Special report: What it will take to vaccinate the world p. 896

Using CRISPR in diagnostic tests pp. 914 & 941

A comprehensive strategy for additive manufacturing p. 932

Science

$15
28 MAY 2021
sciencemag.org

AAAS

PROTECTING CORALS

Local actions defend reefs against climate change

pp. 908 & 977

EMPOWER WITH EVIDENCE

After a year of attending virtual convenings, we've come to appreciate many benefits, but we still long for the things that can only happen in person.

As vaccinations increase, case numbers decline, and states, localities, and institutions begin the process of reopening, we are all left wondering what the newest "new normal" will look like. Like most scientific organizations, we are actively confronting these questions, especially as they relate to our 2022 Annual Meeting.

We write now to share with you some of the preliminary plans for the 2022 meeting, and to ask for your patience and understanding as we work to finalize some of the outstanding details in the months ahead.

- **Philly, here we come!** We invite those who are comfortable doing so to join us February 17-20 in Philadelphia. Our in-person experience will feature live Plenary Lectures, Topical Lectures, special programming, and great social and networking opportunities—all in a city that is ready to welcome visitors—safely—with open arms.
- **Scientific Sessions will be virtual.** To ensure that as many people as possible are able to participate in the Scientific Sessions, they will be entirely virtual. Speakers can participate from wherever they feel comfortable, including Philly. Scientific Sessions will begin in January with the release of Spotlight Videos, which will prepare all meeting participants to engage in active discussion during the live panel sessions.

  For those who join us in Philly, we'll have Community Viewing Pods where attendees can view the Scientific Sessions with friends and colleagues, enabling those spontaneous hallway connections and conversations that we've all missed.
- **A first-rate virtual experience.** Not able to join us in Philly? No problem! You'll be able to participate in all of our Plenary Lectures, Topical Lectures, Scientific Sessions, and Workshops—live!—via a new and improved virtual meeting platform.
- **Deadlines remain the same.** The submission site for Scientific Sessions and Workshops will remain open through June 17. This year's meeting theme "Empower With Evidence" aims to inspire sessions that illustrate the transformative power of evidence-based discoveries and that address the challenges of rebuilding the public trust in science. Meeting registration will open in September.

The Annual Meeting team will continue to work tirelessly to create a convening that features compelling content, powerful professional development, and unparalleled networking opportunities in the safest possible settings.

We look forward to seeing you in Philly or online.

aaas.org/meetings | #AAASmtg

Susan Amara
President, AAAS
Chair, 2022 AAAS Annual Meeting

AAAS | ANNUAL MEETING

Workshop speakers. Top row (L to R): Yigong Shi, Narry Kim, and Ling-Ling Chen. Bottom row (L to R): Reinhard Lührmann, Adrian R. Krainer, and Xiaohua Shen.

Tsinghua–*Science* Workshops: Spliceosomes and RNA

As part of a recent series of online workshops organized by Tsinghua University and Science*/AAAS, an international panel of experts described the latest advances in our understanding of how RNA-protein complexes called spliceosomes play a critical role in disease and viral pathogenesis.*

Spliceosomes play a fundamental role in the production of life-sustaining proteins by accurately splicing pre-messenger RNA (pre-mRNA) into mature mRNA just before it's translated into new proteins. Variations in splicing or errors caused by dysfunctional spliceosomes, however, can also cause deadly diseases.

On December 2, 2020, structural biologist Yigong Shi of Tsinghua University/Westlake University, China, discussed the role of Prp2, one of the eight ATPase/helicases involved in spliceosome remodeling. Shi and colleagues analyzed the atomic structure of Prp2 alone and in complexes. They showed that elaborate mechanisms lie behind the impact of spliceosome remodeling on pre-mRNA splicing.

Spliceosome processing goes through several precatalytic stages called complexes, ending with the B complex that is turned into the activated spliceosome. In his presentation, Reinhard Lührmann—a chemist at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany—explained how small nuclear ribonucleoproteins (snRNPs) drive these structural rearrangements. By blocking the spliceosome's assembly at intermediate stages and analyzing them with cryo-electron microscopy, Lührmann and his colleagues determined the structure of two preactivated B complexes. They also revealed how spliceosome proteins participate in the formation of the activated B complex, including a conformational change in Prp8 that is crucial to the folding that activates the complex.

