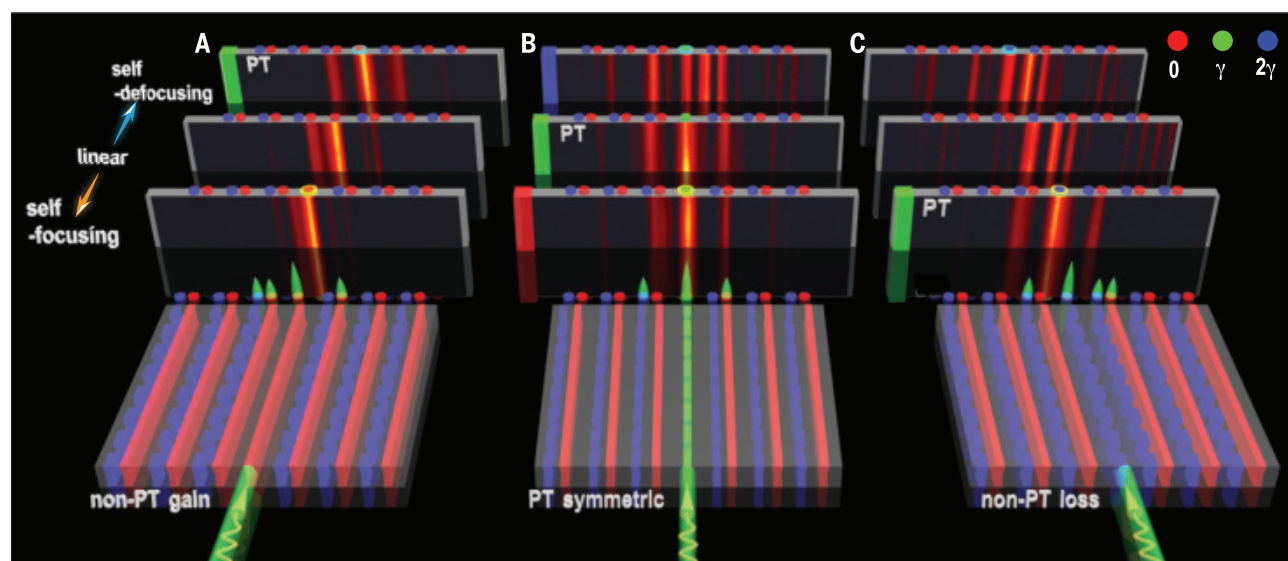


Fig. 3. Demonstration of single-channel nonlinear control of PT symmetry and topological states. (A to C) Top three rows: Experimental results showing output transverse patterns of a probe beam launched into the defect channel for three distinct cases. The NNH-SSH is fabricated with a “gain” (A), “neutral” (B), or “loss” (C) interface waveguide. Bottom row: Schematic of the excitation beam and the corresponding linear mode profile. In the linear regime, only the “neutral” lattice (B) initially has PT symmetry. In (A) and (C), the lattice is non-PT-symmetric, but PT symmetry and topological states are restored under the action of nonlinearity.

linear propagation (i.e., without the bias field), a symmetric topological interface state (corresponding to point A in Fig. 2B) is observed in the “neutral” lattice (Fig. 3B, second row), which indicates that in this case the NNH-SSH respects the PT symmetry (12). However, under the action of a self-defocusing nonlinearity, the probe beam induces anti-guiding so that more of its energy escapes from the defect channel, equivalently introducing leakage (“loss”) to the center waveguide. The NNH-SSH turns into a non-PT-symmetric structure, and the excited mode becomes asymmetric (Fig. 3B, first row) as more light goes to the “loss” waveguide, corresponding to point D in Fig. 2B. In contrast, when a self-focusing nonlinearity is used, it induces self-guiding, such that the diffraction is suppressed, equivalently providing “gain” to the center waveguide. Again, the beam becomes asymmetric but now more light goes to the “gain” waveguide, corresponding to point E in Fig. 2B. Because evidently the change in the real part of the index potential alone does not result in asymmetric modes, our results represent a nonlinearity-induced transition from a PT-symmetric lattice to a non-Hermitian lattice without PT symmetry.

As illustrated in Fig. 1A, the transition can also be reversed by nonlinearity. Such results are shown in Fig. 3A with an initial “gain” NNH-SSH, and in Fig. 3C with an initial “loss” NNH-SSH. In the non-PT “gain” lattice, a probe beam evolves linearly into an asymmetric distribution. Under self-defocusing nonlinearity, a symmetric profile with the characteristic feature of a topological interface state is restored (Fig. 3A, first row), as nonlinearity entails the retrieval of lattice PT symmetry. However, in this case, it cannot be restored with self-focusing nonlinearity, which increases the gain-loss imbalance (Fig. 3A, third row). The scenario corresponding to an inversed transition starting from a non-PT “loss” to a PT-symmetric lattice is shown in Fig. 3C. These results, corroborated by numerical simulations (29), clearly demonstrate nonlinearity-mediated control of PT symmetry and topological states in the NNH-SSHs.

So far, we have shown that local nonlinearity can be used to control the loss in the defect waveguide, thereby affecting the global lattice properties. We now discuss tuning of the NNH-SSHs close to the EPs, where intriguing properties are expected for non-Hermitian systems (30–32). To this end, in our model we considered a scan through the EPs by changing the global gain or loss amplitude; our findings are summarized in Fig. 4, A and B. For the three lattices defined earlier, we kept the potential of the central defect waveguide fixed (as set by the nonlinearity in our experiment) and varied the gain/loss ratio for all other waveguides. The complex eigenvalue spectra λ_n exhibit a markedly different bifurcation

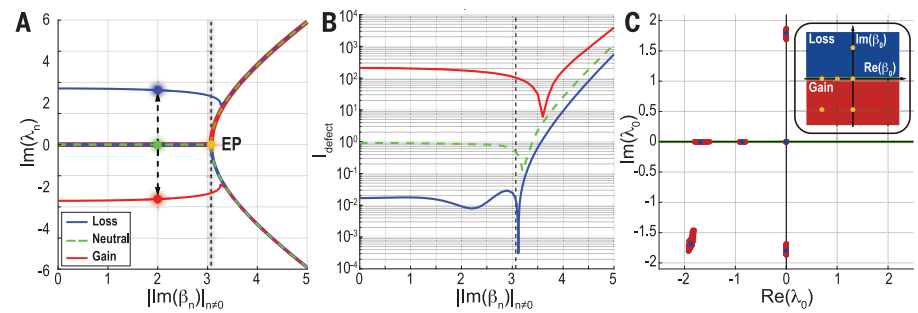


Fig. 4. Theoretical analysis of complex spectra around the EP and zero-mode robustness. (A) The range of the imaginary part of the complex eigenvalues (maximal/minimal magnitudes) as a function of the global gain-loss amplitude for three NNH-SSHs. The EP is marked for the “neutral” PT-symmetric lattice. Colored dots before the EP correspond to experimental gain/loss parameters ($\beta_0 = 2i, 0, -2i$) at the defect waveguide; they can be switched via nonlinearity, as indicated by the dashed arrow. Other lattice parameters are the same as for Fig. 2. (B) Plot of defect-channel intensity in the NNH-SSHs tuned across the EP under the same initial conditions. (C) Sensitivity of defect-mode eigenvalues to the perturbation on coupling coefficients when $|\text{Im}(\beta_n)|_{n \neq 0}$ is fixed to 1. Red dots denote the eigenvalue fluctuation for 100 realizations of added perturbation; the blue dots correspond to the initial defect eigenvalue. The inset depicts corresponding values of the central defect potential β_0 on the complex plane. Notice the robustness of the exact zero mode at the origin.

feature: Before reaching the EP of the PT-symmetric lattice, which has only real eigenvalues, they come in complex conjugate pairs (with different ranges of imaginary magnitudes) for the other two lattices; after a certain value of gain/loss ratio beyond the pertinent EP, all three lattices exhibit the same range of imaginary eigenvalues determined by the bulk modes of the lattices (Fig. 4A). This is a direct outcome of the inherent connection between the Hamiltonians (29). At the EP, the bandgap closes and the topologically protected mode becomes extended, which suggests that a phase transition should have taken place. This can be clearly seen from the sharp drop of the transmitted intensity through the defect channel after a long propagation distance in all three lattices (Fig. 4B): The cusp followed by a continuous rise of the intensity indicates the presence of an EP in each lattice. Moreover, with the nonlinear tuning illustrated in Fig. 4A, our scheme can give rise to the birth or death of EPs for the NNH-SSHs (29), which merits further investigation.

A direct outcome of the topological nature of the SSH model is the robustness of zero modes to perturbation in the off-diagonal of the Hamiltonian. One may wonder which of the two opposite tendencies will prevail: the sensitivity or the robustness, especially when close to the EP (30–32). We theoretically address this question in Fig. 4C, where the defect mode eigenvalues are plotted on the complex plane for various values of defect potential but the gain/loss ratio is fixed for the rest of the lattice. We examine the robustness of the defect mode to off-diagonal perturbations (15% on the coupling coefficients) that respect the lattice chiral symmetry. Strictly

speaking, only the PT-symmetric lattice supports the exact zero mode with complete topological robustness. Once the eigenvalue of the defect mode is driven away from the origin in the complex plane, topological protection is gradually lost. Interestingly, this loss of protection is not “isotropic” (in a sense that the instability of the defect mode grows in a preferred direction in the complex spectra), but is enhanced when the parameters are tuned close to the EP.

Topology and PT symmetry typically describe the global properties of a system, whereas most optical nonlinearities are local. Therefore, their interplay in some sense is a manifestation of the interplay of local and global characteristics. It is natural to ask: Is there a general theory for the NNH PT-symmetric systems driven by nonlinearity (29)? Beyond that, there are many fundamental questions yet to answer. For instance, how can one characterize topological invariants for finite NNH-SSHs driven by nonlinearity, given that the concept developed for linear Hermitian systems is not equally applicable? More intriguingly, can nonlinearity induce crossing or encircling of the EP (31, 32) in order to control the dynamics of topological modes? Is it possible to tune the bandgap structures and remotely control the power flow at a target destination far from a topological defect in NNH systems? How would local nonlinearity affect overall topological features and classifications of symmetry and topology (33)? We envisage that nonlinear control may play a unique role in integrating these concepts to harness complex systems beyond photonics, including but not limited to acoustics, plasmonics, polaritonics, and ultracold atoms.

REFERENCES AND NOTES

1. F. D. Haldane, S. Raghu, *Phys. Rev. Lett.* **100**, 013904 (2008).
2. K. G. Makris, R. El-Ganainy, D. N. Christodoulides, Z. H. Musslimani, *Phys. Rev. Lett.* **100**, 103904 (2008).
3. T. Ozawa *et al.*, *Rev. Mod. Phys.* **91**, 015006 (2019).
4. R. El-Ganainy *et al.*, *Nat. Phys.* **14**, 11–19 (2018).
5. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljacic, *Nature* **461**, 772–775 (2009).
6. M. C. Rechtsman *et al.*, *Nature* **496**, 196–200 (2013).
7. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, *Nat. Photonics* **7**, 1001–1005 (2013).
8. A. Guo *et al.*, *Phys. Rev. Lett.* **103**, 093902 (2009).
9. C. E. Rüter *et al.*, *Nat. Phys.* **6**, 192–195 (2010).
10. A. Regensburger *et al.*, *Nature* **488**, 167–171 (2012).
11. C. Poli, M. Bellec, U. Kuhl, F. Mortessagne, H. Schomerus, *Nat. Commun.* **6**, 6710 (2015).
12. S. Weimann *et al.*, *Nat. Mater.* **16**, 433–438 (2017).
13. H. Zhao *et al.*, *Science* **365**, 1163–1166 (2019).
14. S. Weidemann *et al.*, *Science* **368**, 311–314 (2020).
15. M. A. Bandres *et al.*, *Science* **359**, eaar4005 (2018).
16. G. Harari *et al.*, *Science* **359**, eaar4003 (2018).
17. M. Wimmer *et al.*, *Nat. Commun.* **6**, 7782 (2015).
18. Y. Lumer, Y. Plotnik, M. C. Rechtsman, M. Segev, *Phys. Rev. Lett.* **111**, 243905 (2013).
19. S. Mukherjee, M. C. Rechtsman, *Science* **368**, 856–859 (2020).
20. L. J. Maczewsky *et al.*, *Science* **370**, 701–704 (2020).
21. D. Smirnova, D. Leykam, Y. Chong, Y. Kivshar, *Appl. Phys. Rev.* **7**, 021306 (2020).
22. W. P. Su, J. R. Schrieffer, A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
23. N. Malkova, I. Hromada, X. Wang, G. Bryant, Z. Chen, *Opt. Lett.* **34**, 1633–1635 (2009).
24. A. Blanco-Redondo, B. Bell, D. Oren, B. J. Eggleston, M. Segev, *Science* **362**, 568–571 (2018).
25. N. Malkova, I. Hromada, X. Wang, G. Bryant, Z. Chen, *Phys. Rev. A* **80**, 043806 (2009).
26. Y. Hadad, A. B. Khanikaev, A. Alù, *Phys. Rev. B* **93**, 155112 (2016).
27. S. Xia *et al.*, *Light Sci. Appl.* **9**, 147 (2020).
28. P. St-Jean *et al.*, *Nat. Photonics* **11**, 651–656 (2017).
29. See supplementary materials.
30. Ş. K. Özdemir, S. Rotter, F. Nori, L. Yang, *Nat. Mater.* **18**, 783–798 (2019).
31. J. Doppler *et al.*, *Nature* **537**, 76–79 (2016).
32. H. Hodaei *et al.*, *Nature* **548**, 187–191 (2017).
33. K. Kawabata, K. Shiozaki, M. Ueda, M. Sato, *Phys. Rev. X* **9**, 041015 (2019).

ACKNOWLEDGMENTS

Funding: Supported by the National Key R&D Program of China under grant 2017YFA0303800; National Natural Science Foundation grants 11922408, 91750204, and 11674180; PCSIRT; the 111 Project (no. B07013); the Sino-German Mobility Programme (M-0198); Deutsche Forschungsgemeinschaft grants SZ 276/9-2, SZ 276/19-1, SZ 276/20-1, and BL 574/13-1 (A.S.); and Croatian Science Foundation grant IP-2016-06-5885 SynthMagIA and the QuantiX Lie Center of Excellence (a project co-financed by the Croatian Government and the European Union through the European Regional Development Fund Competitiveness and Cohesion Operational Programme; grant KK.01.1.1.01.0004) (H.B.). **Author contributions:** S.X. and D.S. performed the experiments and numerical simulations. D.K. and I.K. assisted in theoretical analysis. Z.C., H.B., and K.G.M. supervised the project. All authors discussed the results and contributed to this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/72/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
References (34–38)

12 November 2020; accepted 18 February 2021
10.1126/science.abf6873

CATALYSIS

Isolated boron in zeolite for oxidative dehydrogenation of propane

Hang Zhou^{1*}, Xianfeng Yi^{2*}, Yu Hui^{3*}, Liang Wang^{1,4†}, Wei Chen², Yucai Qin³, Ming Wang⁵, Jiabi Ma⁵, Xuefeng Chu⁶, Yeqing Wang⁷, Xin Hong⁷, Zifeng Chen⁷, Xiangju Meng^{7†}, Hai Wang¹, Qiuyan Zhu⁷, Lijuan Song³, Anmin Zheng^{2†}, Feng-Shou Xiao^{1,4,7†}

Oxidative dehydrogenation of propane (ODHP) is a key technology for producing propene from shale gas, but conventional metal oxide catalysts are prone to overoxidation to form valueless CO_x. Boron-based catalysts were recently found to be selective for this reaction, and B–O–B oligomers are generally regarded as active centers. We show here that the isolated boron in a zeolite framework without such oligomers exhibits high activity and selectivity for ODHP, which also hinders full hydrolysis for boron leaching in a humid atmosphere because of the B–O–SiO_x linkage, achieving superior durability in a long-period test. Furthermore, we demonstrate an isolated boron with a –B[OH...O(H)–Si]₂ structure in borosilicate zeolite as the active center, which enables the activation of oxygen and a carbon–hydrogen bond to catalyze the ODHP.

Propene (1–4) has been extensively used as a building unit for the production of plastics, fibers, and intermediates for the synthesis of fine chemicals. The current method for propene production predominantly relies on the steam cracking of heavy hydrocarbons in naphtha, but production using this method is still far from meeting the demands for propene, ~116,000,000 tons in 2020. Recently, shale gas, which contains abundant propane, has been regarded as a promising source for propene to reduce dependence on petroleum. Nonoxidative dehydrogenation of propane has emerged as industrial technologies such as Oleflex by Honeywell UOP and Catofin by ABB Lums (1, 2, 5–9) have been developed, but these have insufficient lifetimes because of thermodynamically favorable coke formation and metal sintering, which require frequent catalyst regeneration. By contrast, oxidative dehydrogenation of propane (ODHP) has substantial advantages regarding exothermic features and hindrance of coke formation but is easily overoxidized to form valueless CO₂ (10–18), and this remains a challenge for practical applications. One solution to this issue is the recent discovery of boron-containing catalysts (e.g., BN, supported boron oxide) that can catalyze dehydrogenation and suppress overoxidation (19–25). For these catalysts, the B–O–B oligomers are generally regarded as active centers (26–28), whereas a single-site boron in zeolite (e.g., B–MWW) is inert for the reaction (29). However, the B–O–B oligomers are associated with complete hydrolysis by water products to form boron with saturated hydroxyl groups (e.g., boric acid), which could deactivate the catalyst once the boric acid leaches from the reaction system because of the strong solubility of such species in water (24, 30).

¹Key Lab of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China. ²National Center for Magnetic Resonance in Wuhan, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics and Mathematics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430071, China. ³Key Laboratory of Petrochemical Catalytic Science and Technology, Liaoning Shihua University, Fushun 113001, China. ⁴Ningbo Research Institute, Zhejiang University, Ningbo 315100, China. ⁵Key Laboratory of Cluster Science of Ministry of Education, Beijing Key Laboratory of Photoelectronic/Electrophotonic Conversion Materials, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, China. ⁶Key Laboratory of Architectural Cold Climate Energy Management, Jilin Jianzhu University, Changchun 130118, China. ⁷Department of Chemistry, Zhejiang University, Hangzhou 310028, China.

*These authors contributed equally to this work.

†Correspondence to: liangwang@zju.edu.cn (L.W.); zhenganm@wipm.ac.cn (A.Z.); mengxj@zju.edu.cn (X.M.); fsxiao@zju.edu.cn (F.-S.X.)

Herein, we report a new active center with a –B[OH...O(H)–Si]₂ structure in the borosilicate MFI framework (BS-1) for this reaction that contains only isolated boron species, which differs from general catalysts containing B–O–B oligomers. Such a structure not only efficiently activates molecular oxygen and propane to promote dehydrogenation but also hinders full boron hydrolysis during catalysis. As a result, the BS-1 catalyst is water tolerant, maintaining high activity and selectivity in a continuous test for a long lifetime, and this represents a large step forward in ODHP technology.

The BS-1 catalyst was synthesized from a solvent-free crystallization strategy in the presence of tetrapropylammonium (TPA⁺) hydroxide from a synthesis gel at 180°C for 3 days. By calcination in air to remove the organic template, BS-1 with an Si/B ratio of ~62 was obtained. BS-1 catalyst was identified by x-ray diffraction (XRD) pattern and N₂ sorption

isotherms, exhibiting typical characteristics in its MFI structure (figs. S1 and S2). A scanning electron microscope (SEM) image of the BS-1 catalyst showed an average zeolite crystal size of 350 nm (fig. S3). A high-resolution transmission electron microscope (TEM) image of the BS-1 catalyst confirmed the open zeolite micropores (fig. S4). Energy dispersive spectroscopic elemental maps of the BS-1 catalyst revealed uniform distribution of boron species on the silica matrix (fig. S5). X-ray photoelectron spectroscopy and inductively coupled plasma (ICP) analysis showed undetectable metal species (fig. S6), which excludes the suspected contribution of impurities on catalytic activity. For comparison, we also loaded the boron species on S-1 zeolite (B/S-1) and amorphous silica (B/SiO₂) by impregnation (fig. S7), with Si/B ratios of 50 and 66, respectively.

The catalytic ODHP was performed in a fixed-bed reactor containing a feed gas of propane and oxygen (C₃H₈/O₂/He = 1/1/8) with a weight-hour-space-velocity (WHSV, L_{C₃H₈} kg_{catalyst}⁻¹ h⁻¹) of 3600 L kg⁻¹ h⁻¹. Figure 1A shows the dependence of propane conversion on reaction temperature over various catalysts. The automatic reaction occurring between propane and oxygen was excluded by the blank run without catalysts. The BS-1 catalyst was active for the reaction, giving propane conversion with increasing reaction temperature. For example,

23.8 and 41.4% of propane conversions were realized on the BS-1 catalyst at 540 and 560°C, respectively. By contrast, B/S-1 and B/SiO₂ catalysts exhibited lower propane conversions at 1.9 and 8.5% at 540°C, respectively.

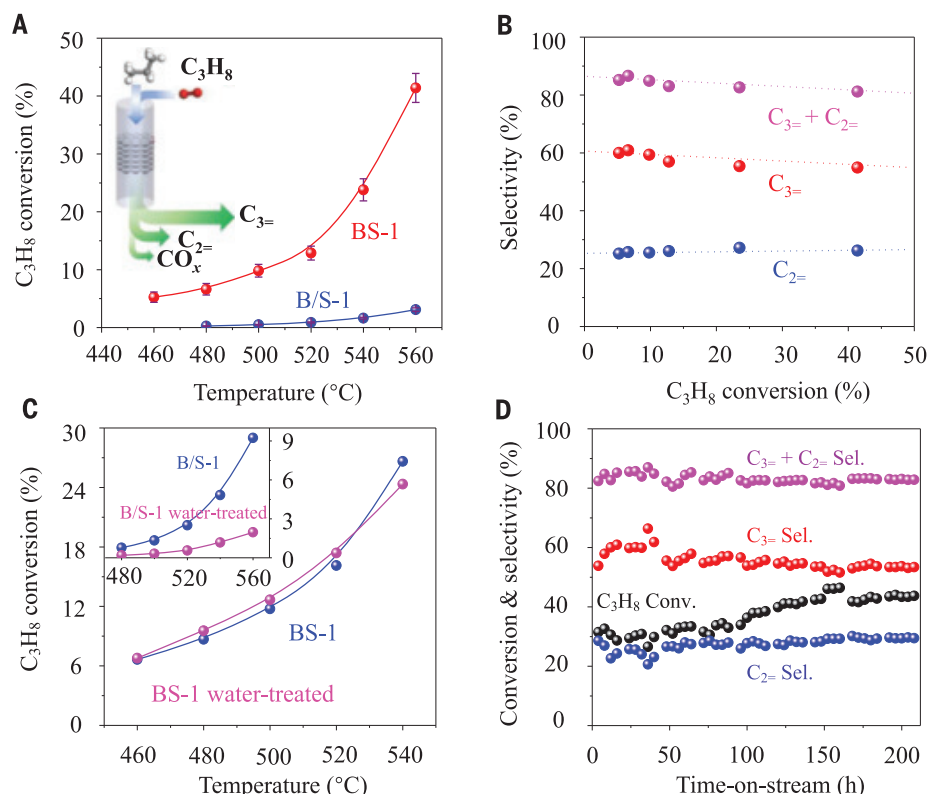
Propene appeared in the boron-catalyzed ODHP as a dominant product, and the side reaction led to the formation of ethene, which is also a valuable olefin product, rather than the valueless CO_x that is usually formed in the oxidation. Olefin selectivity as a function of the propane conversion is shown in Fig. 1B and fig. S8. S-1 zeolite without boron displayed propane combustion as a major reaction to form CO_x. The BS-1 catalyst exhibited propene and ethene selectivities (based on carbon atoms) of 60.9 and 25.7% at a low propane conversions of 5.3%, giving an overoxidized propane selectivity of <0.5% [undetectable organic oxygenates, carbon balance >96%, olefin molecular selectivity >99.5%, (moles of produced olefin/moles of transformed propane)*100%]. This excellent selectivity was mostly maintained with raising the propane conversion (Fig. 1B). For instance, the propene and ethene selectivities were 55.4 and 27.2%, respectively, under 23.8% propane conversion. When the propane conversion reached as high as 41.4%, the propene and ethene selectivities were still 54.9 and 26.3%, with selectivity for light olefins at 81.2% (Fig. 1B and fig. S9). Under these reaction conditions, the one-pass

yield of olefins reached 33.6% (22.7% for propene and 10.9% for ethene). By comparison, the B/S-1 and B/SiO₂ catalysts with similar boron content to BS-1 yielded much lower activity (fig. S8).

In ODHP, high propane conversion usually leads to easy overoxidation (18). For example, VO_x-based catalysts have selectivity for light olefins of <50%, with CO₂ as a major product when propane conversion exceeds 30% in the ODHP (fig. S8). By contrast, high olefin selectivity (>81.2%) was always obtained over the BS-1 catalyst in a wide range of propane conversions (3.8 to 41.4%). The hexagonal carbon nitride (*h*-BN) catalyst, which has been regarded as a highly efficient catalyst for ODHP, showed slightly lower olefin selectivity (79.9%) under the comparable propane conversion (40.2%; fig. S8). In particular, propene selectivity over BS-1 (54.9%) was higher than that over *h*-BN (47.0%). Because of this reaction trend, the olefin formation rate over the BS-1 catalyst was much higher than that of the *h*-BN and B/S-1 catalysts (fig. S10).

The supported boron catalysts were usually deactivated by washing with water because of boron leaching and dissolving (21, 22). For example, the water-treated B/S-1 catalyst was markedly deactivated (Fig. 1C), giving a propane conversion at 1.9%, which is much lower than that of the fresh catalyst (9.2%, 560°C; fig. S11). The BS-1 catalyst retained this ability

Fig. 1. Catalysis. (A) Dependences of propane conversion on reaction temperature over BS-1 and B/S-1 catalysts. (B) Dependences of olefin selectivity on propane conversion over BS-1. Reaction conditions were as follows: reaction temperature at 460 to 560°C for adjusting the conversions, feed gas with ratio of C₃H₈/O₂/He at 1/1/8, 0.1 MPa, WHSV = 3600 L kg⁻¹ h⁻¹. (C) Data characterizing the performances of BS-1 and B/S-1 before and after water treatment. Reaction conditions were as follows: feed gas with ratio of C₃H₈/O₂/He at 1/1/4, 0.1 MPa, WHSV = 3600 L kg⁻¹ h⁻¹. (D) Data characterizing the durability of BS-1. Reaction conditions were as follows: 570°C, feed gas with ratio of C₃H₈/O₂/He at 1/1/4, WHSV = 9000 L kg⁻¹ h⁻¹. Carbon balances were >100 ± 5% during the tests.



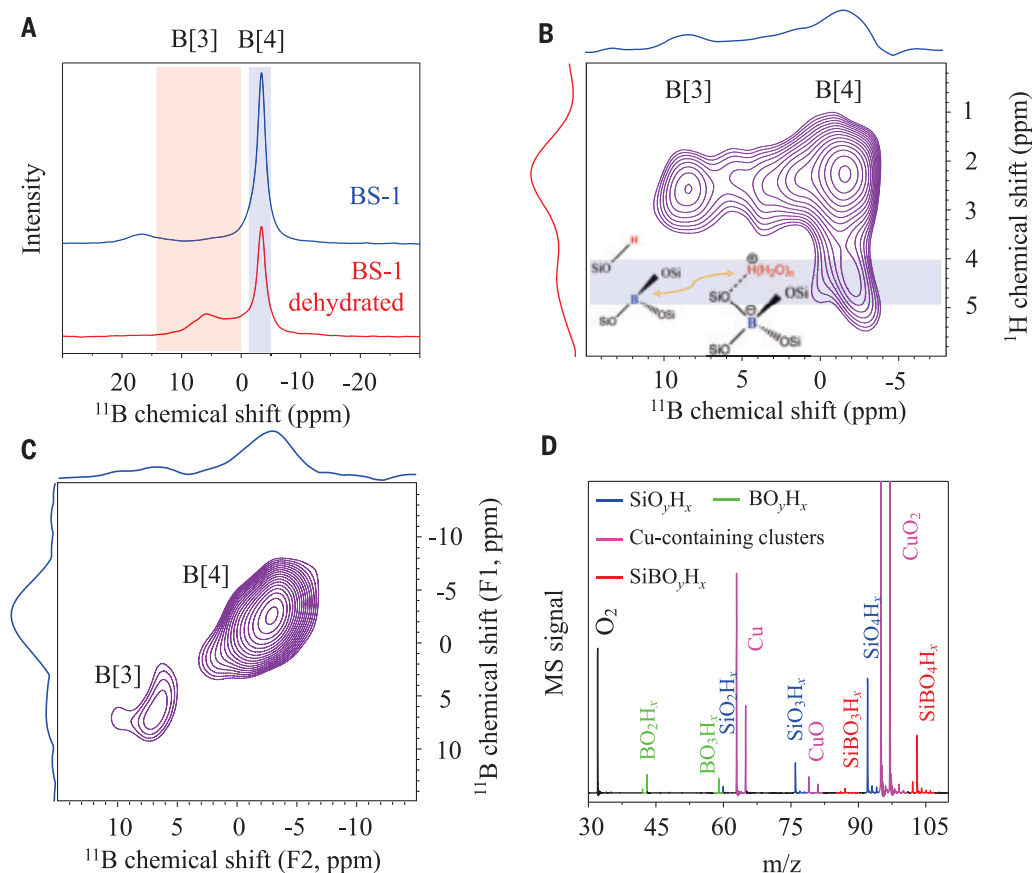


Fig. 2. Identification of isolated boron in zeolite. (A) ^{11}B NMR spectra of as-synthesized and dehydrated BS-1. (B and C) 2D ^{11}B - ^1H HETCOR (B) and 2D ^{11}B - ^{11}B PDSD MAS NMR (C) spectra of dehydrated BS-1. (D) ToF-MS spectra of BS-1. B-, Si-, and Cu-containing species were characterized by a set of peaks because of the isotopes of these elements. The isotope effect allowed differentiating a few fragment ions with similar m/z values from each other.

after the equivalent treatment (Fig. 1C). Further evaluation of the durability of the BS-1 catalyst was performed in a continuous ODHP test with high propane conversion (32.9 to 43.7%) that might produce ~10 vol% water in the gas. As shown in Fig. 1D, the propane conversion and propene and/or ethene selectivities were 32.5% and 53.8% and/or 28.6%, respectively, at the beginning of the reaction. Even after reaction for 88 hours, the BS-1 catalyst exhibited similar propane conversion (32.9%) and propene and ethene selectivities (57.0 and 28.0%, respectively). Further reaction for longer times led to increased propane conversion, reaching 42.6% at 148 hours, whereas the olefin selectivity was almost unchanged. Even after reaction for 210 hours, the catalyst still exhibited propane conversion at 43.7%, with propene and ethene selectivities at 53.3 and 29.5%, respectively, showing the good durability of the BS-1 catalyst in the ODHP. The zeolite structure was maintained and boron leaching was almost negligible in the reaction for 208 hours, as confirmed by the ICP analysis, N_2 sorption, XRD pattern, SEM images, and TEM images of the used catalyst (figs. S12 to S15). Coke formation was undetectable, as evidenced by the thermogravimetry analysis (figs. S16 and S17).

By contrast, under the equivalent reaction conditions, the *h*-BN catalyst was deactivated substantially in the tests for 12 hours (fig. S18), in agreement with the previous phenomena (24, 30). XRD and SEM results showed that the *h*-BN was partially dissolved into the boric acid and boron trioxide species (figs. S19 and S20). Although the boric acid was also active for the ODHP, it would be leached into a humid atmosphere, and its glue-like state is unsuitable for practical utilization because of corrosion and blockage to the reactor. These data all confirm the substantially improved water tolerance of the BS-1 compared with the general boron catalysts, where the zeolite framework hinders the hydrolysis and/or dissolution of boron species by water under humid ODHP conditions. This conclusion is further supported by the study on the apparent activation energies of various catalysts (figs. S21 and S22).

The unusual catalytic performance of BS-1 motivated our investigation of the structure using ^{11}B magic angle spinning (MAS) nuclear magnetic resonance (NMR) spectroscopy (Fig. 2A), where the fresh BS-1 exhibited a strong peak at ~3.9 ppm assigned to tetra-coordinated boron (B[4]) at the TO_4 site of the zeolite framework (31–34). An additional broad signal appeared at 15 to 20 ppm, which was observed on the borosilicate zeolite and as-

signed to the extra-framework boron species [e.g., $\text{B}(\text{OH})_3$] with close silanol and/or $(\text{H}_2\text{O})_n$ (35). The dehydration treatment partially removed the water, resulting in the formation of tricoordinated boron (B[3]) with ^{11}B signals at ~0 to 15 ppm (fig. S23), which is very consistent with the isolated boron in the zeolite framework (31). In addition, the extra-framework boron signal disappeared in the dehydrated BS-1 because of easy condensation with adjacent silanol to form B–O–Si linkage, in good agreement with previous observations (35).

The two-dimensional (2D) ^{11}B - ^1H heteronuclear correlation (HETCOR) spectrum of the dehydrated BS-1 provides insight into the spatial proximities between B and H species. B[4] species are always adjacent to hydroniums ($\text{H}^+[\text{H}_2\text{O}]_n$) because of the negative charge on the B[4] centers, which was confirmed here by the correlation peak of $\delta(^1\text{H})$ at ~4 to 5 ppm and $\delta(^{11}\text{B})$ at ~0 to ~4 ppm (Fig. 2B). Correlations between the ^1H NMR signal of hydronium at ~4 to 5 ppm adjacent to B[4] and the ^{11}B NMR signals of B[3] were unobservable, suggesting that the B[3] and B[4] sites should be well separated in the zeolite (Fig. 2B) (29). Further identification was performed by the 2D ^{11}B - ^{11}B proton driven spin diffusion (PDSD) homonuclear correlation

spectrum (Fig. 2C), where the off-diagonal signal assigned to spatial correlations of different boron species was undetectable. Considering that dehydration of a portion of B[4] sites ($\text{H}^+[\text{H}_2\text{O}]_n$ removal) could form B[3] sites, which is a reversible and stochastic process, the B[3]–B[4] spatial correlations would be inevitable if there were agglomerated boron species (e.g., $-\text{B}-\text{O}-\text{B}-$). Therefore, the undetectable B[3]–B[4] spatial correlations suggest isolated boron sites on BS-1. The framework boron species were further identified by Fourier transform infrared (FTIR) of BS-1 under vacuum at different temperatures, showing bands at ~ 905 to 915 and 1380 to 1400 cm^{-1} (fig. S24), which characterize the isolated boron species with tetrahedral and trigonal coordination in the zeolite framework, respectively (31, 36). Raising the temperature substantially decreased tetrahedral boron signals, confirming the transformation of B[4] to B[3] at high temperature, in good agreement with the typical features of framework boron sites (fig. S25) (36).

In addition, the structure of BS-1 was investigated using gas-phase cluster analysis from the solid structure, which is bombarded from the catalyst and randomly formed to contain compositional and structural information. Time-of-flight mass spectra (ToF-MS) showed different gas-phase clusters from the BS-1 (with metallic Cu as sample support), where the boron-containing clusters of BO_2 and BO_3 and a variety of borosilicate ions [$\text{Si}_z\text{B}_y\text{O}_x\text{H}_n$, $z = 1, 2$] were detected (Fig. 2D and figs. S26 to S28). The clusters containing the B–O–B structure (e.g., $\text{B}_2\text{O}_3\text{H}_y$) were completely undetectable (figs. S26, S28, and S29). By contrast, the ToF-MS spectra of amorphous B_2O_3 showed abundant boron oligomers (figs. S30 and S31). These data also suggest isolated boron in the BS-1, in good agreement with the NMR results.

The hydroxyl groups on the isolated boron were identified using NMR. Figure 3A shows the ^1H MAS NMR spectra of BS-1 and B/S-1, with signals at 1.8 ppm assigned to terminal silanol hydroxyl groups (Si–OH) on the zeolite matrix (Fig. 3B, I) (35, 37, 38). In addition, B/S-1 has another signal at 2.9 ppm related to the fully hydrolyzed boron species [e.g., $\text{B}(\text{OH})_3$], whereas the BS-1 exhibits four additional ^1H signals at 2.1, 2.6, 3.5, and 4.5 ppm, which are assigned to Si–OH adjacent to oxygen on B–OH (2.1 ppm, Si–OH...O(H)–B; Fig. 3B, II), Si–OH adjacent to boron (2.6 ppm; Fig. 3B, III), B–OH adjacent to oxygen on Si–OH [3.5 ppm, B–OH...O(H)–Si; Fig. 3B, II], and the hydroniums [4.5 ppm, $\text{H}^+(\text{H}_2\text{O})_n$; Fig. 3B, IV] (35, 37, 38). The signals at 3.5 and 2.1 ppm should be coproduced from the hydrolysis of the Si–O–B linkage in zeolite framework. The used BS-1 had similar signals at 3.5 and 2.1 ppm after the ODHP test (fig. S32),

confirming the high stability of the B–OH species during the ODHP.

Advanced 2D ^1H – ^1H double quantum (DQ) MAS NMR spectroscopy was used to explore the proton-proton proximities in the fresh and used BS-1 catalysts (Fig. 3C). The autocorrelation signal at (2.1, 4.2) ppm appeared on the diagonal, indicating that the H in Si–OH...O(H)–B were in close proximity to each other. A similar phenomenon also occurred in the B–OH...O(H)–Si species, with signal at (3.5, 7.0) ppm, where the H species were also in close proximity. In addition, the off-diagonal peak pair at (3.5, 5.6) and (2.1, 5.6) ppm indicates the spatial correlation between the Si–OH...O(H)–B and B–OH...O(H)–Si groups [δ_{DQ} (5.6 ppm) = 2.1 ppm + 3.5 ppm]. Considering that boron species are isolated in BS-1, the aforementioned signals should result from the $\text{B}[\text{OH}...\text{O}(\text{H})-\text{Si}]_2$ structure (Fig. 3B, II) (37). Compared with as-synthesized BS-1, the used sample exhibited enhanced signal intensity at (3.5, 5.6) ppm and (2.1, 5.6) ppm (Fig. 3C). This phenomenon is reasonably related to the further hydrolysis of Si–O–B bonds during the ODHP to form more spatially correlated B–OH and Si–OH species (e.g., isolated $\text{B}[\text{OH}...\text{O}(\text{H})-\text{Si}]_2$), which should be crucial for the high activity of BS-1 zeolite (37, 38). This phenom-

enon was also confirmed by FTIR spectroscopy of the fresh and used BS-1 catalyst (fig. S33), which explained the slight increase of propane conversion after a long-period ODHP test (Fig. 1D). More importantly, Si–OH nests (1.8, 3.6 ppm) and fully hydrolyzed boron species (2.9 ppm) were undetectable in the ^1H – ^1H DQ spectrum of the fresh and used BS-1 (Fig. 3C) (33), confirming that the deep hydrolysis of Si–O–B species and boron leaching were effectively inhibited (fig. S34); this was supported by the in situ FTIR study (fig. S35) showing that hydrolysis of the Si–O–B linkage is reversible. This feature contributes to the good durability during water treatment and reaction for long time.

The combined NMR and IR results indicate that the boron sites anchored within the MFI zeolite framework might proceed to hydrolysis to form more $-\text{B}[\text{OH}...\text{O}(\text{H})-\text{Si}]_2$ groups for boosting catalysis, and such species could reverse into the Si–O–B linkage, which is further supported by the reaction rates under a different atmosphere (fig. S36). The isotope-labeled experiment also confirmed the importance of hydroxyl groups for the reaction (fig. S36). This feature simultaneously provided the necessary active species for the reaction and hindered full hydrolysis and dissolution (figs.

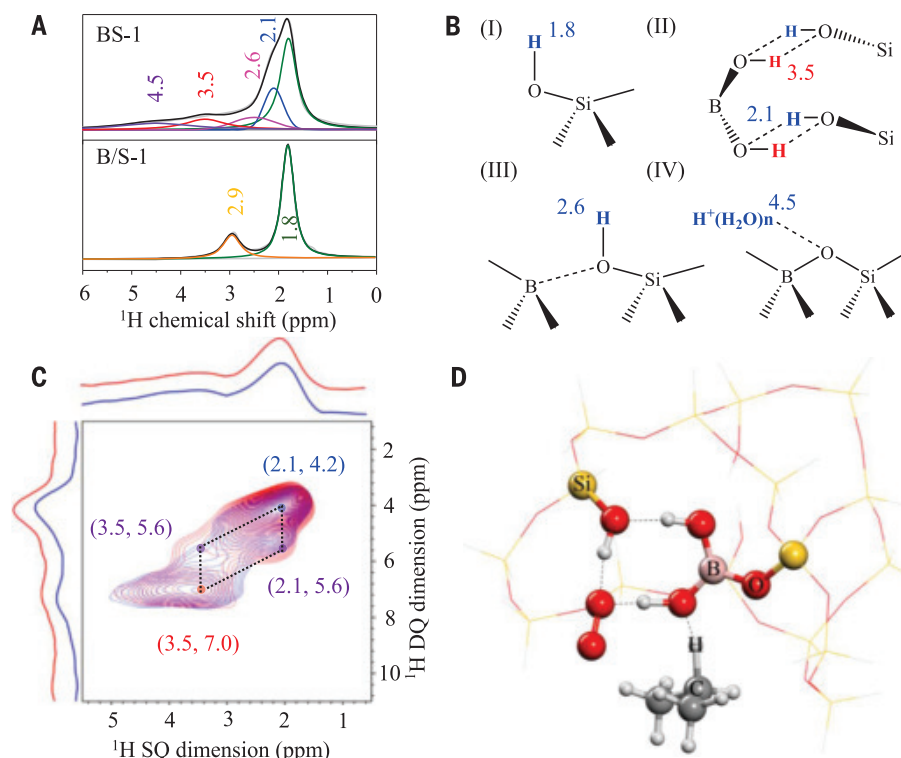


Fig. 3. Identification of boron hydroxyl groups in BS-1. (A) ^1H MAS NMR spectra of BS-1 and B/S-1. (B) Different H species and their chemical shifts. (C) 2D ^1H – ^1H DQ MAS NMR of BS-1 before (blue) and after (red) ODHP. (D) Transition state structure of $-\text{B}[\text{OH}...\text{O}(\text{H})-\text{Si}]_2$ reacting with oxygen and propane.