The active spliceosome drives a range of outcomes. As an example, changes in transcripts can accelerate the growth of SARS-CoV-2, as described by virus expert Narry Kim of Seoul National University, South Korea. She and her colleagues sequenced cells infected with the virus to determine the architecture of the transcriptome, particularly the canonical transcripts, which consist of a genomic RNA and nine subgenomic RNAs. Their work revealed that discontinuous mechanisms in translation create a complex transcriptome, which includes the encoding of unknown open reading frames. The team's research also revealed many proteins that impact SARS-CoV-2's growth, and thereby its pathogenicity, through interactions with viral RNA.

Actions of the spliceosome contribute to diseases beyond infectious ones, such as neurodegenerative conditions. At Cold Spring Harbor Laboratory in New York, Adrian Krainer studies ways to repair defects in splicing in neurodegenerative conditions, such as spinal muscular atrophy (SMA). Using antisense approaches for targeted splicing modulation, he helped to develop nusinersen (Spinraza), the first approved drug for SMA. As he explained, nusinersen corrects pre-mRNA splicing, increasing the level of survival motor neuron (SMN) protein, which is reduced in people with SMA. This approach is also promising for treating other neurodegenerative diseases.

In addition to diseases, RNA-based mechanisms participate in the normal developmental complexity of eukaryotes. Xiaohua Shen—an expert in noncoding portions of the genome at Tsinghua University in Beijing, China—described how genetic repeats, noncoding RNA, and RNA-binding proteins participate in the regulation of transcription and genome organization.

Studies of RNA splicing also reveal previously unknown regulators of gene expression. For example, Ling-Ling Chen—an RNA biologist at the Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences—works on circular and long noncoding RNAs (lncRNAs). Distinct pathways produce these forms of RNA. For instance, many lncRNAs are spliced like mRNAs, but some processes lead to circular forms. The action of all these RNAs, including gene regulation, depends on their structure and subcellular location.

Although these experts unveiled many new mechanisms involved in the creation and participation of the spliceosome in healthy and diseased conditions, their work also provided the basis for many more studies. This research will lead scientists even closer to understanding how the spliceosome determines the construction of mRNA and the ultimate production of proteins, and may also help them manipulate that knowledge to treat a range of diseases.

Sponsored by

FRCBS
北京生物結構前沿研究中心
Beijing Frontier Research Center for Biological Structure

IMAGES PROVIDED BY TSINGHUA UNIVERSITY

ADVERTISEMENT

May 28, 2021

Stop hate crimes and racism

We are physician and scientist organizations in North America. We vehemently condemn all forms of hatred, violence and racism in our society. While, sadly, these are not new issues, the recent uptick in discrimination and hate crimes against the Asian American and Pacific Islander community is devastating and unsettling to us all as a society.

The increasing hate crimes and violence towards Asian Americans and Pacific Islanders have incited fear and terror within our community. All human beings are free and equal in dignity and rights. No one should have to fear for their safety or be the victims of any kind of physical or emotional attacks due to their physical appearance or cultural and ethnic background. The right to freedom from discrimination is a fundamental human right—one that we must uphold.

Built upon and thriving on immigration, the United States has always been a beacon of hope and justice to attract talented and hardworking immigrants from all over the world. Immigrants and their descendants have been and will continue to make tremendous and incontrovertible contributions to the growth and prosperity of the US. Just for the scientific endeavor alone, over 30% of the Nobel Prize Laureates in the United States during 1901-2013 were born in other countries, including 6 from Greater China, 5 from Japan, and 7 from India[1, 2]. Asian Americans and Pacific Islanders are a vital and integral part of this diverse nation.

We call on you to denounce in the strongest terms all forms of discrimination and crimes against any racial/ethnic group, including Asian Americans and Pacific Islanders. We call on you to pursue racial equality and tolerance in the United States and beyond. We call on you to stand strong and together with us.

- American Association of Chinese Rehabilitation Physiatrists (AACRP)
- Asian American Accountable Care Organization (AAACO)
- Asian American Health Center (AAHC)
- Association of Chinese American Physicians (ACAP)
- Association of Chinese Americans in Cancer Research (ACACR)
- Association of Chinese Geneticists in America (ACGA)
- BayHelix Group
- Chinese American Allergy Professionals (CAAP)
- Chinese American BioMedical Association (CABA)
- Chinese American Biopharmaceutical Society (CABS)
- Chinese American Diabetes Association (CADA)
- Chinese American Doctor Association (Greater Baltimore region) (CADA)
- Chinese American Heart Association (CnAHA)
- Chinese American Hematologist and Oncologist Network (CAHON)
- Chinese American Independent Practice Association (CAIPA)
- Chinese American Lung Association (CALA)
- Chinese American Medical Association of California (CAMAC)
- Chinese American Neurological Association (CANA)
- Chinese American Pathologists Association (CAPA)
- Chinese American Society of Anesthesiology (CASA)
- Chinese Biological Investigators Society (CBIS)
- Chinese Biopharmaceutical Association, USA (CBA)
- Chinese Canadian Medical Society B.C. (CCMSBC)
- Chinese Community Accountable Care Organization (CCACO)
- Eastern Chinese American Physicians IPA, Inc. (ECAP)
- Excelsior Integrated Medical Group (EIMG)
- San Diego Chinese American Physicians Association (SDCAPA)
- Sino-American Biotechnology and Pharmaceutical Professional Association (SABPA)
- Sino-American Network for Therapeutic Radiology and Oncology (SANTRO)
- Society for Chinese Neuroscientists (SCN)
- Society of Chinese American Physician Entrepreneurs (SCAPE)
- Society of Chinese Bioscientists in America (SCBA)
- Society of Chinese Bioscientists in America Hematology Division (SCBA HD)
- Washington Association of Chinese American Physicians (WACAP)

References:

1. Institute for Immigration Research, Research brief, November 2013. http://s3.amazonaws.com/chssweb/documents/20864/original/Nobel_Prize_Research_Brief_Final.pdf?1447975594
2. http://www.leadthecompetition.in/GK/nobel-prize-winners-from-india.html

CONTENTS

28 MAY 2021 • VOLUME 372 • ISSUE 6545

ILLUSTRATION: STEPHAN SCHMITZ/FOLIOART

NEWS

IN BRIEF

IN DEPTH

FEATURES

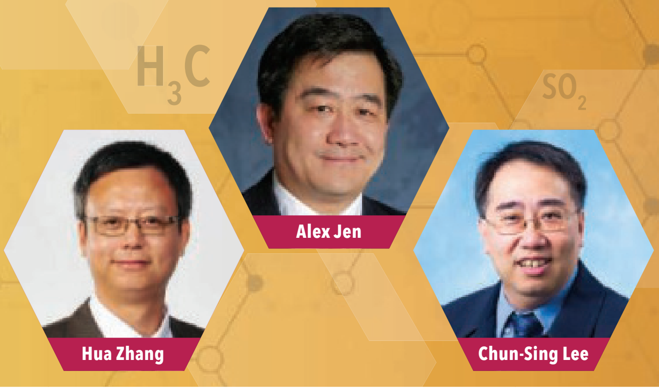
INSIGHTS

PERSPECTIVES

POLICY FORUM

Forging new paths at the intersection of chemistry and materials science

Close to 1,000 scientists attended the virtual Distinguished Lecture Series on Chemistry organized by the Hong Kong Institute for Advanced Study (HKIAS) at the City University of Hong Kong (CityU) during March and April 2021. Discussion topics ranged from the future of printable solar-cell technologies to the role of nanomaterials in developing clean energy solutions.



According to metrics compiled by Stanford University, CityU has over 140 full-time faculty members—eight of whom are HKIAS members—listed among the top 2% of the world's most highly cited scientists, reflecting its excellent academic stature.