S37 and S38), giving good durability in the reaction process.

At the time of this investigation of boron-catalyzed ODHP, active sites are still not fully understood and controversy remains. Amorphous boron oxide species with multiple boron nuclei in close proximity (e.g., B–O–B species) have been proposed as active sites, whereas the isolated boron in zeolite is regarded to be inactive. For example, in a previous study, the MWW-type borosilicate zeolite (B-MWW) was observed to be catalytically inactive in the ODHP reaction (29), and this result was reproduced in our laboratory (figs. S39 to S42). However, we found that BS-1 had high activity, which differs substantially from previous knowledge. NMR characterizations showed that $\text{B}[\text{OH} \dots \text{O}(\text{H})\text{-Si}]_2$ sites were absent in the inactive B-MWW catalyst (fig. S41) (29), and such active sites might be related to the B [4] species in the zeolite framework because B[4] was nearly absent in B-MWW but abundant in BS-1 (fig. S42). It is possible that the $\text{B}[\text{OH} \dots \text{O}(\text{H})\text{-Si}]_2$ sites were from the hydrolysis of B[4] because B[4] is easily transformed into B[3] with abundant hydroxyl groups at higher temperature on BS-1 (figs. S24, S33, and S35). In addition, the borosilicate zeolite with a *BEA structure (B-Beta; fig. S43) also had abundant $\text{B}[\text{OH} \dots \text{O}(\text{H})\text{-Si}]_2$ sites exhibiting high activity and selectivity for ODHP (figs. S44 to S47). The kinetic parameters of B-Beta were in good agreement with those of BS-1 (fig. S48). The B-MWW has boron sites with a monohydroxyl group $[(\text{Si}-\text{O})_2\text{-B-OH}]$ but exhibits undetectable $\text{B}[\text{OH} \dots \text{O}(\text{H})\text{-Si}]_2$ species (29), which also supports the active site of dihydroxyl boron species. In addition, even if a small amount of $\text{B}(\text{OH})_3$ species exists on BS-1, such a species would not contribute to propane conversion, as supported by the extremely low activity of the B/S-1 catalyst (boric acid loaded on S-1 zeolite; Fig. 1A).

To better understand the active sites, the kinetic parameters in ODHP have been measured (figs. S49 and S50). The BS-1 catalyst had a first-order rate dependence with respect to propane partial pressure, whereas the supported boron and *h*-BN had a second-order rate dependence (fig. S49) (19). Although boron-catalyzed ODHP is still not fully understood (23), the first-order rate dependence might suggest a different reaction behavior on BS-1 compared with the generally supported boron and *h*-BN. The reaction pathway of propane dehydrogenation on BS-1 was explored using density functional theory calculation (figs. S51 to S54), where the catalyst framework was taken into account by the extended cluster model with isolated boron functionalized with one or two hydroxyl groups. For the catalyst with a dihydroxyl group on boron, O_2 and C_3H_8 were initially coadsorbed on one hydroxyl group through weak van der Waals interaction

(R; figs. S51 and S52). The other hydroxyl group in boron was able to form a hydrogen bond with Si–OH and further constrain O_2 , as shown in Fig. 3D. In this case, the constructed eight-member ring in the transition state by three hydrogen bonds benefits the synergistic conversion of O_2 ($^3\Sigma_g$) and propane through the decreasing of energy barriers, indicating the importance of $\text{B}(\text{OH})_2$ groups and the nearby silanol species (figs. S53 and S54). By contrast, the catalyst with monohydroxyl group on boron formed a weaker interaction between BOH and O_2 , as confirmed by the H...O distance and the interaction strength of 2.424 Å and 0.165 eV, which cannot build the desired transition state illustrated in the dihydroxyl system (fig. S55). In the reaction over dihydroxyl group model (fig. S51), the intermediate (**Int**) was formed with a barrier (**TS1**) of 1.583 eV through the proton transfer from propane to BOH with C_3H_7 intermediate formation, and the hydrogen atom of BOH synchronously interacted with O_2 . It is noteworthy that the C_3H_7 intermediate (**Int**) is highly reactive, and one hydrogen of the CH_3 group will quickly transfer to the O_2H species to form the products (**P**, i.e., C_3H_6 and H_2O_2) by overcoming an energy barrier (**TS2**) of 0.928 eV. Finally, C_3H_6 and H_2O_2 were desorbed in sequence. H_2O_2 is an unstable intermediate at the reaction temperature, which should quickly transform into H_2O and O_2 in the practical reaction. The confinement effect of the zeolite microporous structure also contributes to the reaction activity illustrated by the results in smaller model without a zeolite framework (fig. S56), which shows a higher TS1 energy barrier at 1.700 eV, as indicated by the stronger van der Waals interaction (fig. S57) in the reduced density gradient isosurface.

Our results demonstrate that the catalytic performances of isolated boron species could be improved by tuning the coordination environment, which provides new opportunities for the rational design of efficient dehydrogenation catalysts by incorporating boron sites in a zeolite framework. In addition, the zeolite structure hinders the full hydrolysis of boron species for excellent durability, which leads to an ideal catalyst with water tolerance. This work represents a large step forward in boron catalysis that may lead the way to scalable applications.

REFERENCES AND NOTES

- J. J. H. B. Sattler, J. Ruiz-Martinez, E. Santillan-Jimenez, B. M. Weckhuysen, *Chem. Rev.* **114**, 10613–10653 (2014).
- R. Ryoo et al., *Nature* **585**, 221–224 (2020).
- F. Jiao et al., *Science* **351**, 1065–1068 (2016).
- L. Liu et al., *Nat. Catal.* **3**, 628–638 (2020).
- L. Liu et al., *Nat. Mater.* **18**, 866–873 (2019).
- Q. Sun et al., *Angew. Chem. Int. Ed.* **59**, 19450–19459 (2020).
- K. Searles et al., *J. Am. Chem. Soc.* **140**, 11674–11679 (2018).

- H. Xiong et al., *Angew. Chem. Int. Ed.* **56**, 8986–8991 (2017).
- L. Shi et al., *Angew. Chem. Int. Ed.* **54**, 13994–13998 (2015).
- L. Zeng, Z. Cheng, J. A. Fan, L.-S. Fan, J. Gong, *Nat. Rev. Chem.* **2**, 349–364 (2018).
- A. S. Bodke, D. A. Olschki, L. D. Schmidt, E. Ranzi, *Science* **285**, 712–715 (1999).
- J. Zhang et al., *Science* **322**, 73–77 (2008).
- S. Chen et al., *J. Am. Chem. Soc.* **141**, 18653–18657 (2019).
- M. C. Simons et al., *J. Am. Chem. Soc.* **141**, 18142–18151 (2019).
- J. Tian et al., *Sci. Adv.* **5**, eaav8063 (2019).
- X. Li et al., *Angew. Chem. Int. Ed.* **55**, 4092–4096 (2016).
- X. Rozanska, R. Fortrie, J. Sauer, *J. Am. Chem. Soc.* **136**, 7751–7761 (2014).
- F. Cavani, N. Ballarini, A. Cericola, *Catal. Today* **258**, 226–229 (2015).
- J. T. Grant et al., *Science* **354**, 1570–1573 (2016).
- L. Shi et al., *Chin. J. Catal.* **38**, 389–395 (2017).
- W.-D. Lu et al., *ACS Catal.* **9**, 8263–8270 (2019).
- B. Qiu et al., *J. Catal.* **385**, 176–182 (2020).
- J. M. Venegas, W. P. McDermott, I. Hermans, *Acc. Chem. Res.* **51**, 2556–2564 (2018).
- J. A. Lolland, Z. Zhao, A. Patel, P. Hazin, *Ind. Eng. Chem. Res.* **58**, 2170–2180 (2019).
- L. Shi et al., *ChemCatChem* **9**, 1788–1793 (2017).
- A. M. Love et al., *J. Am. Chem. Soc.* **141**, 182–190 (2019).
- Z. Zhang, E. Jimenez-Izal, I. Hermans, A. N. Alexandrova, *J. Phys. Chem. Lett.* **10**, 20–25 (2019).
- A. M. Love et al., *J. Phys. Chem. C* **123**, 27000–27011 (2019).
- N. R. Altwater et al., *Angew. Chem. Int. Ed.* **59**, 6546–6550 (2020).
- P. Chaturvedi, M. Ahamed, M. Eswaramoorthy, *ACS Omega* **3**, 369–374 (2018).
- R. Millini, G. Perego, G. Bellussi, *Top. Catal.* **9**, 13–34 (1999).
- Y.-T. Angel Wong, D. L. Bryce, *Annu. Rep. NMR Spectrosc.* **93**, 213–279 (2018).
- S. Kroecker, J. F. Stebbins, *Inorg. Chem.* **40**, 6239–6246 (2001).
- C. Fild, H. Eckert, H. Koller, *Angew. Chem. Int. Ed.* **37**, 2505–2507 (1998).
- S.-J. Hwang, C.-Y. Chen, S. I. Zones, *J. Phys. Chem. B* **108**, 18535–18546 (2004).
- J. C. Jansen, R. de Ruiter, E. Biron, H. van Bekkum, *Stud. Surf. Sci. Catal.* **49**, 679–688 (1989).
- P. V. Wiper, J. Amelse, L. Mafra, *J. Catal.* **316**, 240–250 (2014).
- H. Koller, C. Fild, R. F. Lobo, *Micro. Meso. Mater.* **79**, 215–224 (2005).

ACKNOWLEDGMENTS

We thank Professor W. M. Lu for kindly helping with catalytic tests. **Funding:** This work was supported by the National Key Research and Development Program of China (2018YFB0604801), the National Natural Science Foundation of China (21822203, 21932006, 22032005, and U1908203), and the Natural Science Foundation of Zhejiang Province (LR18B030002). **Author contributions:** H.Z. performed the catalyst preparation, characterization, and catalytic tests and prepared the draft manuscript. X.Y., W.C., and A.Z. performed the NMR analysis and theoretical simulations. Y.H., Y.Q., X.C., and L.S. participated in the structure investigation. M.W. and J.M. performed the ToF study. Y.W., X.H., Z.C., H.W., Q.Z., and X.M. participated in zeolite synthesis. L.W. and F.-S.X. designed the study, analyzed the data, and wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/76/suppl/DC1
Materials and Methods
Figs. S1 to S57
References (39–55)

15 September 2020; accepted 9 February 2021
10.1126/science.abe7935

PESTICIDES

Applied pesticide toxicity shifts toward plants and invertebrates, even in GM crops

Ralf Schulz^{1,2*}, Sascha Bub¹, Lara L. Petschick¹, Sebastian Stehle^{1,2}, Jakob Wolfram¹

Pesticide impacts are usually discussed in the context of applied amounts while disregarding the large but environmentally relevant variations in substance-specific toxicity. Here, we systemically interpret changes in the use of 381 pesticides over 25 years by considering 1591 substance-specific acute toxicity threshold values for eight nontarget species groups. We find that the toxicity of applied insecticides to aquatic invertebrates and pollinators has increased considerably—in sharp contrast to the applied amount—and that this increase has been driven by highly toxic pyrethroids and neonicotinoids, respectively. We also report increasing applied toxicity to aquatic invertebrates and pollinators in genetically modified (GM) corn and to terrestrial plants in herbicide-tolerant soybeans since approximately 2010. Our results challenge the claims of a decrease in the environmental impacts of pesticide use.

Recent debates about the potential side effects of pesticides on humans (1) and the environment (2, 3) have been dominated by the comparison of use rates (e.g., kilograms per hectare) or applied amounts (e.g., kilograms per year) (4–9). These weight-based measures are not necessarily informative from an environmental perspective because toxicity among pesticides varies over several orders of magnitude (tables S1 to S3). This suggests that environmental effects strongly depend on the shares of individual pesticides in the total applied amount (1).

In this study, we extended a weight-based assessment of 381 pesticides for the years 1992 to 2016 (figs. S1 to S3) by 1591 regulatory threshold levels (RTLs) [as officially derived thresholds indicative of potential biodiversity impacts (3)] for eight different groups of nontarget species (10). We multiplied the annually applied amount (i.e., mass) of individual pesticides [data from the US Geological Survey (USGS)] with the reciprocal of the pesticide- and species group-specific RTLs (10) [data mainly from the US Environmental Protection Agency (EPA), see tables S1 and S2; species groups were unequally represented, see table S5] to derive the total applied toxicity (TAT) per substance, species group, and year (for TAT sensitivity, see fig. S4). The TAT is predictive of the potential pesticide impact (fig. S5). The annual TAT values were aggregated over different sets of substances (e.g., pesticide use types, chemical classes, and modes of action) to derive relative measures of temporal trends in agriculture both overall and for genetically modified (GM)-dominated crops specifically.

A comparison of the applied pesticide amount and the TAT reveals different temporal phases

for the different species groups. Regarding vertebrate toxicity (Fig. 1A), great reductions in acute toxicity have been achieved over the past few decades, driven almost entirely by insecticides (fig. S6A), whose TATs decreased by approximately a factor of 9 for mammals (Fig. 1C) and birds (Fig. 1D) through the replacement of organophosphorus and carbamate insecticides by pyrethroids and neonicotinoids (figs. S2 and S7, A and B). This development, which coincided with a proportional decrease in the applied amount (Fig. 1A, phase a) and an increase in corn acreage (fig. S10A), occurred in response to the high toxicity found in vertebrates (9). The fish TAT (TAT_{fish}) (Fig. 1B) remained constant overall since 2004 because of pyrethroid toxicity, which is relevant for this group (fig. S7C).

In sharp contrast, the invertebrate TAT has markedly increased since approximately 2005 (Fig. 1E, phase c). Both aquatic invertebrate TAT (TAT_{aqua-inverts}) and TAT_{pollinators} more than doubled, with an increase of ~8% per year between 2005 and 2015 (Fig. 1, F and G; fig. S6B; and fig. S8, A and C), whereas the terrestrial arthropod TAT (TAT_{terr-arthropods}) (referring here to nonpollinating species) increased less (Fig. 1H and figs. S6B and S8B; note, there is lower data availability for terrestrial arthropods, table S5). The TAT was driven solely by insecticides in all invertebrate groups (fig. S6B), coinciding with a proportional increase in cultivated area in relevant crops (figs. S10, B and C, and S11). Simultaneously, the applied insecticide amount decreased by ~40% (fig. S1B).

Although pollinators and aquatic invertebrates show similar temporal patterns regarding the applied amount of pesticides and TAT (Fig. 1E, phases b and c), the toxicities are driven by distinct classes of insecticides. For pollinators—e.g., bees or bumble bees—neonicotinoids are increasingly responsible for the TAT (Fig. 1G and fig. S8A). Neonicotinoids have been documented as being highly

toxic to bees (11, 12), and some of them have therefore been banned in the European Union (EU). Although seed treatments constitute >80% of all neonicotinoid use in the US (4) and restrictions on postbloom applications in perennial (tree) crops receiving spray applications have reduced pollinator risks (13), neonicotinoid use remains problematic (12), for example because of oral-based bee toxic load, which has increased particularly in heartland corn and soybeans (13). For aquatic invertebrates (e.g., crustaceans, mayflies, caddisflies, and dragonflies), pyrethroid insecticides have dominated the TAT since 1992, and they have also become increasingly relevant for terrestrial arthropods (nonpollinators such as mites, flies, and beetles; Fig. 1, F and H, and fig. S8, B and C). Figure 1E highlights a regime shift in the evolution of pesticide use in the mid-2000s (phase bc), when a phase of prevailing structural change in insecticide use (b) turned into a phase of TAT growth that is completely decoupled from trends in the total applied amount (c). Increases in applied pyrethroid toxicity have previously been implied only for fish (14). In the case of TAT_{aqua-inverts}, just four pyrethroids explained >80% of the increase since 2006. Because the detection limits of these four compounds in water are more than two orders of magnitude higher than their respective RTLs (table S3), it appears virtually impossible to track them at the entire range of ecologically relevant concentrations through scientific monitoring efforts (3, 15, 16). The highly effective, low-use rate insecticides (tables S3 and S4), often associated with an environmentally benign character (6, 9), increase in toxicity and use (fig. S12) and have the potential to be a considerable but widely unrecognized threat to both terrestrial and aquatic invertebrates (2, 3, 16).

The TAT for nontarget plants, which has been driven solely by herbicide use, showed an upward trend since approximately 2006 (Fig. 1I, phases a' and f, and figs. S6C and S9), likely related to resistance in crops (17). Although no single mode of action dominates plant toxicity, growth regulators (e.g., acetochlor) contribute mainly to the terrestrial and aquatic plant TAT, and amino acid synthesis inhibitors, such as the increasingly used glyphosate (fig. S3B) and cell membrane disruptors (e.g., oxyfluorfen), contribute to the TAT_{terr-plants} (Fig. 1, J and K, and fig. S9). The increases in plant TAT may have major impacts on terrestrial food webs, for example through reduced plant seed production (18) or plant species decline (19), requiring a systemic evaluation of previously unrecognized aspects of pesticide use.

Toxicity-weighted use is the strongest predictor of the potential impact of a pesticide on the environment (20). Its application in the

¹IES Landau, Institute for Environmental Sciences, University Koblenz-Landau, 76829 Landau, Germany. ²Eusserthal Ecosystem Research Station, University Koblenz-Landau, 76857 Eusserthal, Germany.

*Corresponding author. Email: schulz@uni-landau.de

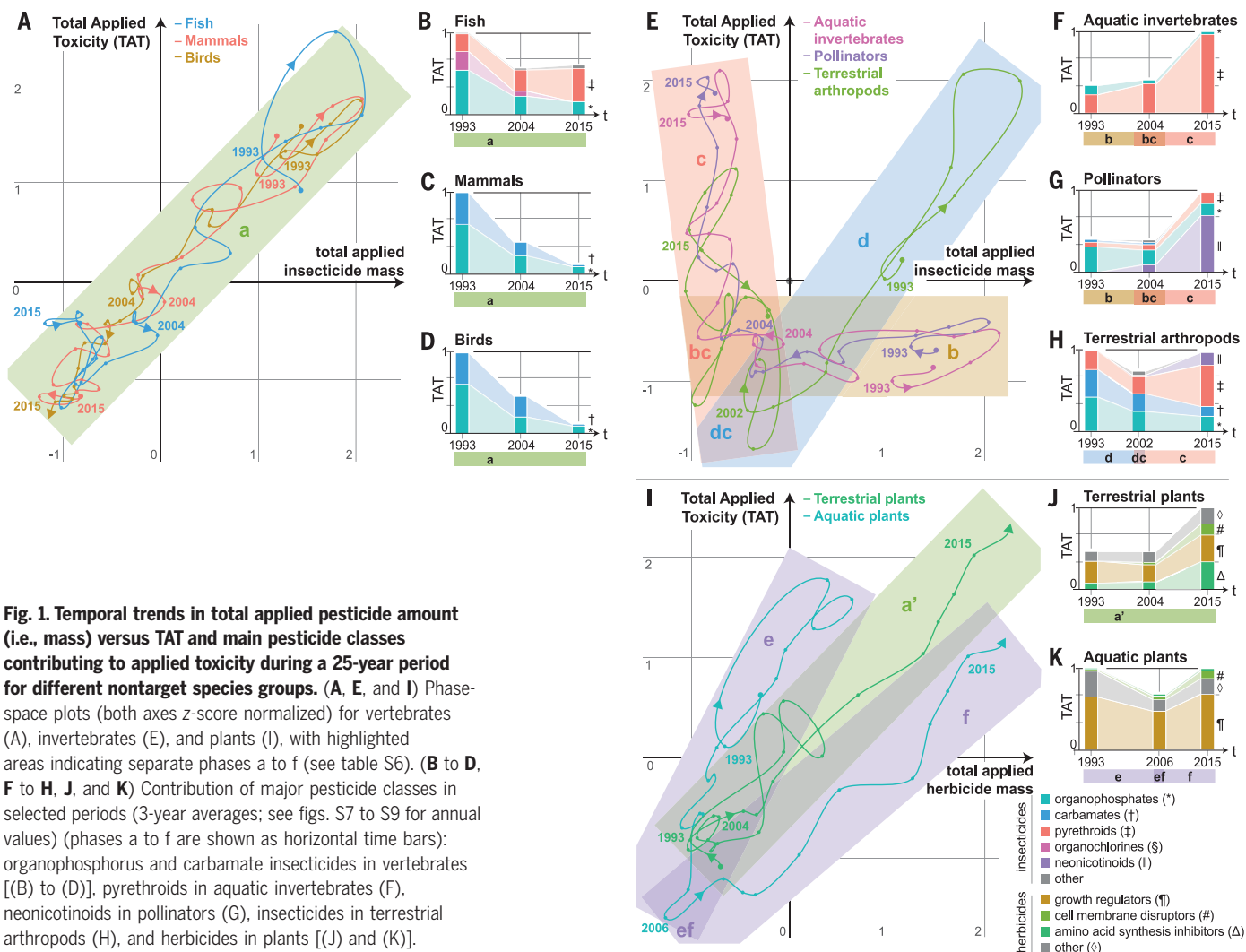


Fig. 1. Temporal trends in total applied pesticide amount (i.e., mass) versus TAT and main pesticide classes contributing to applied toxicity during a 25-year period for different nontarget species groups. (A, E, and I) Phase-space plots (both axes z-score normalized) for vertebrates (A), invertebrates (E), and plants (I), with highlighted areas indicating separate phases a to f (see table S6). (B to D, F to H, J, and K) Contribution of major pesticide classes in selected periods (3-year averages; see figs. S7 to S9 for annual values) (phases a to f are shown as horizontal time bars): organophosphorus and carbamate insecticides in vertebrates [(B) to (D)], pyrethroids in aquatic invertebrates (F), neonicotinoids in pollinators (G), insecticides in terrestrial arthropods (H), and herbicides in plants [(J) and (K)].

present study relies on the assumption that pesticide use and its effects on organisms are robustly connected to each other at large scales, even though there is tremendous variability in substance properties (table S4), application patterns, and local exposure situations. This assumption is, however, supported by multiple lines of evidence, even in the crucial case of pyrethroid risk to aquatic invertebrates. Monitoring data from a total of 89 available peer-reviewed studies [1977 insecticide concentrations from 231 different surface waters across the US (20, 21)] show that the rate at which measured insecticide and pyrethroid concentrations exceed the $RTL_{\text{aqua-inverts}}$ is significantly correlated with the applied toxicity to aquatic invertebrates (fig. S5). RTL exceedance in surface waters is indicative of negative effects on aquatic biodiversity (3), and pyrethroids exhibit the highest RTL exceedance rates (3, 21). Pyrethroids show adverse effects in midwestern streams (15), occur regularly in stream biofilms (22), and even cause resistance in nontarget freshwater am-

phipods (23). Taken together, multiple lines of evidence provide a clear link between the use of, exposure to, and effects of pyrethroid insecticides in aquatic systems. This link likely also applies to other pesticide and species groups, although further investigation in this field is needed.

From a broader perspective, decreases in vertebrate TAT were achieved at the cost of increased invertebrate TAT (Fig. 1, A and E). Additionally, ecologically linked pollinators and terrestrial plants (11, 24) are among those with the largest TAT increases (Fig. 1, E and I). The cumulative direct impact of modern insecticides on invertebrates and the indirect impact of herbicides on invertebrates through the food chain thus likely contribute to the currently debated decline in arthropods (2, 3, 5, 25–27). This decline may ultimately lead to indirect effects on vertebrate predators (2).

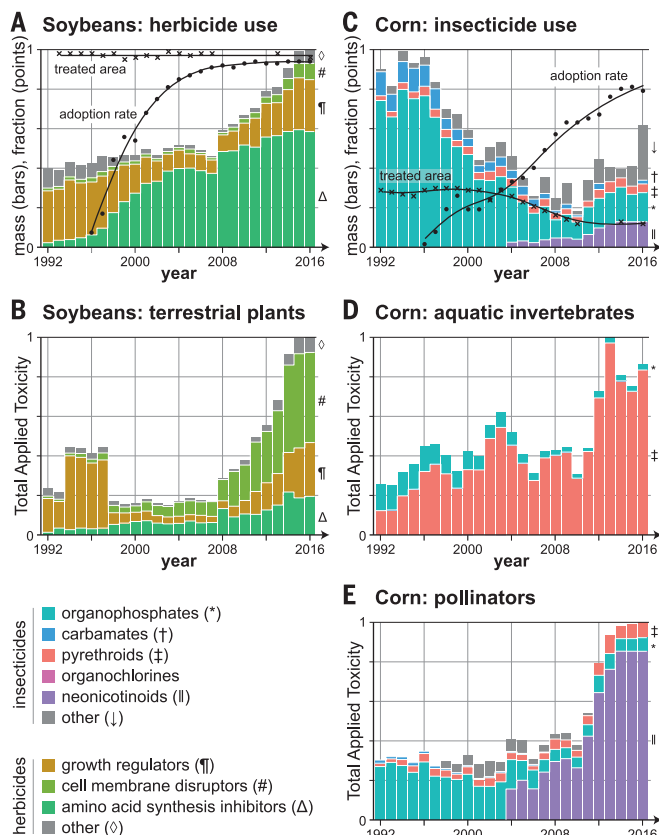
The TAT increased even in GM crops (Fig. 2, B, D, and F). Herbicide use has undergone substantial changes with the implementation of herbicide-tolerant GM crops (Fig. 2A),

which has led to a strong increase in the use of glyphosate (8, 28) (Fig. 2A and fig. S3B). $TAT_{\text{terr-plants}}$ has increased steadily since approximately 2008 for herbicides in herbicide-tolerant soybeans (Fig. 2B and fig. S14, A and B), likely in response to glyphosate resistance (17). However, downward trends have been reported for GM soybean herbicide toxicity to humans (1).

In the most widely grown GM crop that produces a *Bacillus thuringiensis* (Bt) toxin, corn, the insecticide TAT increased. Considering only data for corn, of which 79% in 2016 was Bt hybrids (Fig. 2C), TAT increased for both aquatic invertebrates (mainly because of pyrethroids; Fig. 2D) and terrestrial pollinators (mainly because of neonicotinoids; Fig. 2E) at the same rate observed for US agriculture as a whole (fig. S8, A and C). We verified that the toxicity per hectare of insecticides applied to Bt corn is equal to that for non-Bt corn (Fig. 2D, fig. S13, and fig. S14, C to F). The increasing insecticide TAT may be a result of preemptive, possibly unnecessary applications (4) or resistance (17). Our analysis suggests that

Fig. 2. Annual applied amount (i.e., mass) and nontarget species toxicity of the main classes of pesticides used in two of the most widely grown GM crops—soybeans and corn—in the US between 1992 and 2016.

(A) Amount of herbicides applied in soybeans (Mann-Kendall tests for monotonic trend: $\tau = 0.873$, $P < 0.001$). (B) TAT_{ter-plants} of 78 herbicides applied in soybeans ($\tau = 0.387$, $P = 0.129$). (C) Amount of insecticides applied in corn ($\tau = -0.553$, $P = 0.019$). (D) TAT_{aqua-inverts} of 72 insecticides applied in corn ($\tau = 0.5$, $P = 0.017$). (E) TAT_{pollinators} of 63 insecticides applied in corn ($\tau = 0.653$, $P = 0.009$). Lines in (A) and (C) were fitted using generalized additive models.



claims of reduced chemical insecticide use in US *Bt* crops (8, 9, 28) simply reflect the considerably lower application rates required for more recently developed, more toxic insecticide classes, whereas the TAT_{pollinators} and TAT_{aqua-inverts} both continue to increase.

TAT values may increase even more in global agriculture than they do in the US as important TAT drivers increase simultaneously. Such global developments include increased pesticide sales in Asia, Latin America, and Europe (9); expanded global cropland area (www.fao.org/faostat); and increased global pesticide use (29). Widespread resistance development (17), increasing field size (7), decreasing crop diversification (7), international market connectedness (30), and increasing temperatures (5) are key drivers of these developments. Although insecticide effects have been repeatedly documented (2, 15, 22), invertebrate biodiversity trends are generally debated (25, 26), and large-scale studies often do not focus on pesticides (25, 26, 31, 32). The unavailability of open-access pesticide use data in many regions, such as Latin America, the EU, China, and Russia, also prevents analyses such as those presented here, which potentially masks a crucial driver of the global biodiversity decline.

Despite being based on high-quality, curated data, our analysis carries some uncertainties.

The link between pesticide use and potential ecological impacts is formed by complex processes at different scales, and the methods applied here can only be indicative of the role pesticide use plays in the degradation of ecosystems, even at large scales. In light of the multiple emergent risks and resistance problems (6, 17), pesticide risks should be more integrated into policy strategies (33) to develop resilient global production systems (30, 31). Advancements in precision agriculture, mixed and organic farming, and nanoscale delivery platforms provide examples of how agriculture can develop productively while reducing environmental impacts (34, 35). Because 61% of US drinking water originates from surface waters, according to the EPA, TAT-based pesticide evaluations may also benefit human health evaluations. Overall, a system-centric view is conducive to understanding the dynamics of pesticide use on medium to large scales (36), and this is even more important given that ecosystem-wide pesticide impacts are becoming increasingly evident (2, 11, 12, 16).

REFERENCES AND NOTES

1. A. R. Kniss, *Nat. Commun.* **8**, 14865 (2017).
2. C. A. Hallmann, R. P. B. Foppen, C. A. M. van Turnhout, H. de Kroon, E. Jongejans, *Nature* **511**, 341–343 (2014).
3. S. Stehle, R. Schulz, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5750–5755 (2015).

4. M. R. Douglas, J. F. Tooker, *Environ. Sci. Technol.* **49**, 5088–5097 (2015).
5. J. A. Ewald et al., *Glob. Change Biol.* **21**, 3931–3950 (2015).
6. C. Lamberth, S. Jeanmart, T. Luksch, A. Plant, *Science* **341**, 742–746 (2013).
7. A. E. Larsen, F. Noack, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5473–5478 (2017).
8. C. D. Osteen, J. Fernandez-Cornejo, *Pest Manag. Sci.* **69**, 1001–1025 (2013).
9. E. Stokstad, G. Grullón, *Science* **341**, 730–731 (2013).
10. See the supplementary materials.
11. D. A. Stanley et al., *Nature* **528**, 548–550 (2015).
12. B. A. Woodcock et al., *Science* **356**, 1393–1395 (2017).
13. M. R. Douglas, D. B. Sponsler, E. V. Lonsdorf, C. M. Grozinger, *Sci. Rep.* **10**, 797 (2020).
14. S. Fong, S. Louie, I. Werner, R. E. Connon, *San Franc. Estuary Watershed Sci.* **14**, 1–34 (2016).
15. H. A. Rogers et al., *Environ. Sci. Technol.* **50**, 11974–11983 (2016).
16. I. Werner, T. M. Young, in *Encyclopedia of the Anthropocene*, D. A. Dellasala, M. I. Goldstein, Eds. (Elsevier, 2018), pp. 119–126.
17. F. Gould, Z. S. Brown, J. Kuzma, *Science* **360**, 728–732 (2018).
18. L. A. Morandin, M. L. Winston, *Ecol. Appl.* **15**, 871–881 (2005).
19. C. Stenoien et al., *Insect Sci.* **25**, 528–541 (2018).
20. J. Wolfram, S. Stehle, S. Bub, L. L. Petschick, R. Schulz, *Environ. Sci. Technol.* **53**, 12071–12080 (2019).
21. J. Wolfram, S. Stehle, S. Bub, L. L. Petschick, R. Schulz, *Environ. Sci. Technol.* **52**, 14452–14460 (2018).
22. B. J. Mahler et al., *Environ. Sci. Technol.* **54**, 5509–5519 (2020).
23. D. P. Weston et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16532–16537 (2013).
24. J. C. Biesmeijer et al., *Science* **313**, 351–354 (2006).
25. C. L. Outhwaite, R. D. Gregory, R. E. Chandler, B. Collen, N. J. B. Isaac, *Nat. Ecol. Evol.* **4**, 384–392 (2020).
26. S. Seibold et al., *Nature* **574**, 671–674 (2019).
27. C. J. Vorosmarty et al., *Nature* **467**, 555–561 (2010).
28. E. D. Perry, F. Ciliberto, D. A. Hennessy, G. Moschini, *Sci. Adv.* **2**, e1600850 (2016).
29. E. S. Bernhardt, E. J. Rosi, M. O. Gessner, *Front. Ecol. Environ.* **15**, 84–90 (2017).
30. M. Nyström et al., *Nature* **575**, 98–108 (2019).
31. D. Tilman et al., *Nature* **546**, 73–81 (2017).
32. R. van Klink et al., *Science* **368**, 417–420 (2020).
33. P. Kudsik, L. N. Jørgensen, J. E. Ørum, *Land Use Policy* **70**, 384–393 (2018).
34. P. Mäder et al., *Science* **296**, 1694–1697 (2002).
35. P. Vega-Vázquez, N. S. Mosier, J. Irudayaraj, *Front. Bioeng. Biotechnol.* **8**, 79 (2020).
36. C. J. Topping, A. Aldrich, P. Berry, *Science* **367**, 360–363 (2020).
37. R. Schulz, S. Bub, L. L. Petschick, S. Stehle, J. Wolfram, Applied pesticide toxicity shifts towards plants and invertebrates, even in GM crops, version v1, Zenodo (2021); <http://doi.org/10.5281/zenodo.4537036>.

ACKNOWLEDGMENTS

R. B. Schaefer, C. A. Bruehl, E. R. Bennett, I. Werner, and E. Bub provided feedback that strengthened this paper. J. P. Zubrod and D. W. Wong provided statistical advice. C. M. Cooper, a research ecologist at the USDA-ARS National Sedimentation Laboratory until 2008, is acknowledged as a pioneer in aquatic pyrethroid monitoring in the US. **Funding:** Research funding is provided by the DFG, German Research Foundation (grant no. SCHU 2271/6-2). The DBU, Deutsche Bundesstiftung Umwelt, provides support for L.L.P. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** R.S., S.B., S.S., and J.W. designed the research. R.S., S.B., L.L.P., and J.W. analyzed the data. R.S. and S.B. wrote the manuscript, and all authors contributed to editing the manuscript. **Competing interests:** S.B. is an employee of the University Koblenz-Landau and also works as a part-time freelance consultant in the field of ecotoxicology and environmental risk assessment. The authors declare no other competing interests. **Data and materials availability:** The USGS applied pesticide use data are available at <https://water.usgs.gov/nawqa/pnsp/usage/maps/>. The toxicity threshold levels (RTLs) used in this study are based on data from <https://cfpub.epa.gov/ecotox/> and <https://ecotox.ipmcenters.org>.

and have been supplemented for some species groups with data from www.efsa.europa.eu/en/data/chemical-hazards-data. The RTLs, the data for additional analysis regarding GM corn, and the code for this analysis can be found at <https://static.magic.eco/TAT> and are archived on Zenodo (37).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/81/suppl/DC1
Materials and Methods
Figs. S1 to S14
Tables S1 to S6

References (38–63)
MDAR Reproducibility Checklist

31 July 2020; accepted 19 February 2021
10.1126/science.abe1148

MARINE CONSERVATION

At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts

Casey C. O'Hara^{1*}, Melanie Frazier², Benjamin S. Halpern^{1,2}

Human activities and climate change threaten marine biodiversity worldwide, though sensitivity to these stressors varies considerably by species and taxonomic group. Mapping the spatial distribution of 14 anthropogenic stressors from 2003 to 2013 onto the ranges of 1271 at-risk marine species sensitive to them, we found that, on average, species faced potential impacts across 57% of their ranges, that this footprint expanded over time, and that the impacts intensified across 37% of their ranges. Although fishing activity dominated the footprint of impacts in national waters, climate stressors drove the expansion and intensification of impacts. Mitigating impacts on at-risk biodiversity is critical to supporting resilient marine ecosystems, and identifying the co-occurrence of impacts across multiple taxonomic groups highlights opportunities to amplify the benefits of conservation management.

The impact on the world's oceans of human activities, including fishing (1), land-based development and runoff (2), and ship strikes (3), coupled with the accelerating effects of climate change (4), are pervasive and increasing (5). Impacts from these anthropogenic stressors threaten marine species across taxa, driving thousands toward extinction (6, 7) and jeopardizing the sustainability of coastal social-ecological systems (7, 8).

Species respond differently to stressors, and multiple stressors can have cumulative impacts on threatened marine species (9). Efforts to assess cumulative human impacts on marine species have been single snapshots in time limited to a few specific taxa and stressors [e.g., (10–13)], leaving most species unassessed. A recent comprehensive, species-level assessment of cumulative impacts on at-risk terrestrial vertebrates (14) did not include marine species and did not capture changes in impact over time. Assessments of the distribution and rate of change of cumulative human impacts on global marine habitats (5, 15) provide valuable insights into ecosystem-level concerns, but habitat-focused assessments do not capture the heterogeneity of species' vulnerability (4, 11), which is crucial for designing conservation strategies.

Here, we present a global assessment of cumulative human impacts on at-risk marine spe-

cies and changes in those impacts over a recent time period. For each of 1271 threatened and near-threatened marine species comprehensively assessed and mapped for the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (16)

(hereafter “at-risk species”), we identified sensitivity to 14 anthropogenic stressors. We then intersected species range maps with relevant maps of annual stressor intensity from 2003 to 2013 to determine the extent of potential impacts [as in (17); hereafter simply “impacts”] across species' ranges, as well as how rapidly these impacts have been expanding in extent and increasing in intensity.

Mapping potential impacts to at-risk species requires understanding which stressors threaten the species (sensitivity) and where those stressors overlap the species' range (exposure) (17). We identified sensitivity to various stressors for each at-risk species based on threat information from IUCN Red List assessments (16). Of the 1271 marine species identified as at risk, 1036 (82%) are sensitive to one or more of our suite of 14 anthropogenic stressors (tables S1 and S2), with 865 species (68%) being sensitive to multiple stressors (Fig. 1). The remaining 235 species (18%) are not classified as sensitive to these stressors, but rather as either sensitive to others (e.g., invasive species, terrestrial hunting) or having insufficient information to

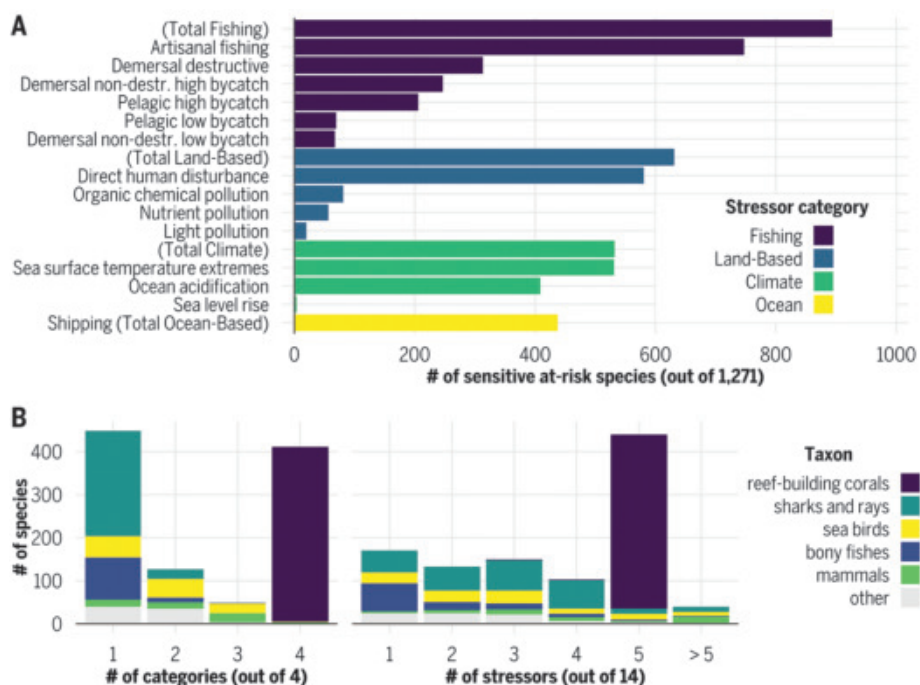


Fig. 1. Number of stressors and stressor categories (fishing, ocean, land-based, and climate) affecting at-risk species. (A) Counts of species classified as sensitive to each anthropogenic stressor or category; category totals count species sensitive to one or more stressors in the category. **(B)** Counts of species by number of stressor categories (left) or stressors (right) to which each is sensitive; the five largest taxa are highlighted.

¹Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA, USA. ²National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, USA.

*Corresponding author. Email: cohara@bren.ucsb.edu

determine sensitivity. The greatest proportion of at-risk species are sensitive to artisanal fishing (59%), direct human disturbance (e.g., trampling or coastal development, 46%), and sea surface temperature extremes (42%). Overall, 70% of at-risk species are sensitive to one or more fishing stressors (Fig. 1).

We then assessed where the range for each at-risk species intersected with the spatial extent of stressors to which it is sensitive (i.e., the footprint of potential impacts on species range, hereafter “affected range”) and found highly heterogeneous patterns, with a much higher number of affected species occurring in the Central Indo-Pacific and Coral Triangle regions (Fig. 2A). This result agrees with general understandings of global marine species richness (18) and patterns of threatened status of marine species (19). Adjusting for local richness of at-risk species (fig. S1), we found additional areas with a high proportion of affected species in the North Atlantic, North Sea, and Baltic Sea; international waters in the eastern Atlantic; and the western Pacific and tropical Indo-Pacific (Fig. 2B).

Areas with a high proportion of affected at-risk species ($\geq 50\%$ of species present in a cell were affected) cover 22% of the global ocean, whereas areas with a low proportion of impacts ($\leq 10\%$ of species affected) cover 26% (Fig. 2B). These regions represent areas of particularly high concern and potentially lower concern, respectively, for managing at-risk species. In 14% of the ocean, including some high-biodiversity areas in Australia’s northern waters, no at-risk species are affected (Fig. 2B), highlighting potential refugia. Fishing stressors dominate impact footprints in most national exclusive economic zones (EEZs); however, there are areas where this pattern is reversed and low fishing effort within certain EEZs gives way to unrestricted effort in areas beyond national jurisdiction (fig. S3A). The footprint of climate stressors on at-risk species range is particularly notable in temperate and polar regions, as well as in the Coral Triangle region (fig. S3B).

Human impacts on at-risk species are changing over time. From 2003 to 2013, impacts were intensifying (i.e., one or more stressors significantly increasing in intensity faster than 0.1% per year) for at least one at-risk species in 70% of the global ocean, and in 4% of the ocean, a high proportion ($>50\%$) of species were experiencing intensifying impacts (Fig. 2C and figs. S2 and S4 by stressor group). Only 4% of the ocean had areas where impacts were abating for at least one at-risk species, and a high proportion of abatement occurred in only 0.5% of the ocean (Fig. 2C and fig. S2).

The footprint of impacts on species ranges was extensive and varied considerably by taxonomic group (fig. S5). In the most recent year of assessment (2013), impacts occurred across

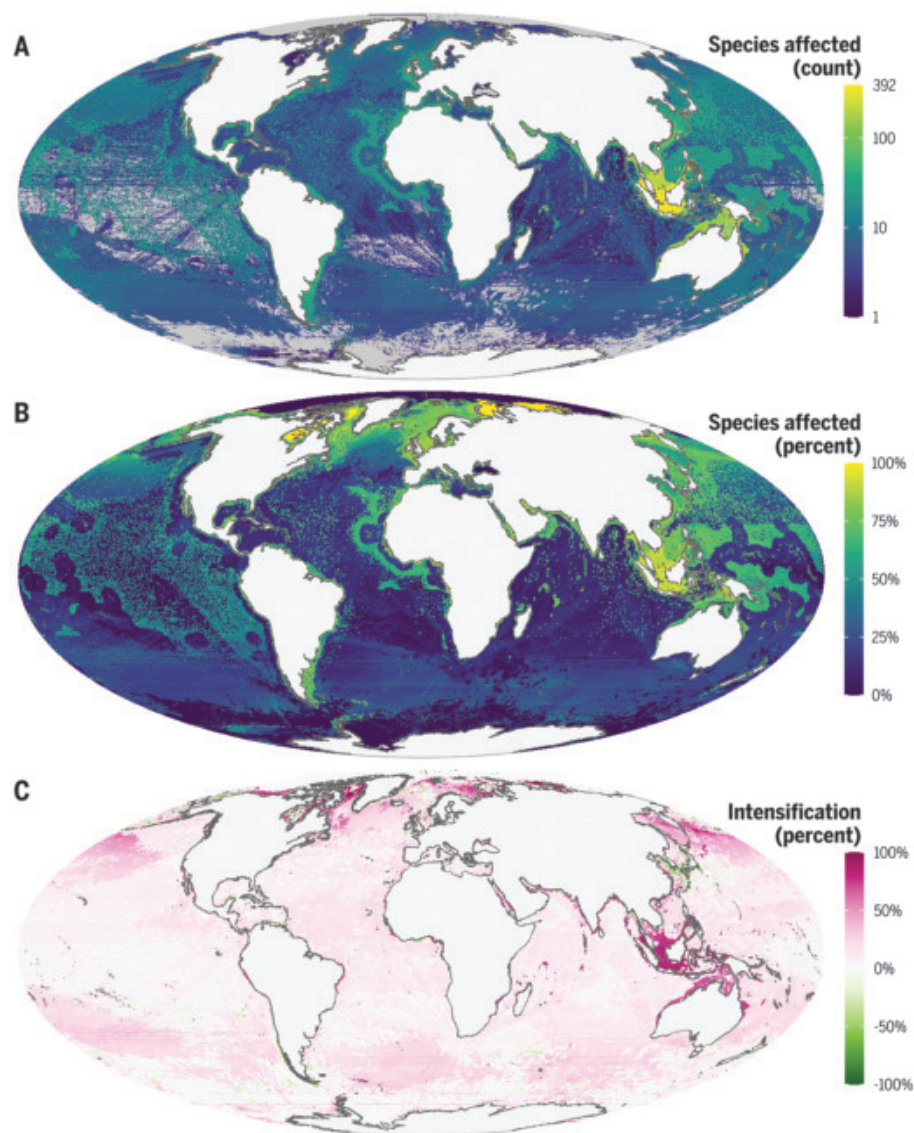


Fig. 2. Proportion of species affected and with intensifying impacts. (A) Number of threatened species affected by one or more stressors in 2013, the most recent year of assessment; gray indicates no affected species. (B) Proportion of threatened species relative to at-risk species richness affected by one or more stressors. (C) Net proportion of affected at-risk species in which stressors intensified at a rate $\geq 0.1\%$ per year over the period 2003 to 2013. See fig. S2 for insets highlighting areas of high intensification and abatement.

$57 \pm 42\%$ (mean \pm SD; median 73%) of the total range of at-risk marine species, with a mean of $19 \pm 35\%$ of range affected by two or more stressors (Fig. 3A). Impacts exceeded half the total range for 59% ($n = 744$) of at-risk species and 90% of the total range for 42% ($n = 540$) of species; the entire range was affected for 7% ($n = 92$) of species. Corals and mangroves bore the largest cumulative impact footprints ($99 \pm 2\%$ and $89 \pm 18\%$ of range, respectively).

Because greater exposure to pressures likely increases extinction risk, there is an expectation that the percentage of range affected

should correlate positively with IUCN risk category. Such a relationship was evident for small-ranged species (i.e., those with range areas in the bottom quartile, $<113,000 \text{ km}^2$) (fig. S6A). However, this correlation broke down for middle-ranged species (second quartile), and for large-ranged species, the affected range ostensibly correlated negatively with extinction risk. A likely driver of these results is that, as range size increases, extinction risk becomes predicted less by overall impact footprint and more by impacts on critical habitats or life stages (20, 21), particularly as stressors and species

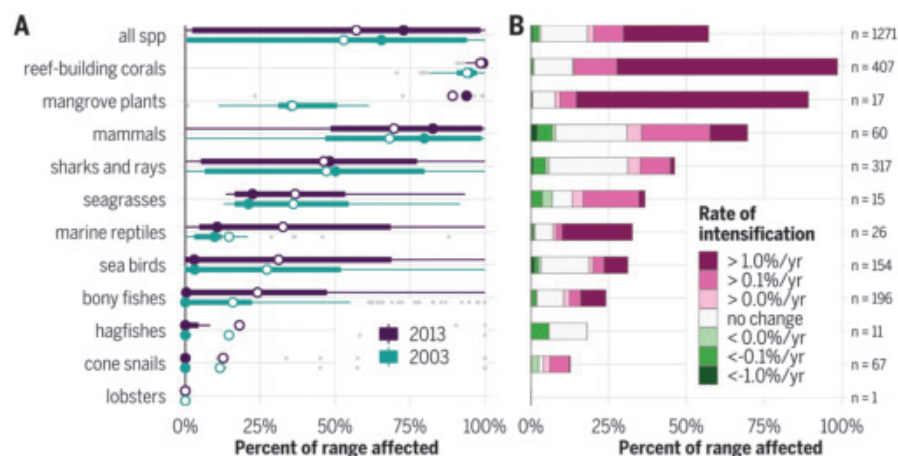


Fig. 3. Proportion of range affected, including intensifying and abating impacts, by taxonomic group. (A) Boxplots of affected range and expansion per taxon for 2003 and 2013. Hollow point is mean, solid point is median, thick line spans the interquartile range (IQR), and thin lines extend to the last observation within $1.5 \times \text{IQR}$. Gray points represent outliers. (B) Taxon-level mean affected range (2013) and average portion of that range intensifying or abating.

are concentrated in coastal areas. Focusing on species-stressor interactions in neritic waters, the results for small-ranged species were essentially unchanged, but the counterintuitive patterns for larger-ranged species were subdued (fig. S6B). Large-ranged species at lower extinction risk were dominated by coast-hugging corals, which were widely affected by the stressors included in this assessment, whereas large-ranged Endangered and Critically Endangered species included a higher proportion of pelagic-ranging turtles and sea birds harmed by terrestrial threats, e.g., hunting and invasive species, which are not assessed here (fig. S6C).

Across all included species, the average impact footprint increased over time, from $53 \pm 41\%$ in 2003 to $57 \pm 42\%$ in 2013 (Fig. 3A) but varied substantially among taxonomic groups. Mangroves (impact footprint $+53\%$), marine reptiles ($+18\%$), and bony fishes ($+8\%$) showed marked expansion of the mean impact footprint from 2003 to 2013; the mean impact footprint for mammals, seagrasses, cone snails, and hagfishes did not significantly change. The impact footprint of sharks and rays showed a mild contraction over time, driven by shifting patterns of fishing pressure.

On average, the impacts experienced by at-risk species intensified faster than 0.1% per year in $37 \pm 39\%$ of their ranges and faster than 1% per year in $27 \pm 35\%$ of their ranges. Overall, only $2 \pm 6\%$ of species' ranges experienced abating impacts and only $1 \pm 2\%$ abated rapidly (Fig. 3B). Mangrove plants and corals in particular experienced intensifying impacts across their ranges ($80 \pm 16\%$ and $85 \pm 11\%$, respectively), largely driven by climate stressors. Although the impact footprint on sharks and rays contracted by a small amount overall (Fig. 3A), impacts intensified over $11 \pm 15\%$ of their ranges (Fig. 3B), particularly from small-scale fishing. On average, the intensifying range exceeded the abating range by a factor of 15.

Although species ranges are dictated by ecological boundaries, effective management of activities that affect those species is dictated by political boundaries. At-risk species in the eight most-affected EEZs on average suffered impacts across 88% of their ranges within those EEZs. In the largest EEZs, the mean impact footprint varied considerably, e.g., Indonesia (84%) versus French Polynesia (12%). Uninhabited or sparsely inhabited islands made up the regions with the smallest mean footprint; within the eight least-affected EEZs, species were on average affected across 8% of their ranges (Fig. 4 and table S3).

Fishing activity, because of its broad reach (fig. S3) and preponderance of sensitive species (Fig. 1), was the dominant contributor to cumulative impact to species' ranges in most EEZs, with some notable exceptions (e.g.,



Fig. 4. Mean footprint of impacts on at-risk species ranges by EEZ. Each panel shows the area-weighted mean proportion of species range affected within each EEZ by stressor category. Yellow and green bars indicate the eight EEZs with the highest and lowest mean cumulative impact footprint, respectively. Purple bars indicate the eight largest EEZs by area, which do not coincide with the eight highest- or lowest-scoring EEZs. Narrow gray bars indicate other EEZs within each geographic region.

Jordan and Australia; Fig. 4). Climate impacts were the second-largest contributor, followed by ocean-based impacts. Land-based impacts were the most extreme in very small EEZs (e.g., Jordan, Singapore, Oecussi Ambeno, and Bosnia-Herzegovina; Fig. 4) but in general were the smallest contributors to species range impacts because they were limited to coastal areas and absent for uninhabited islands.

IUCN Red List assessments (16) have long provided critical information on the status and key threats to at-risk species. Leveraging that work and recent advancements in mapping the location and pace of change of human-induced stressors to the oceans (5), we show where species are being affected and how the impacts are changing in intensity over time. This spatially and temporally resolved information helps to inform conservation strategies aimed at species and locations with the greatest conservation need. It can also inform effective ecosystem-based management strategies such as protections for flagship or keystone species or taxa-specific mandates such as the Marine Mammal Protection Act of the United States, which leverage particular species to benefit the ecosystem more broadly. Co-occurrence analysis of taxa-level impacts highlights where such species-based strategies potentially confer co-benefits broadly across the ecosystem. For example, conservation efforts to reduce pressures on corals can simultaneously generate considerable benefits for marine mammals, bony fishes, sharks, seabirds, and marine reptiles (fig. S7 and table S4). Additionally, by assessing at-risk marine species across all comprehensively assessed taxa, our approach provides a window into broader ecosystem health with greater resolution into impacts across ecosystem structure complementary to impacts mapped onto representative habitats. Stressors that selectively affect some species over others potentially disrupt the “biostructure” of an ecosystem (22), resulting in reductions in the biomass of exploited species (23), reduced ecosystem functioning (22, 24), and general loss of resilience (25) that can lead to ecosystem collapse to an undesirable stable state (25–27).

Reactive conservation measures are urgent where impacts on at-risk species are pervasive and intensifying (28) to allow for ecosystem recovery (29). Of particular concern is the tropical Indo-Pacific, where accelerating climate impacts are exacerbated by intensifying fishing, shipping, and land-based stressors affecting most species (fig. S4). Areas of low and/or abating impacts may indicate opportunities for proactive conservation to maintain existing patterns and trends (28); e.g., the legal designation of the Phoenix Islands Protected Area in 2008 locked in already low impacts to species and enabled further reductions in impacts over time.

Because most marine species ranges cross international boundaries (30), effective conservation in one country may be undone by ineffective management in the next, and the fate of an at-risk species depends on managing impacts throughout its range. For example, despite low fishing pressure in Jordan's waters (mean impact range due to fishing, 18%; Fig. 4), higher fishing pressures in bordering Egypt (32%), Israel (68%), and Saudi Arabia (82%) (table S3) may reduce the capacity of at-risk populations to rebuild after a regional disturbance. Quantifying the effects of marine, land-based, and climate change stressors helps to link drivers of impacts to the management actions best suited to address them (31, 32). Because climate impacts do not respect political boundaries, it is especially important to manage those impacts that can respond to localized policy, e.g., marine protected areas or fisheries management, to improve the resilience of at-risk species and populations to climate change (33).

Within the 1271 species included in our study, marine vertebrates are well represented, and this includes most large marine predators (e.g., sharks, cetaceans), which are widely considered as useful proxies for ecosystem health (34) (table S5). Habitat-building species are also well represented, including reef-building corals, seagrasses, and mangroves (table S5). Sharks and rays ($n = 312$) and corals ($n = 407$) together represent 56% of the at-risk species included. Although range maps and occurrence data exist for many more marine species than are included here, our dataset is limited by current state of knowledge of threatened species and species-specific sensitivity to anthropogenic stressors. Whereas frameworks have been developed to estimate species sensitivity to climate stressors [e.g., (4, 17)], a general framework for estimating species sensitivity to a comprehensive set of stressors that is based on physiological and life history traits would enable a thorough global assessment across many more species and taxonomic groups.

Our analysis reveals that human activity and climate change are affecting at-risk marine species within most of the global ocean and across most of their ranges, and these impacts are expanding and increasing in intensity for most species. However, areas of the ocean remain that harbor at-risk species free of impacts, including areas rich in biodiversity. If we hope to reverse the course of species extinction and recover populations of at-risk species, then we need to know where species are exposed to the threats to which they are sensitive and how those threats are changing. Our results provide that information and can be embedded within a wide range of management and conservation strategies, including marine protected areas, fisheries reform, land-sea conservation, and climate change mitigation efforts.

REFERENCES AND NOTES

1. R. L. Lewison et al., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5271–5276 (2014).
2. C. J. Brown et al., *J. Appl. Ecol.* **56**, 1106–1116 (2019).
3. R. P. Schoeman, C. Patterson-Abrolat, S. Plön, *Front. Mar. Sci.* **7**, 292 (2020).
4. M. C. Jones, W. W. L. Cheung, *Glob. Chang. Biol.* **24**, e719–e731 (2018).
5. B. S. Halpern et al., *Sci. Rep.* **9**, 11609 (2019).
6. S. H. M. Butchart et al., *Science* **328**, 1164–1168 (2010).
7. S. Diaz et al., *Science* **366**, eaax3100 (2019).
8. N. J. Bennett, *Coast. Manage.* **47**, 244–252 (2019).
9. C. M. Crain, K. Kroeker, B. S. Halpern, *Ecol. Lett.* **11**, 1304–1315 (2008).
10. S. M. Maxwell et al., *Nat. Commun.* **4**, 2688 (2013).
11. I. C. Avila, K. Kaschner, C. F. Dormann, *Biol. Conserv.* **221**, 44–58 (2018).
12. N. Queiroz et al., *Nature* **572**, 461–466 (2019).
13. J. Stockbridge, A. R. Jones, B. M. Gillanders, *Sci. Rep.* **10**, 11934 (2020).
14. J. R. Allan et al., *PLOS Biol.* **17**, e3000158 (2019).
15. B. S. Halpern et al., *Nat. Commun.* **6**, 7615 (2015).
16. International Union for Conservation of Nature and Natural Resources, “The IUCN Red List of Threatened Species, Version 2020-1” (2020); <https://api3.iucnredlist.org/>.
17. S. E. Williams, L. P. Shoo, J. L. Isaac, A. A. Hoffmann, G. Langham, *PLOS Biol.* **6**, 2621–2626 (2008).
18. E. R. Selig et al., *PLOS ONE* **9**, e82898 (2014).
19. C. C. O'Hara, J. C. Villaseñor-Derbez, G. M. Ralph, B. S. Halpern, *Conserv. Lett.* **12**, 12651 (2019).
20. J. L. Payne, S. Finnegan, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10506–10511 (2007).
21. K. S. Collins, S. M. Edie, G. Hunt, K. Roy, D. Jablonski, *Proc. Biol. Sci.* **285**, 20181698 (2018).
22. K. McCann, *Nature* **446**, 29 (2007).
23. M. L. D. Palomares et al., *Estuar. Coast. Shelf Sci.* **243**, 106896 (2020).
24. C. E. Benkwitt, S. K. Wilson, N. A. J. Graham, *Nat. Ecol. Evol.* **4**, 919–926 (2020).
25. C. Folke et al., *Annu. Rev. Ecol. Syst.* **35**, 557–581 (2004).
26. T. P. Hughes, *Science* **265**, 1547–1551 (1994).
27. K. Filbee-Dexter, R. E. Scheibling, *Mar. Ecol. Prog. Ser.* **495**, 1–25 (2014).
28. T. M. Brooks et al., *Science* **313**, 58–61 (2006).
29. J.-C. Ortiz et al., *Sci. Adv.* **4**, eaar6127 (2018).
30. L. Roberson et al., *Preprints* 2020080525 (2020).
31. J. G. Álvarez-Romero et al., *Annu. Rev. Ecol. Syst.* **42**, 381–409 (2011).
32. V. J. D. Tulloch et al., *Biol. Conserv.* **245**, 108527 (2020).
33. U. R. Sumaila, T. C. Tai, *Front. Mar. Sci.* **7**, 523 (2020).
34. F. Sergio et al., *Annu. Rev. Ecol. Syst.* **39**, 1–19 (2008).
35. Code and results from data analysis for: C. C. O'Hara, M. Frazier, B. S. Halpern, At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts, Knowledge Network for Biocomplexity (2020); <https://knbi.econinformatics.org/view/doi:10.5063/SJ1J03>.

ACKNOWLEDGMENTS

We thank the National Center for Ecological Analysis and Synthesis (NCEAS) for computational support. **Funding:** We gratefully acknowledge financial support from NCEAS, the National Philanthropic Trust, and a fellowship from the Bren School of Environmental Science and Management. **Author contributions:** All authors conceptualized the study goals and methodology. C.C.O. performed the analysis, software coding, data curation, and visualization. B.S.H. provided supervision. C.C.O. wrote the initial draft, and all authors contributed to reviewing and editing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All raw data are freely available from the original sources. All code and results from this analysis are available at the Knowledge Network for Biocomplexity (35).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/84/suppl/DC1
Materials and Methods
Tables S1 to S5
Figs. S1 to S7
References (36–46)
MDAR Reproducibility Checklist

8 September 2020; accepted 3 March 2021
10.1126/science.abe6731

QUANTUM OPTICS

Observation of a non-Hermitian phase transition in an optical quantum gas

Fahri Emre Öztürk¹, Tim Lappe², Göran Hellmann¹, Julian Schmitt^{1*}, Jan Klaers^{1†}, Frank Vewinger¹, Johann Kroha², Martin Weitz^{1*}

Quantum gases of light, such as photon or polariton condensates in optical microcavities, are collective quantum systems enabling a tailoring of dissipation from, for example, cavity loss. This characteristic makes them a tool to study dissipative phases, an emerging subject in quantum many-body physics. We experimentally demonstrate a non-Hermitian phase transition of a photon Bose-Einstein condensate to a dissipative phase characterized by a biexponential decay of the condensate's second-order coherence. The phase transition occurs because of the emergence of an exceptional point in the quantum gas. Although Bose-Einstein condensation is usually connected to lasing by a smooth crossover, the observed phase transition separates the biexponential phase from both lasing and an intermediate, oscillatory condensate regime. Our approach can be used to study a wide class of dissipative quantum phases in topological or lattice systems.

Creating and understanding phases of systems that are dissipatively coupled to the environment is of importance in research fields ranging from optics to biophysics (1–6). One intriguing aspect of this openness is the possible existence of quantum states that are not otherwise accessible (7–10). Near-equilibrium physics (11, 12) has been studied in optical quantum gases (13), such as photons or polaritons (strongly coupled, mixed states of light and matter), despite their driven-dissipative nature. In

particular, Bose-Einstein condensates of photons, realized in dye-filled microcavities by multiple photon absorption and reemission cycles, provide a platform to study quantum dynamics in an open, grand canonical situation where the condensate particles are coupled to a reservoir of the photoexcitable dye molecules (14). Photon condensates have the macroscopic mode occupation in common with lasers, but they operate near thermal equilibrium, in distinct contrast to lasers. Naïvely, a smooth crossover between lasing and condensation

might be expected given that both phenomena exhibit spontaneous symmetry breaking (8, 15).

Recently, oscillatory dynamics in open dye microcavity systems have been observed (6, 16), a phenomenon that at large resonator losses crosses over to the relaxation oscillations known in laser physics. Unlike in a laser, the stochastic driving induced by grand canonical condensate fluctuations makes the system dynamics observable in stationary-state operation, which characterize the system's state by its second-order coherence. In contrast to closed systems governed by time-reversal symmetric—i.e., Hermitian—dynamics, the dissipative coupling to the environment is described by a non-Hermitian time-evolution operator with complex eigenvalues. Of special interest are exceptional points, where the eigenvalues and the corresponding eigenmodes coalesce (1, 17–20). Such points are well known to enable phase transitions (21); a first-order phase transition between a photon laser and a polariton condensate has recently been proposed (8, 22).

¹Institut für Angewandte Physik, Universität Bonn, Wegelerstr. 8, 53115 Bonn, Germany. ²Physikalisches Institut, Universität Bonn, Nussallee 12, 53115 Bonn, Germany.

*Corresponding author. Email: schmitt@iap.uni-bonn.de (J.S.); martin.weitz@uni-bonn.de (M.W.)

†Present address: Complex Photonic Systems (COPS), MESA+ Institute for Nanotechnology, University of Twente, Drienerlolaan 5, 7522 NB Enschede, Netherlands.

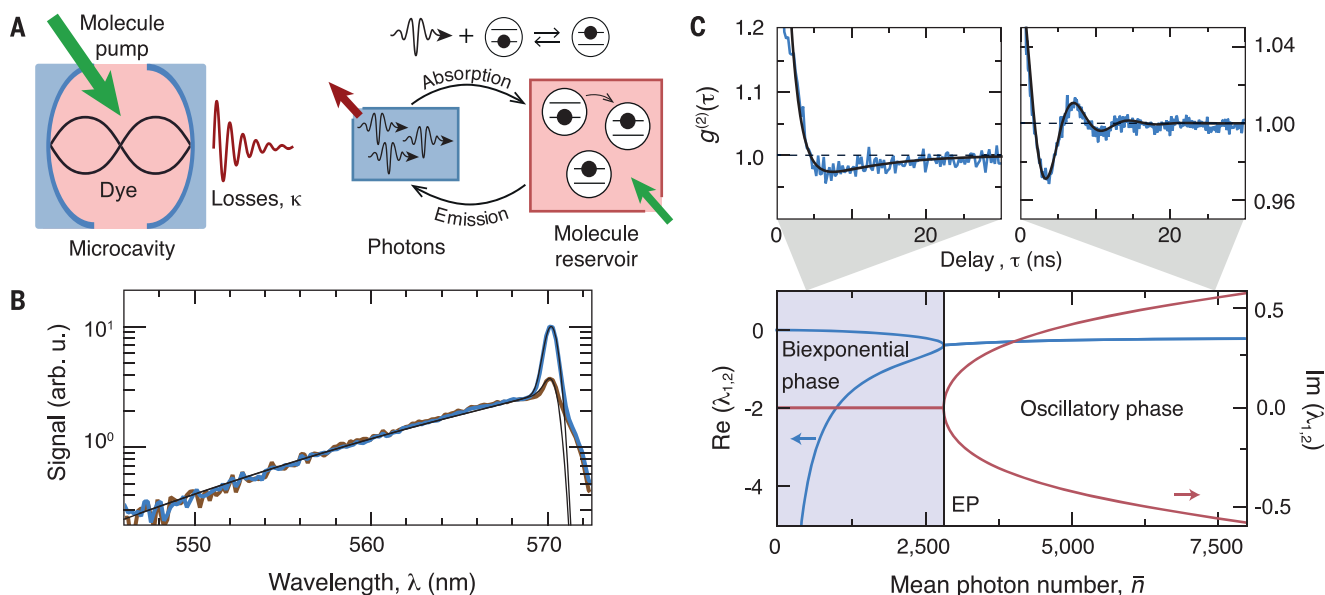


Fig. 1. Experimental principle. (A) Photons are trapped within a dye-filled microcavity, where losses κ are compensated by pumping the dye molecules with a laser. The photon gas is coupled to this reservoir by the exchange of excitations between photons and electronically excited dye molecules (right panel). (B) Spectra of the emission for average photon numbers $\bar{n} \approx 2100$ (orange line) and 10,300 (blue line), showing a thermalized photon gas with a condensate peak at the position of the low-energy cutoff, closely following the expected (experimentally

broadened) Bose-Einstein distributions at 300 K (black lines). arb. u., arbitrary units. (C) Second-order correlations $g^{(2)}(\tau)$ of the condensate, recorded at $\bar{n} \approx 2300$ (left) and $\bar{n} \approx 14,000$ (right), with fitted theory curves (black lines) (25), showing oscillatory behavior for large photon numbers and a biexponential decay for small photon numbers. The bottom panel shows predictions of real (blue) and imaginary (red) parts of the eigenvalues $\lambda_{1,2}$ (for a molecule number $M = 5 \times 10^9$), which are real below the exceptional point (EP) and complex above it.

To prepare an open photon Bose-Einstein condensate coupled to a reservoir, we use a dye microcavity apparatus (11, 23–25) (Fig. 1A). The short mirror spacing of a few wavelengths discretizes the longitudinal wave vector, such that only modes with a fixed longitudinal mode number are accessible to the photon gas at room temperature. This imposes a quadratic dispersion as a function of the transverse wave numbers, and the photon gas becomes formally equivalent to a harmonically trapped two-dimensional gas of massive bosons, which supports Bose-Einstein condensation (25). Photons are injected by pumping with a laser beam. They thermalize to the dye temperature by absorption-reemission cycles before being lost by, for example, mirror transmission (Fig. 1A, right panel). The rhodamine dye fulfills the Kennard-Stepanov relation $B_{\text{em}}/B_{\text{abs}} \propto e^{-h\omega_0/k_B T}$, a Boltzmann-type frequency scaling between the Einstein coefficients for absorption B_{abs} and emission B_{em} . Experimental spectra show agreement with an equilibrium Bose-Einstein distribution within experimental accuracy (Fig. 1B).

The steady-state particle flux from the pump beam through the dye microcavity condensate and out to the environment induces a modified behavior of the particle number fluctuations. In this open system, the sum X of the condensate photon number $n(t)$ and dye molecular excitations $M_e(t)$ is conserved only on average (14, 25), $\bar{X} = \bar{n} + \bar{M}_e$ is constant, where the bar denotes the time average. The dynamics of the corresponding fluctuations Δn and ΔX around the mean can be derived from a Lindblad equation that incorporates the thermally driven fluctuations of the grand canonical system (equivalent to a Langevin equation). For small deviations Δn and ΔX , this leads to a set of equations (25)

$$\frac{d}{dt} \begin{pmatrix} \Delta n \\ \Delta X \end{pmatrix} = \hat{A} \begin{pmatrix} \Delta n \\ \Delta X \end{pmatrix} \quad (1)$$

with the non-Hermitian matrix

$$\hat{A} = \begin{pmatrix} -2\delta & \omega_0^2/\kappa \\ -\kappa & 0 \end{pmatrix} \quad (2)$$

where $\delta = \frac{1}{2}B_{\text{em}}(\bar{M}_e/\bar{n} + \bar{n})$ is the damping rate of the photon number fluctuations, $\omega_0 = \sqrt{\kappa B_{\text{em}}\bar{n}}$ is an oscillation frequency, and the rate constant κ models photon loss. It is instructive to first discuss the expected response to an instantaneous fluctuation at a time t_0 . With the exponential ansatz $(\Delta n_0, \Delta X_0) \cdot e^{\lambda(t-t_0)}$, one obtains solutions characterized by the matrix eigenvalues $\lambda_{1,2} = -\delta \pm \sqrt{\delta^2 - \omega_0^2}$. For a damp-

ing δ below the natural angular frequency ω_0 of the undamped system, the eigenvalues become complex, corresponding to a (damped) oscillatory solution, while in the opposite regime of a large damping ($\delta > \omega_0$) we arrive at

Fig. 2. Expected phase diagram. (A) Illustration of the hierarchy of phases for increasing losses, with fixed condensate size and molecule number. The exceptional point introduces a well-defined phase boundary between photon Bose-Einstein condensates with weakly dissipative (biexponential) and dissipative (oscillatory) correlation dynamics. At loss rates exceeding the thermalization rate, a crossover connects the oscillatory phase to the lasing regime. BEC, Bose-Einstein condensate. (B) Calculated phase boundary ($\delta = \omega_0$) between the two condensate phases (25), as a function of mean condensate and molecule numbers.

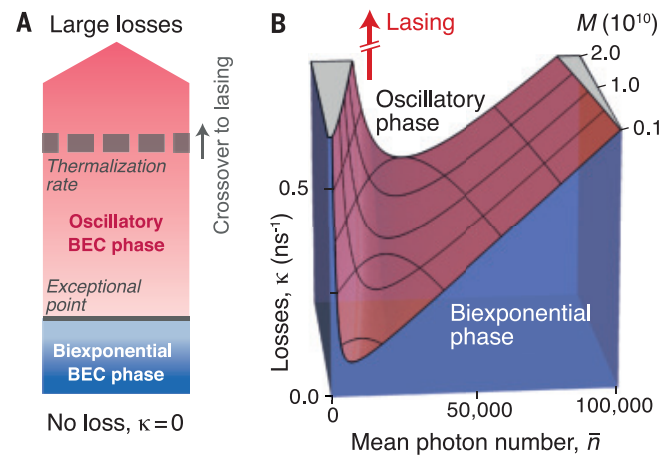
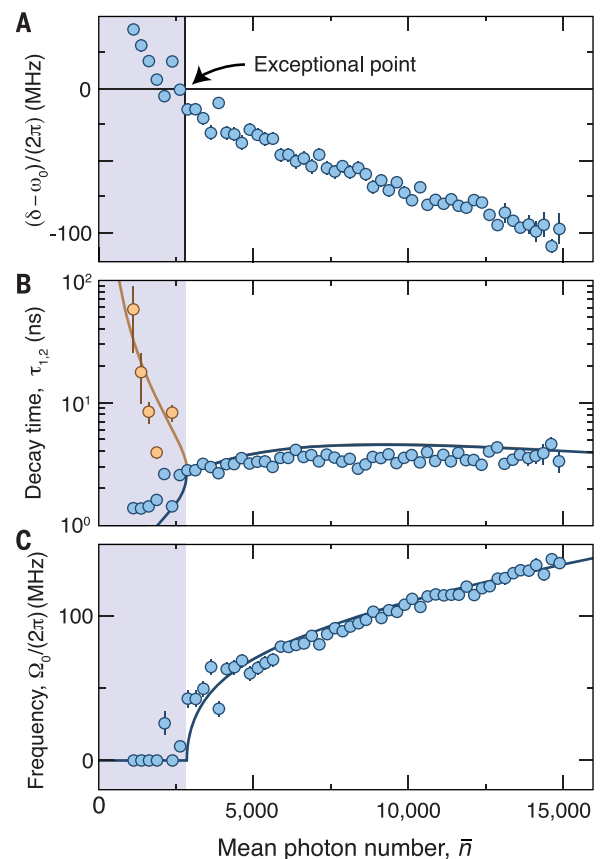


Fig. 3. Non-Hermitian phase transition. (A) Variation of the difference $\delta - \omega_0$, (B) the decay times, and (C) the oscillation frequency with average photon number, determined from the correlation data. Negative (positive) values of $\delta - \omega_0$ (A) indicate an oscillatory (biexponential, blue shading) coherence function. With increasing condensate size, the two decay times (B) merge toward a single value near a photon number $\bar{n}_{\text{EP}} \approx 2800$, which displays the phase transition expected from $\delta - \omega_0 = 0$ in (A). Accordingly, above \bar{n}_{EP} the oscillation frequency becomes non-vanishing. The fits yield $\kappa \approx 2.2(2) \text{ ns}^{-1}$, $M \approx 4.76(3) \times 10^9$. Error bars are calculated from the uncertainties of the fit parameters. (Cutoff wavelength $\lambda_c = 571.3 \text{ nm}$).



real eigenvalues, implying a biexponential decay. At $\delta = \omega_0$, the eigenvalues and the corresponding solutions coalesce, marking an exceptional point. For stationary conditions, i.e., constant pumping and loss, the dynamics of the grand canonical system driven by thermal fluctuations become stochastic; the modes described above can thus only be observed in the number correlations of the

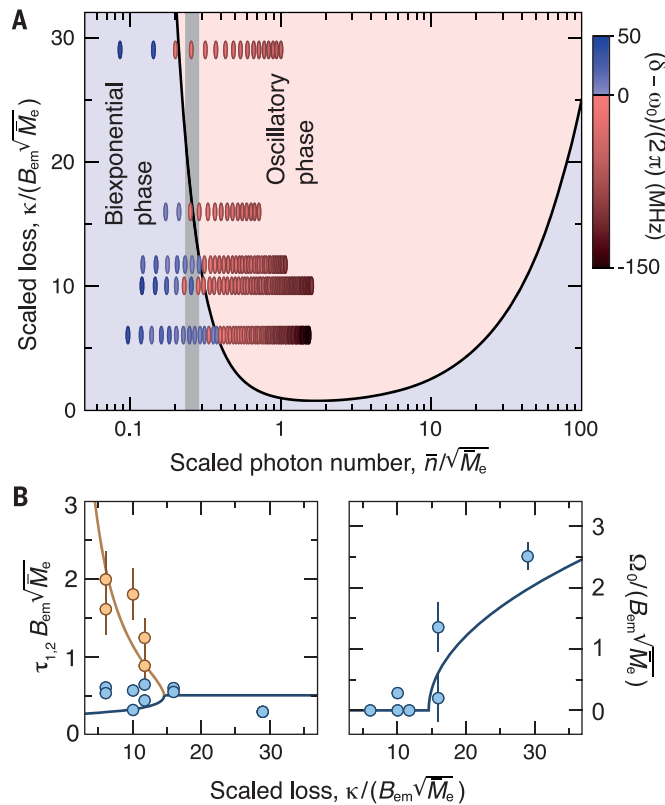
condensate mode, described by $g^{(2)}(\tau) = 1 + e^{-\delta\tau} \left(C_1 e^{-\sqrt{\delta^2 - \omega_0^2}\tau} + C_2 e^{\sqrt{\delta^2 - \omega_0^2}\tau} + \text{c.c.} \right)$, with constants C_1 and C_2 (c.c., complex conjugate) (25).

Tuning between the different regimes—damped oscillatory for $\delta < \omega_0$ and biexponentially decaying correlations for the opposite case—is experimentally achieved by varying the average photon number \bar{n} , which with

Fig. 4. Exploring the phase diagram.

(A) The phase boundary between the two regimes (black line) is mapped out by recording different datasets with different cutoff wavelengths and dye concentrations. From the corresponding coherence functions (as in Fig. 3), we identify a condensate in the biexponential (blue points) or oscillatory (red points) phase, respectively.

(B) Variation of the normalized decay times (left) and oscillation frequency (right) with loss for a scaled photon number $\bar{n}/\sqrt{\bar{M}_e} = 0.27(2)$ [gray shaded area in (A)], showing the transition into the biexponential phase when reducing losses, in good agreement with theory (solid lines). Error bars are calculated from the uncertainties of the fit parameters.



$\omega_0 = \omega_0(\bar{n})$ and $\delta = \delta(\bar{n})$ serves as a control parameter. For small variations of \bar{n} , the photon condensate will remain in the oscillatory or the biexponential regime when being far from the exceptional point ($\delta \ll \omega_0$ or $\delta \gg \omega_0$). At the exceptional point ($\delta = \omega_0$), however, the condensate dynamics, as observed in $g^{(2)}(\tau)$, become very sensitive to changes in \bar{n} and may qualitatively change abruptly. We attribute the exceptional point as marking a non-Hermitian phase transition, separating two dynamical condensate phases. The phase transition mechanism draws analogies with that of quantum phase transitions in closed systems, where typically two energy eigenvalues cross; see also (26) for a proposal of a dissipative phase transition more closely resembling that of (usual) Hermitian systems. For our system, deep in one of the two condensate phases, the eigenvalues of the fluctuation matrix $A(\bar{n})$ are gapped in the complex plane, and we have $\text{Re}(\lambda_1 - \lambda_2) \neq 0$ or $\text{Im}(\lambda_1 - \lambda_2) \neq 0$ on different sides the transition respectively (Re, real; Im, imaginary; see Fig. 1C, bottom panel). The gap closes ($\lambda_1 = \lambda_2$) at the exceptional point. Notably, critical fluctuations known from equilibrium phase transitions are here replaced by an enhanced sensitivity in the correlation dynamics to changes of the control parameter at the phase boundary. No spontaneous symmetry breaking occurs, which is a property shared with, for example, the fermionic Mott-Hubbard

transition. Thermal, reservoir-induced fluctuations of the photon condensate are crucial for the emergence of the described non-Hermitian phase transition.

To experimentally determine the second-order coherence of the photon condensate around the exceptional point, the microcavity emission passes a mode filter to separate the condensate mode from the higher transverse modes. The transmitted light is polarized and directed onto a fast photomultiplier, whose electronic output allows for correlation analysis. Typical obtained traces of the second-order correlations are shown for a cutoff wavelength $\lambda_c \approx 571.3$ nm (Fig. 1C). Whereas for the larger condensate photon number of $\bar{n} \approx 14,000$ the second-order coherence is oscillatory (Fig. 1C, top right panel), for the smaller photon number of $\bar{n} \approx 2100$ it exhibits biexponential behavior (Fig. 1C, top left panel), in good agreement with theory. The difference of damping constant and undamped oscillation frequency, as determined from the fits, is $(\delta - \omega_0)/2\pi = -99(7)$ MHz and $19(2)$ MHz for the two datasets, and consequently the data can be assumed to be in the oscillatory phase for the former and in the biexponential condensate phase for the latter dataset. The presence of both the thermal cloud and the condensate peak in the observed spectra (Fig. 1B), which are a consequence of Bose-Einstein (quantum) statistics,

are attributed as evidence for the quantum many-body character of the phases.

Figure 2A schematically shows the hierarchy of phases for fixed values of the average photon and molecule number, and Fig. 2B gives a three-dimensional plot of the full expected phase diagram. The indicated crossover between lasing and condensation occurs when for $\kappa \gg M_g B_{\text{abs}}$ the loss rate becomes so large that photons leak from the cavity before thermalizing via reabsorption (27, 28). The phase transition between the intermediate oscillatory and the biexponential phases for $\delta = \omega_0$ occurs at $\sqrt{\kappa B_{\text{em}}} \bar{n} \approx \frac{1}{2} B_{\text{em}} (\bar{M}_e / \bar{n} + \bar{n})$ and features a grand canonical ($\bar{M}_e \gg \bar{n}^2$) and a canonical branch ($\bar{M}_e \ll \bar{n}^2$) of the phase boundary, corresponding to the first or second term in the sum being dominant. Here, the first (grand canonical) term, understood to arise from re-trapping of spontaneous emission, is absent in usual laser theory. Experimentally, with $\kappa / (M_g B_{\text{abs}}) \approx 1.1 \times 10^{-3}$ photon thermalization dominates over photon loss, meaning that the exceptional point resides well inside the Bose-Einstein condensed regime.

To explore the phase transition between the two different condensate phases, we have recorded the photon number correlations at different average photon numbers. Figure 3A shows the variation of $\delta - \omega_0$, as determined from the fits of the correlation data. While for condensate sizes above $\bar{n}_{\text{EP}} \approx 2800$, the second-order coherence shows a damped oscillation, for smaller photon numbers, the data exhibits a biexponential decay of the correlations. Figure 3, B and C, shows the obtained decay times and, for the case of the oscillatory phase, the oscillation frequency depending on the condensate size. Both datasets give evidence for the photon condensate undergoing a non-Hermitian phase transition to the biexponential phase at a critical condensate occupation \bar{n}_{EP} . The deviation of the observed decay times from the prediction for short times is attributed to the 500-ps resolution of the detection system. Notably, when approaching the phase transition from below, the two characteristic decay times merge toward a single one, and when approaching the transition from above, the oscillation frequency converges to zero. This is in good agreement with the expectation that at the exceptional point, a (single) exponential decay of the second-order coherence occurs owing to the coalescence of the two eigenvalues, $\lambda_1 = \lambda_2 = \delta$. The revealed phase transition is visible in the temporal correlations but not in the average values.

Next, we recorded data at different cavity low-frequency cutoffs and dye concentrations, to explore the phase diagram beyond a single control parameter. The resulting change of the wavelength of condensate photons modifies both the loss as well as the Einstein coefficient. Because of the shape of the phase boundary at

$\delta = \omega_0$, upon rescaling the photon number as $\bar{n}/\sqrt{M_e}$ and the loss rate as $\kappa/B_{em}\sqrt{M_e}$, the phase diagram in Fig. 2B collapses to a two-dimensional one (25). Corresponding data are summarized in Fig. 4A. To obtain κ and the molecule number M_e , curves similar to those of a single dataset shown in Fig. 3 were fitted to all data. Our experimental data maps out the non-Hermitian phase transition between the oscillatory and biexponential phase within the investigated parameter range, in good agreement with expectations (black line). The variation of the normalized decay times and oscillation frequency versus the scaled loss rate in Fig. 4B for a fixed value of $n/\sqrt{M_e} \approx 0.27$ demonstrate the branching of the eigenvalues when reducing the loss toward the idealized case of a perfect photon box.

The state of a macroscopic quantum system on different sides of an exceptional point can be in two distinct regimes. We have observed the associated dissipative phase transition from an oscillatory to a biexponential dynamical phase of a dye microcavity photon Bose-Einstein condensate and mapped out the corresponding phase diagram. This reveals a state of the light field, which, contrary to the usual picture of Bose-Einstein condensation, is separated by a phase transition from the phenomenon of lasing.

REFERENCES AND NOTES

- M. A. Miri, A. Alù, *Science* **363**, eaar7709 (2019).
- T. Fink, A. Schade, S. Höfling, C. Schneider, A. Imamoglu, *Nat. Phys.* **14**, 365–369 (2017).
- G. Makey et al., *Nat. Phys.* **16**, 795–801 (2020).
- N. Dogra et al., *Science* **366**, 1496–1499 (2019).
- I. Prigogine, *Science* **201**, 777–785 (1978).
- F. E. Öztürk et al., *Phys. Rev. A* **100**, 043803 (2019).
- S. Diehl et al., *Nat. Phys.* **4**, 878–883 (2008).
- R. Hanai, A. Edelman, Y. Ohashi, P. B. Littlewood, *Phys. Rev. Lett.* **122**, 185301 (2019).
- K. Kawabata, K. Shiozaki, M. Ueda, M. Sato, *Phys. Rev. X* **9**, 041015 (2019).
- E. Altman, L. M. Sieberer, L. Chen, S. Diehl, J. Toner, *Phys. Rev. X* **5**, 011017 (2015).
- J. Klaers, J. Schmitt, F. Vewinger, M. Weitz, *Nature* **468**, 545–548 (2010).
- Y. Sun et al., *Phys. Rev. Lett.* **118**, 016602 (2017).
- I. Carusotto, C. Ciuti, *Rev. Mod. Phys.* **85**, 299–366 (2013).
- J. Schmitt et al., *Phys. Rev. Lett.* **112**, 030401 (2014).
- A. Griffin, D. W. Snoke, S. Stringari, Eds., *Bose-Einstein Condensation* (Cambridge Univ. Press, 1995).
- B. T. Walker et al., *Nat. Commun.* **11**, 1390 (2020).
- T. Kato, *Perturbation Theory of Linear Operators* (Springer, 1966).
- C. M. Bender, S. Boettcher, *Phys. Rev. Lett.* **80**, 5243–5246 (1998).
- T. Gao et al., *Nature* **526**, 554–558 (2015).
- F. Minganti, A. Miranowicz, R. W. Chhajlany, F. Nori, *Phys. Rev. A* **100**, 062131 (2019).
- W. D. Heiss, *J. Phys. A* **45**, 444016 (2012).
- R. Hanai, P. B. Littlewood, *Phys. Rev. Res.* **2**, 033018 (2020).
- J. Marelic, R. A. Nyman, *Phys. Rev. A* **91**, 033813 (2015).
- S. Greveling, K. L. Perrier, D. van Oosten, *Phys. Rev. A* **98**, 013810 (2018).
- See supplementary materials.
- E. M. Kessler et al., *Phys. Rev. A* **86**, 012116 (2012).
- P. Kirtan, J. Keeling, *Phys. Rev. Lett.* **111**, 100404 (2013).
- J. Schmitt et al., *Phys. Rev. A* **92**, 011602 (2015).
- F. E. Öztürk et al., Observation of a non-Hermitian phase transition in an optical quantum gas, Version 1, Zenodo (2020); <http://doi.org/10.5281/zenodo.4522437>.

ACKNOWLEDGMENTS

We thank S. Diehl, M. Scully, and H. Stöf for discussions. **Funding:** We acknowledge support by the DFG, under SFB/TR 185 (277625399) and the Cluster of Excellence ML4Q (EXC 2004/1–390534769); the EU, under the Quantum Flagship project PhoQuS; and the DLR, with funds provided by the BMWi (50WM1859). J.S. thanks the University of Cambridge for support during the early stages of this work, and M.W. thanks the CAIQE for providing a guest stay at UC Berkeley. **Author contributions:** F.E.Ö., T.L., J.S., and F.V. analyzed the data. J.S., J.K.L., and M.W. conceived of and designed the experiments. F.E.Ö., T.L., and J.Kr. contributed materials and/or analysis tools. F.E.Ö. and G.H. performed the experiments.

F.E.Ö., T.L., J.S., F.V., J.Kr., and M.W. wrote the paper.

Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** Data shown in the figures are available in the Zenodo database (29).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/88/suppl/DC1
Materials and Methods
Supplementary Text
Table S1
References (30–41)

26 September 2020; accepted 23 February 2021
10.1126/science.abe9869

NEUROGENOMICS

Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons

Dylan A. Reid^{1,†}, Patrick J. Reed^{1,†}, Johannes C. M. Schlachetzki^{2,†}, Ioana I. Nitulescu^{1,†}, Grace Chou^{3,†}, Enoch C. Tsui^{1,†}, Jeffrey R. Jones^{1,†}, Sahaana Chandran⁴, Ake T. Lu⁵, Claire A. McClain¹, Jean H. Ooi¹, Tzu-Wen Wang³, Addison J. Lana², Sara B. Linker¹, Anthony S. Ricciardulli¹, Shong Lau¹, Simon T. Schafer¹, Steve Horvath^{5,6}, Jesse R. Dixon⁴, Nasun Hah³, Christopher K. Glass², Fred H. Gage^{1,*}

Neurons are the longest-lived cells in our bodies and lack DNA replication, which makes them reliant on a limited repertoire of DNA repair mechanisms to maintain genome fidelity. These repair mechanisms decline with age, but we have limited knowledge of how genome instability emerges and what strategies neurons and other long-lived cells may have evolved to protect their genomes over the human life span. A targeted sequencing approach in human embryonic stem cell–induced neurons shows that, in neurons, DNA repair is enriched at well-defined hotspots that protect essential genes. These hotspots are enriched with histone H2A isoforms and RNA binding proteins and are associated with evolutionarily conserved elements of the human genome. These findings provide a basis for understanding genome integrity as it relates to aging and disease in the nervous system.

Neurons are highly specialized postmitotic cells and comprise the major functional cell type of the nervous system. Although there is a limited capacity to generate new neurons throughout life, the majority of neurons age in parallel with the organism, making them especially susceptible to decline from age-related disruptions in cellular homeostasis (1). In total, neurons repair on the order of $\sim 10^4$ to 10^5 DNA lesions each day, or more than 1 billion repairs over the life span

of humans (2). Deficiencies in DNA repair and genome instability have been linked to both developmental and age-associated neurodegenerative diseases (3, 4).

Studies of genome integrity in neurons suggest that their DNA repair efforts focus on transcribed genes at the expense of inactive regions of the genome (5). Accumulation of DNA lesions drives age-associated changes in transcription that lead to a decline in neuronal function (6, 7). Additionally, neuronal activity correlates with the generation of DNA double-strand breaks (DSBs), potentially contributing to genome instability (8, 9). Although genomics approaches have made the study of mitotic neural progenitor cells and the role of somatic mosaicism in neurons accessible, methods to detect DNA damage remain technically challenging, limiting their use (10, 11). Despite the link between genome maintenance and neuronal health, we know surprisingly little about the genome protection strategies with which neurons have evolved to ensure their distinctive longevity.

To better understand genome integrity in neurons, we developed a sequencing method capable of capturing a genomic distribution of

¹Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037-1002, USA. ²Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92037-0651, USA. ³Next Generation Sequencing Core, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037-1002, USA. ⁴Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037-1002, USA. ⁵Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA. ⁶Department of Biostatistics, School of Public Health, University of California, Los Angeles, Los Angeles, CA 90095, USA.

*Corresponding author. Email: dylreid@gmail.com (D.A.R.); gage@salk.edu (F.H.G.). †These authors contributed equally to this work. ‡These authors contributed equally to this work.

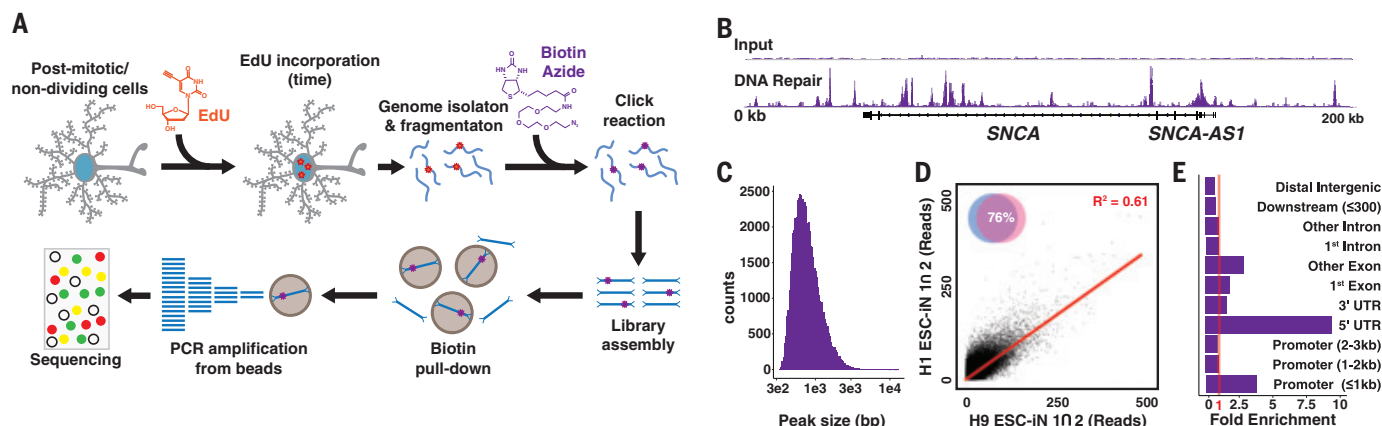


Fig. 1. EdU incorporated into the genomes of postmitotic neurons by DNA repair can be mapped by next-generation sequencing. (A) Repair-seq workflow: Neurons are cultured with EdU for 24 hours, the genomes are isolated and fragmented with sonication, a click reaction adds biotin to the EdU, and biotin-DNA fragments are enriched on streptavidin beads and subsequently amplified (12). PCR, polymerase chain reaction. (B) DNA

repair peaks from the *SNCA* locus in EdU-fed neurons compared with input genomes sequenced to the same depth show a site with substantial enrichment. (C) Histogram of DNA repair hotspot peak widths. bp, base pairs. (D) Comparison of intersecting peaks in two H1 and two H9 ESC-iN samples. (E) Fold enrichment of DRHs over predicted genome distribution. UTR, untranslated region.

all DNA repair by the nonreplicative incorporation of the nucleoside analog 5-ethynyl-2'-deoxyuridine (EdU). We generated human embryonic stem cell-induced neurons (ESC-iNs) that assume a postmitotic neuron identity after the addition of doxycycline through *NEUROG2* expression (fig. S1) (12). ESC-iNs were labeled with EdU for 24 hours, and sites of DNA repair synthesis were identified by the enrichment of next-generation sequencing libraries containing EdU (Fig. 1A) (12). Our method, Repair-seq, revealed many sites enriched for EdU incorporation relative to whole-genome sequencing to the same depth and was relatively free of mitochondrial reads (Fig. 1B and figs. S2 and S3, A and B). EdU-enriched sites appeared as well-defined peaks of ~500 base pairs (Fig. 1C). We applied genome peak calling to our data and found 61,178 reproducible peaks, or DNA repair hotspots (DRHs), covering ~1.6% of the genome (Fig. 1D; fig. S3, C to E; and table S1). These DRHs were distributed throughout the genome on all chromosomes and were enriched in promoters of ≤1 kb, 5'-untranslated regions, and gene bodies (Fig. 1E and fig. S3, F to H).

We compared the location of DRHs with open chromatin and active regulatory regions in neurons mapped by ATAC-seq (assay for transposase-accessible chromatin using sequencing) and histone 3 lysine 27 acetylation (H3K27Ac) ChIP-seq (chromatin immunoprecipitation sequencing), and we observed that ~23.5% of hotspots were located within these genomic regions (Fig. 2A and fig. S4, A to C). Intersecting peaks in open regions correlated with greater DNA repair signal strength (Fig. 2, B and C, and fig. S4D). Promoters were enriched for repair, ATAC, and H3K27Ac peak intersections, whereas DRHs not associated

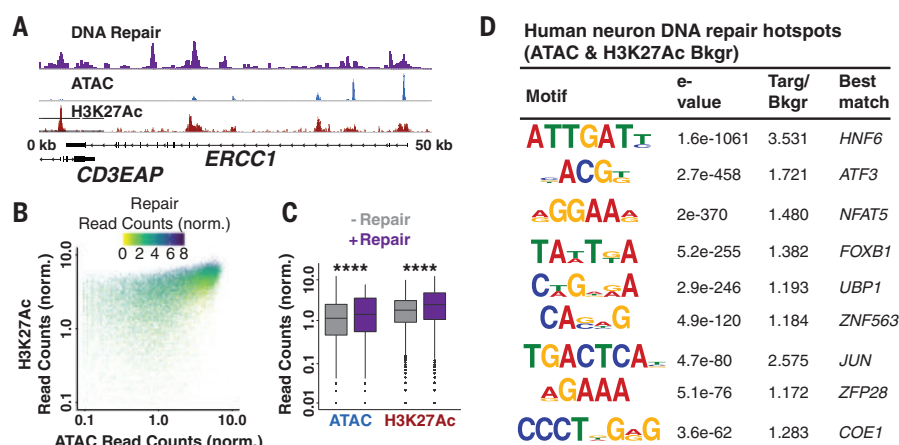


Fig. 2. Chromatin accessibility controls the placement of repair hotspots. (A) Repair-seq, ATAC-seq, and H3K27Ac ChIP-seq data at the *ERCC1* locus demonstrate overlap between DNA repair, chromatin accessibility, and histone acetylation. (B) Scatter plot of Repair-seq-normalized read counts compared to ATAC- and H3K27Ac-normalized read counts. (C) Box plots of ATAC and H3K27Ac peaks with and without DNA repair. The horizontal black lines represent the medians, whereas the whiskers are displayed at the largest and smallest values no more than 1.5 times the interquartile range from quartiles 3 and 1, respectively. **** $P < 2.2 \times 10^{-16}$ by Kruskal-Wallis test. (D) DNA sequence motifs identified de novo and predicted as enriched in DRHs relative to randomized sequence. Targ/Bkgr, target/background.

with open chromatin were predominantly located in intergenic and intronic elements of the genome (fig. S4E). De novo DNA sequence motif analysis identified significantly enriched sequences in DRHs when considering sequence bias and ATAC or H3K27Ac peaks as background to correct for the contributions of open chromatin (Fig. 2D, fig. S5, and table S2).

Repair-seq allowed us to compare all DNA repair- and transcription-associated reads. Most Repair-seq reads (~67%) could be assigned to genes, with the majority of the neuronal transcriptome exhibiting some level of maintenance

that increased with expression levels (Fig. 3A and fig. S6, A and B). This finding corroborates prior work suggesting that in neurons, global DNA repair is attenuated and consolidated to actively transcribed genes, presumably to suppress the accumulation of lesions and mutations (5). However, when we examined DRH reads (~23% of all Repair-seq reads), we observed that many genes lacked recurrent DNA repair sites and showed no relationship with expression (Fig. 3B, fig. S6C, and table S3). Comparison of the locations of DRHs with transcribing RNA polymerases [global run-on sequencing (GRO-seq)] showed strong promoter

Fig. 3. Transcriptional output correlates with total DNA repair in genes but not repair hotspots. (A and B) All DNA repair-associated reads in genes (A) and repair peak-associated reads in genes (B) compared with RNA-associated reads from total RNA-seq. TPM, transcripts per kilobase million. (C and D) All DNA repair-associated reads in genes (C) and repair peak-associated reads in genes (D) compared with RNA-associated reads from total RNA-seq in length-normalized TADs.

(E) Select biological process (BP) GO terms for genes containing DRHs. Terms are neuron projection development, neuron development, neuron projection morphogenesis, genesis of neurons, neuron differentiation, axon development, neurogenesis, and nervous system development. (F) Line plot of transcription start sites (TSSs) to DRHs in each gene compared with total gene length (colored by total DNA repair level). (G) String network representation of peptides enriched for histones (green), RNA binding proteins (RBPs; blue), and some chaperones and ubiquitin (purple). (H) LFQ proteomics data for H2AX and NONO abundance in cognitively normal (CN), asymptomatic Alzheimer's disease (AsymtAD), and Alzheimer's disease (AD). Horizontal black lines represent mean $\log_2(\text{LFQ})$. ns, not significant; * $P < 0.05$; ** $P < 0.01$; and **** $P < 0.0001$ by analysis of variance (ANOVA) with Tukey's multiple comparison test.

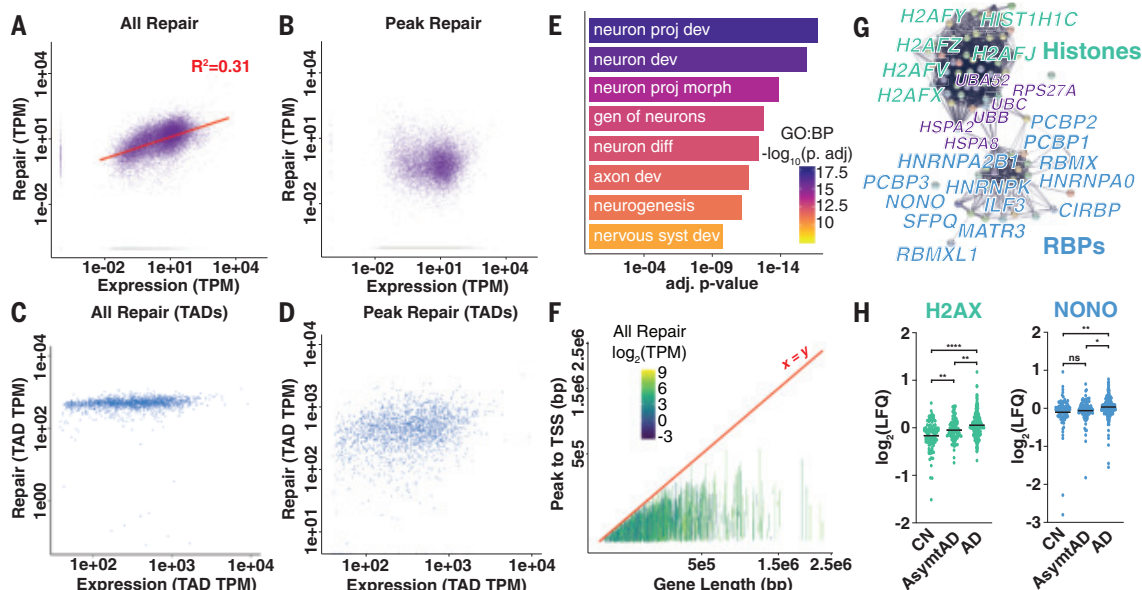
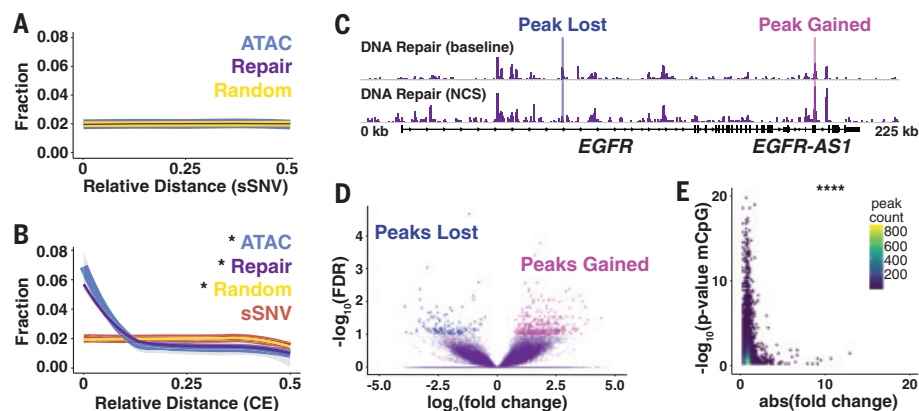


Fig. 4. Repair hotspots protect evolutionarily constrained regions of the human genome from epigenetic drift. (A) Relative distance measurement from postmortem human neuron sSNVs to nearest DRH or randomly generated peaks. (B) Relative distance measurement from GERP CEs to nearest sSNV, DRH, ATAC, or random peaks. (C) Representative browser view of DRHs at baseline and 24 hours after 10 min of NCS treatment demonstrates that peaks are lost and gained. (D) Volcano plot for NCS differential peaks using a false discovery rate (FDR) of <0.1 from four samples. (E) Heat map of the DRH stability (absolute fold change after NCS treatment) compared with epigenetic clock mCpG sites from sorted human neurons. * $P < 0.01$ by Jaccard distance test; **** $P < 8.52 \times 10^{-5}$ by hypergeometric test.

(A) Relative distance measurement from postmortem human neuron sSNVs to nearest DRH or randomly generated peaks. (B) Relative distance measurement from GERP CEs to nearest sSNV, DRH, ATAC, or random peaks. (C) Representative browser view of DRHs at baseline and 24 hours after 10 min of NCS treatment demonstrates that peaks are lost and gained. (D) Volcano plot for NCS differential peaks using a false discovery rate (FDR) of <0.1 from four samples. (E) Heat map of the DRH stability (absolute fold change after NCS treatment) compared with epigenetic clock mCpG sites from sorted human neurons. * $P < 0.01$ by Jaccard distance test; **** $P < 8.52 \times 10^{-5}$ by hypergeometric test.



enrichment (fig. S7). Almost one-third of DRHs were located in intergenic regions and could not be assigned to transcription of single genes.

To address the potential contribution of these sites to transcription-associated repair, we generated Hi-C contact maps for ESC-iNs to assign intergenic peaks to genes using features of three-dimensional (3D) genome organization, such as topologically associating domains (TADs) (13). DNA repair levels in most TADs were uniform (Fig. 3C). Assignment of intergenic peaks did not substantially alter the finding that DRHs were not correlated with the level of gene transcription (Fig. 3D and fig. S8). A comparison of the distribution

of either all DNA repair-associated reads or Repair-seq peaks with genome-wide features of 3D genome organization, such as A/B compartments, displayed an enrichment of DNA repair in the “active” A compartments (fig. S9).

We found that DRH genes were enriched for specific cellular processes irrespective of expression level, because they were correlated with genes essential for neuronal identity and function (Fig. 3E, fig. S10, and table S4). We explored whether gene length played a role in DRH density and found that both total repair and transcription were independent of gene length (fig. S11, A and B). However, when we examined reads that were only from DRHs in

relationship to length, the total level of repair in these sites, as well as total peak density, paradoxically diminished in relationship with gene length (Fig. 3F and fig. S11, C and D). These findings suggest that DRHs in neuronal genes might arise from the requirements of maintaining transcriptional elongation and splicing in genes containing large introns (14).

To investigate if DRHs were linked to splicing in neurons, we performed rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) on chromatin that had undergone repair (12), and we detected 79 enriched proteins (table S5). Proteins identified by RIME were largely grouped into histone H2A isoforms and RNA binding proteins by network

and gene ontology (GO) analysis (Fig. 3G, fig. S12, and table S6). The presence of the marker histone H2AZ was validated by ChIP-seq (fig. S13). We used the Consensus Brain Protein Coexpression Study dataset to compare protein abundance [label-free quantitation (LFQ)] in cognitive normal and asymptomatic or symptomatic Alzheimer's disease patients (15). We found that 21 of the identified proteins showed differences in neurodegenerative disease ($P < 2.67 \times 10^{-10}$ by hypergeometric test), suggesting a role for changes to DNA repair in the etiology and progression of Alzheimer's disease (Fig. 3H and table S7) (16).

In mice, neuronal activity generates DSBs and the damage marker γ H2AX in select genes to initiate transcription for learning and memory (8, 9). Repair-seq was used on KCl-stimulated ESC-iNs to find activity-induced break sites in human neurons (12). No substantial changes were observed in DRHs after neuron depolarization, in contrast to cells where spontaneous activity was inhibited with tetrodotoxin (TTX) (fig. S14, A to D). Genes linked to activity-induced DSBs in mice showed minimal changes in DNA repair levels with either neuronal stimulation or inhibition (fig. S14E). Activity-induced breaks were linked to topoisomerase II β at CCCTC-binding factor (CTCF) sites; however, we found minimal intersection between CTCF ChIP-seq and Repair-seq peaks (fig. S14, F and G). This lack of increased DNA repair linked to neuronal stimulation suggests species-specific differences in how these genes are transcribed (17), that their repair might be highly reliable and not incorporate new nucleotides, or that the γ H2AX that is associated with activity may not be a reliable marker of DSBs (18). However, given the ability of TTX to suppress many DRHs, we believe that a substantial fraction of DNA repair is linked to neuronal identity established by activity. Finally, we noted that *FOS* and *NPAS4* contained predicted G-quadruplex structures, and we performed an analysis that suggested that these could be key regulatory features of neuron promoters that might be vulnerable to damage (fig. S15 and table S8) (19).

As cells age, the activity of DNA repair mechanisms declines, leading to an increase in somatic mutations and the accumulation of unrepaired lesions (20). Direct intersection and relative distance comparison between DRHs and somatic single-nucleotide variants (sSNVs) identified from single neurons isolated from postmortem human brains showed no proximal enrichment (Fig. 4A and fig. S16, A to C) (17), suggesting that mutations occurred randomly throughout the genome, irrespective of DRHs. We next used genomic evolutionary rate profiling (GERP)-defined constrained elements (CEs) in humans and compared the maximum GERP score and CE location with DRHs; the DRHs were enriched near CEs and

more likely to have a single base under strong conservation, in contrast to sSNV sites (Fig. 4B; fig. S16, D to F; and fig. S17). These data suggest that DRHs might protect essential elements from both erroneous repairs and going unrepaired.

Aging drives fundamental changes in the epigenome (21), and biological age can be quantified with epigenetic clock models created using changes in the methylation patterns on CpG dinucleotides (22). Despite the accuracy of such models, no satisfying biological explanation exists as to why these DNA modifications are linked to aging (22). We compared the locations and proximity of DRHs with CpG sites and methylated CpG dinucleotides (mCpG) statistically associated with aging in neurons from human prefrontal cortex (23), and we found that they were closely associated (fig. S18). Genome instability in the form of DSBs is a primary driver of biological aging (24). We treated ESC-iNs with the DNA-damaging agent neocarzinostatin (NCS) to assay injury-induced changes to DRHs. Acute NCS treatment triggered both the gain and loss of DRHs in neurons in a stochastic fashion, although at the dosage used, relatively few peaks were detected (Fig. 4, C and D, and table S9). In the context of aging, genome instability could potentially redistribute repair efforts away from hotspots to other locations in the genome, similar to NCS treatment. A comparison of absolute fold change for NCS and other DNA damage-treated samples with statistically significant mCpG sites indicated that the most stable DRHs were those associated with the epigenetic clock and CEs (Fig. 4E, figs. S19 and S20, and tables S10 to S15). Therefore, as DNA repair capacity declines with age and pathways become overtaxed, these sites could be susceptible to dysregulation.

Our results suggest that DRHs are established in neurons and play a key role in identity and function. Going forward, Repair-seq will be a powerful tool to explore how age and disease disrupt genome integrity in the nervous system. Finally, whether DRHs are specific to neurons, particular developmental lineages, or other nondividing cells or are found in only some long-lived species remains an open question. The discovery of these sites in other cell types might further aid in our understanding of how age-related changes in their organization could drive differential aging or the development of disease in other tissue types.

Note added in proof: It was brought to our attention that a closely related paper by Wu *et al.* (25) is in press.

REFERENCES AND NOTES

- M. P. Mattson, T. Magnus, *Nat. Rev. Neurosci.* **7**, 278–294 (2006).
- S. P. Jackson, J. Bartek, *Nature* **461**, 1071–1078 (2009).
- H. M. Chow, K. Herrup, *Nat. Rev. Neurosci.* **16**, 672–684 (2015).
- P. J. McKinnon, *Nat. Neurosci.* **16**, 1523–1529 (2013).

- T. Nospikel, P. C. Hanawalt, *Mol. Cell. Biol.* **20**, 1562–1570 (2000).
- T. Lu *et al.*, *Nature* **429**, 883–891 (2004).
- W. P. Vermeij *et al.*, *Nature* **537**, 427–431 (2016).
- E. Suberbielle *et al.*, *Nat. Neurosci.* **16**, 613–621 (2013).
- R. Madabhushi *et al.*, *Cell* **161**, 1592–1605 (2015).
- P. C. Wei *et al.*, *Cell* **164**, 644–655 (2016).
- M. A. Lodato *et al.*, *Science* **359**, 555–559 (2018).
- Materials and methods are available as supplementary materials.
- J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
- A. Takeuchi *et al.*, *Cell Rep.* **23**, 1326–1341 (2018).
- E. C. B. Johnson *et al.*, *Nat. Med.* **26**, 769–780 (2020).
- V. Swarup *et al.*, *Cell Rep.* **31**, 107807 (2020).
- P. Pruunsild, C. P. Bengtson, H. Bading, *Cell Rep.* **18**, 122–135 (2017).
- N. M. Shanbhag *et al.*, *Acta Neuropathol. Commun.* **7**, 77 (2019).
- R. Hånsel-Hertsch *et al.*, *Nat. Genet.* **48**, 1267–1272 (2016).
- S. Maynard, E. F. Fang, M. Scheibye-Knudsen, D. L. Croteau, V. A. Bohr, *Cold Spring Harb. Perspect. Med.* **5**, a025130 (2015).
- C. López-Otin, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, *Cell* **153**, 1194–1217 (2013).
- S. Horvath, K. Raj, *Nat. Rev. Genet.* **19**, 371–384 (2018).
- A. Kozlenkov *et al.*, *Genes (Basel)* **8**, 152 (2017).
- R. R. White *et al.*, *Nat. Commun.* **6**, 6790 (2015).
- W. Wu *et al.*, *bioRxiv* 2020.12.16.423085 [Preprint].
- 16 December 2020; <https://doi.org/10.1101/2020.12.16.423085>.
- P. J. Reed, Code used in the analysis of Repair-seq data. Synapse (2021); <http://doi.org/10.7303/syn24202004>.

ACKNOWLEDGMENTS

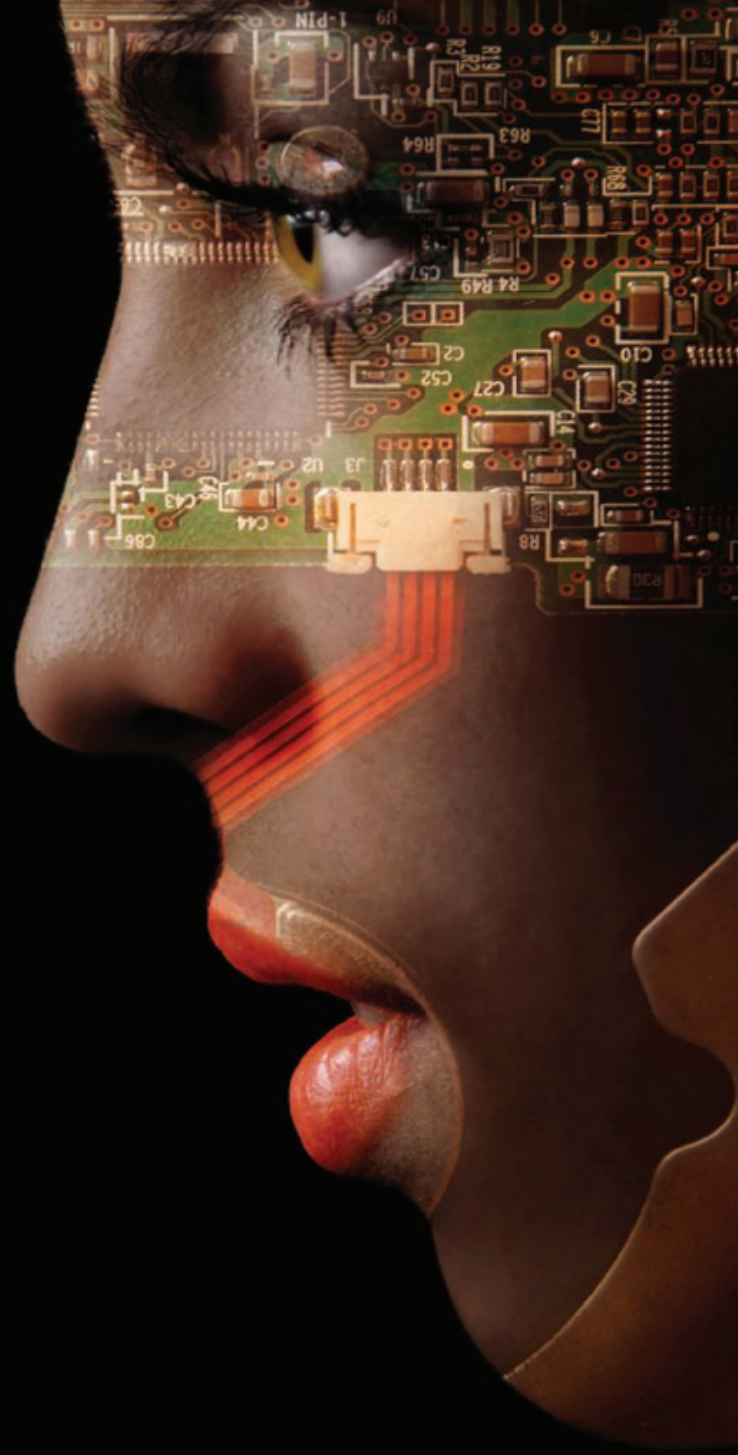
We thank L. Moore, I. Guimont, K. E. Diffenderfer, W. Travis Berggren, J. Diedrich, and A. Pinto for technical assistance as well as M. L. Gage for editorial comments. We also acknowledge the Salk Institute Stem Cell Core, Next Generation Sequencing Core, and Mass Spectrometry Core for technical support. **Funding:** D.A.R. is an Alzheimer's Association Research Fellow (AARF-17-504089). This work was supported by an American Heart Association (AHA)-Allen Initiative in Brain Health and Cognitive Impairment award made jointly through the AHA and The Paul G. Allen Frontiers Group (19PABH134610000); the JPB Foundation; the Dolby Foundation; the Helmsley Charitable Trust; NIH AG056306 to F.H.G.; NIH R01AG056511-02 to C.K.G.; NIH DP5OD023071-03 to J.R.D.; and P30 014195 to the Mass Spectrometry Core. **Author contributions:** D.A.R. conceived of the project, generated the data, helped analyzed the results, and supervised the project in coordination with N.H., C.K.G., and F.H.G. Repair- and RNA-seq libraries were generated by D.A.R., G.C., C.A.M., J.H.O., and T.-W.W. ATAC- and ChIP-seq experiments were performed by J.C.M.S. and A.J.L. Hi-C experiments were performed by S.C. RIME experiments were performed by I.I.N., D.A.R., and E.C.T. GRO-seq was performed by N.H. Additional experiments and reagents were contributed to the study by J.R.J., A.S.R., E.C.T., S.L., and S.T.S. Analysis of repair-, ATAC-, ChIP-, and RNA-seq was performed by P.J.R. and S.B.L. Analysis of Hi-C was performed by P.J.R. and J.R.D. DNA methylation aging analysis was performed by A.T.L. and S.H. Analysis of RIME data was performed by I.I.N. The manuscript was written by D.A.R., P.J.R., and F.H.G. and edited by J.C.M.S., J.R.J., J.R.D., I.I.N., and C.K.G. **Competing interests:** The Salk Institute for Biological Studies has filed a patent governing the use of Repair-seq for the detection of DNA repair in nondividing cells. **Data and materials availability:** Primary data are available at www.ncbi.nlm.nih.gov/bioproject/PRJNA699309, and code is available at www.synapse.org/repairseq (26).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6537/91/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S15
References (27–59)
MDAR Reproducibility Checklist

25 March 2020; resubmitted 13 April 2020
Accepted 18 February 2021
10.1126/science.abb9032


Share Your Robotics Research with the World.



Shaping the future of robotics with high impact research!

As a multidisciplinary online-only journal, *Science Robotics* publishes original, peer-reviewed, research articles that advance the field of robotics. The journal provides a central forum for communication of new ideas, general principles, and original developments in research and applications of robotics for all environments.

Submit your research today. Learn more at: **robotics.sciencemag.org**

 Twitter: @SciRobotics

 Facebook: @ScienceRobotics

Science Robotics

 AAAS

ScienceCareers

FROM THE JOURNAL SCIENCE  AAAS

Confused about your next career move?



**Download Free Career
Advice Booklets!**

ScienceCareers.org/booklets





**SOMETIMES
THE GRASS
REALLY
IS GREENER
SOMEPLACE
NEW.**

 **Find your next job at [ScienceCareers.org](https://www.sciencecareers.org)**

ScienceCareers
FROM THE JOURNAL SCIENCE 

The National
Academies of

SCIENCES
ENGINEERING
MEDICINE



ARL DISTINGUISHED POSTDOCTORAL FELLOWSHIPS

The Army Research Laboratory (ARL) Distinguished Postdoctoral Fellowships provide opportunities to pursue independent research in ARL laboratories. Fellows benefit by working alongside some of the nation's best scientists and engineers, while enhancing the mission and capabilities of the U.S. Army and the warfighter in times of both peace and war.

Fellows must display extraordinary abilities in scientific research and show clear promise of becoming future leaders. Candidates are expected to have already successfully tackled a major scientific or engineering problem or to have provided a new approach or insight, evidenced by a recognized impact in their field.

Fellowships are one-year appointments, renewable for up to three based on performance. The award includes a \$100,000 annual stipend, health insurance, paid relocation, and a professional travel allowance. Applicants must have completed all requirements for a Ph.D. or Sc.D. degree by **October 1, 2021**, and may not be more than five years beyond their doctoral degree as of the application deadline. For more information and to apply visit www.nas.edu/ar1.

Online applications must be submitted by **June 15, 2021 at 5 PM Eastern Time**.



**Who's the top
employer for 2020?**

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

**Science 2020
TOP EMPLOYER**

Read the article and employer profiles at [sciencecareers.org/topemployers](https://www.sciencecareers.org/topemployers)

By Phil De Luna

Choosing from the heart

After I returned from a monthslong research trip, my partner sat me down to talk. We'd moved in together shortly before my departure and, in my absence, she'd been pondering our future. "I don't want to leave Toronto. I love it here and so do you," she said. I was in the third year of my Ph.D., and she worried my career plans would lead us elsewhere, first for a postdoc and then for a faculty position. Deep down, I knew this conversation was coming. Until then I hadn't given the issue much thought, but I knew she was right. I wanted to be with her more than I wanted to be a professor.

My partner's career as an operating room nurse was mobile in theory. But she loved the hospital where she worked, loved Toronto (the city where she grew up), and wanted to stick close to friends and family. She wanted to build our life where we were, to make the city our home.

During my time in grad school, I had grown to love Toronto as well. The energy in the streets was exhilarating, and as a Filipino immigrant to Canada, I enjoyed the cultural diversity. But I wasn't aware of jobs that would allow me to stay without compromising the academic career I had been working toward for my entire life.

My partner's concerns about our future forced me to focus on my career plans and to reassess the academic career path I'd been blindly following up to that point. And the more I thought about it, the more I realized that path might not be right for me after all. There were aspects of academia I didn't like, such as the power disparity between students and professors and the pervasive view that only academic success was real success. I also realized many careers outside academia—in industry and policy, for instance—offered a chance to make a more rapid impact in the real world.

I started to explore options by contacting Ph.D. graduates who had followed career paths I found interesting, asking for informational interviews. I spoke with venture capitalists, energy company executives, management consultants, former politicians, and startup founders. I'd always end the phone call with the same question: "If you could talk to yourself when you were in your mid-20s, knowing everything you know now, what advice would you give?" Often, the answer was that career paths are not straight or neatly assembled. It's important to be willing to fail and pivot to



"The conversations opened my eyes to a universe of career options."

something new, as doing so often leads you to better places.

The conversations opened my eyes to a universe of career options and showed me that no one path interested me most; rather, I wanted to experience them all! From then on, I started to envision my career as a set of chapters: Perhaps I'd start with one option, then move on to another at some point in the future. That perspective freed me to explore many possibilities without worrying that diving into one meant giving up a chance to pursue another.

When I was in my last year of graduate school, I looked into starting a company to build the clean-energy technology I'd been researching during my Ph.D. But my adviser felt the technology

was too early for commercialization, so I started to look for jobs. One day on LinkedIn, I noticed a job posting for a government position. A national lab wanted someone to lead the development of technologies to help Canada reach net zero carbon emissions. I hadn't been seeking government jobs, but I was excited. The work had the potential to make a real impact—and it was based in Toronto. I applied, was offered the job, and accepted.

Two years later, I can say with confidence that my career transition was right for me. I might not stay in this job forever, but it feels like the perfect first step for me to take postgraduation. I love what I do—and I get to wake up every morning with a smile on my face next to the person I love most. I wouldn't trade that for anything. ■



Phil De Luna is a director at a national lab in Toronto, Canada. Do you have an interesting career story that you would like to share? Send it to SciCareerEditor@aaas.org.

READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

ScienceImmunology



 Twitter: @SciImmunology
 Facebook: @ScienceImmunology



EMPOWER WITH EVIDENCE

 **AAAS** | ANNUAL MEETING

SUBMIT A SCIENTIFIC SESSION OR WORKSHOP PROPOSAL FOR THE 2022 AAAS ANNUAL MEETING!

Theme of the 2022 AAAS Annual Meeting—Empower With Evidence—
focuses on science literacy, evidence-based decision making, and the use
of validated knowledge to drive public policy for the benefit of all.

FIND OUT MORE:

aaas.org/meetings | [#AAASmtg](https://twitter.com/AAASmtg)