Among them are Sir Colin Blakemore, Philippe G. Ciarlet, Way Kuo, Chain-Tsuan Liu, Jian Lu, and David J. Srolovitz (listed alphabetically), who are senior fellows of HKIAS; Tei-Wei Kuo, former visiting fellow of HKIAS; and Jacob C. Huang, executive director of HKIAS. The three speakers of the HKIAS Distinguished Lecture Series on Chemistry are also listed.

A printable green energy source

On March 3, 2021, Alex Jen, Lee Shau-Kee Chair Professor of Materials Science and Chair Professor of Chemistry and Materials Science at CityU, delivered a lecture titled "Printable Solar Cells for Transformative Clean Energy and Sustainable Society."

"Solar power is a potential sustainable solution because it can generate the largest amount of energy," said Jen. "The question is 'do we have enough devices to store or generate the power?'"

He introduced printable solar cells—an innovative energy-saving application—highlighting the advantages of organic solar cells and perovskite solar cells as two significant printable materials with immense potential. Barriers to their application, and possible solutions, were also discussed.

"The applications for printable solar cells are broad, including wearable gadgets and portable power sources," he said, expressing his hope that printable solar cells will be easy to scale up, comparable to printing a newspaper.

The future of nanomaterials

In the second lecture of the series, on March 31, 2021, Hua Zhang, Herman Hu Chair Professor of Nanomaterials at CityU, delivered a lecture titled "Phase Engineering of Nanomaterials (PEN)."

Zhang reviewed his group's recent research on PEN, published in *Nature Reviews Chemistry*, which focuses on the rational design and synthesis of novel nanomaterials with unconventional phases for various promising applications.

In response to a question about how nanomaterials are synthesized, he stressed that conventionally, the control of composition, morphology, size, dimension, and facet is important. By contrast, his group focuses on the phase control of nanomaterials, particularly unconventional crystal phases and amorphous structures.

"Although PEN is still under development, it's an important topic," said Zhang. "Its future applications are diverse, such as in catalysis, surface-enhanced Raman scattering, waveguides, and clean energy."

Finding new material properties

In the final lecture, on April 8, 2021, head of the CityU chemistry department and Chair Professor of Materials Chemistry Chun-Sing Lee delivered a lecture titled "Charge-Transfer Complexes and Their Applications."

Lee described charge-transfer complexes (CTCs): "People define CTCs in different ways, but there is still no agreed-upon definition," he explained. Adopting a simple definition, he called CTCs "a substantial charge transfer between donor and acceptor that achieves different properties from the parents."

Characteristic signatures of CTCs, such as a red-shift and broadened emission and absorption, were introduced. Lee also highlighted how CTCs enhance solar cells and LEDs when applied to these devices. Additionally, he described their wide application in optoelectronic devices, biomedicine, and energy and environment, providing relevant examples.

"The formation of CTCs is a simple way of getting new and unconventional properties from organic materials," he said, stressing their exciting potential in novel applications.

This lecture series at HKIAS is supported in part by the Kwang Hua Educational Foundation, which brings the work of its internally acclaimed scholars to a wide audience of young researchers in order to deepen the academic community's understanding of key areas of scientific endeavor.

Hong Kong Institute for Advanced Study
Phone: +852-3442-6611
Email: hkias@cityu.edu.hk

PHOTOS: COURTESY OF HKIAS

BOOKS *ET AL.*

922 Getting to the root of forest symbioses

An ecologist's memoir emphasizes the interdependencies of life

By J. B. Yoder

923 Moving physics forward

Weaving together her own story with reflections on the field, a physicist calls for progress *By M. Muzio*

LETTERS

925 Assisted colonization risk assessment

By A. Ricciardi and D. Simberloff

925 Response

By J. F. Brodie et al.

926 Inclusion and equity through STEM training

By M. H. Zaman

RESEARCH

IN BRIEF

929 From *Science* and other journals

REVIEW

932 Additive manufacture

Material-structure-performance integrated laser-metal additive manufacturing *D. Gu* et al.

REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG1487

RESEARCH ARTICLES

933 Neuroscience

Multiscale representation of very large environments in the hippocampus of flying bats *T. Eliav* et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG4020

PERSPECTIVE p. 913

934 Coronavirus

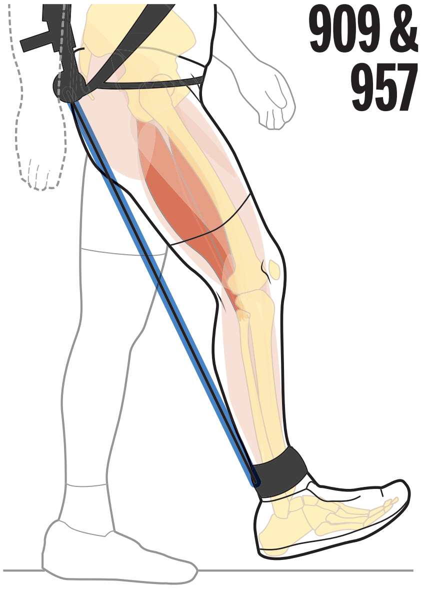
Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile *G. E. Mena* et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG5298

935 Cellular microbiology

Dynamic remodeling of host membranes by self-organizing bacterial effectors

T.-S. Hsieh et al.



941 Coronavirus

Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9 *C. Jiao* et al.

PERSPECTIVE P. 914

REPORTS

948 Quantum computing

Quantum walks on a programmable two-dimensional 62-qubit superconducting processor *M. Gong* et al.

952 Zeolite catalysis

Role of the ionic environment in enhancing the activity of reacting molecules in zeolite pores *N. Pfriem* et al.

957 Exoskeletons

Removing energy with an exoskeleton reduces the metabolic cost of walking

M. Shepertycky et al.

PERSPECTIVE p. 909

961 Ferroelectrics

Control of polarization in bulk ferroelectrics by mechanical dislocation imprint

M. Höfling et al.

964 Spin physics

Free coherent evolution of a coupled atomic spin system initialized by electron scattering *L. M. Veldman* et al.

968 Metabolism

Mitochondrial NADP(H) generation is essential for proline biosynthesis

J. Zhu et al.

973 Quantum criticality

Nematic quantum criticality in an Fe-based superconductor revealed by strain-tuning

T. Worasaran et al.

977 Coral reefs

Local conditions magnify coral loss after marine heatwaves

M. K. Donovan et al.

PERSPECTIVE p. 908

980 Ecosystem sentinels

Hemispheric asymmetry in ocean change and the productivity of ecosystem sentinels

W. J. Sydeman et al.

984 Genomic evolution

3D genomics across the tree of life reveals condensin II as a determinant of architecture type

C. Hoencamp et al.

990 Coronavirus

Resurgence of SARS-CoV-2: Detection by community viral surveillance

S. Riley et al.

DEPARTMENTS

885 Editorial

Managing Colorado River risk

By John Fleck and Brad Udall

1006 Working Life

Family first

By Robert Blasiak

ON THE COVER

Climate change threatens coral reefs by causing heat stress events, leading to widespread coral bleaching and mortality. Yet some coral reefs, such as this one in the Red Sea, are more resilient. A global analysis of 223 sites found that local stressors act synergistically with climate change to kill corals, but effective local management, alongside global efforts to mitigate climate change, can help reefs survive. See pages 908 and 977.

Photo: Tom Shlesinger

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. 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.sciencemag.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.

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

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

News Editor Tim Appenzeller

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

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **MULTIMEDIA MANAGING PRODUCER** Joel Goldberg **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Kelly Franklin **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **WEB DESIGNER** Jennie Pajerowski **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **VIDEO PRODUCER** Meagan Cantwell **SENIOR PODCAST PRODUCER** Sarah Crespi

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman **FINANCIAL ANALYST** Isacco Fusi **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy **PRODUCTION SPECIALIST** Kristin Wowk **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **DIGITAL MARKETING SPECIALIST** Asleigh Rojanavongse **SENIOR DESIGNER** Kim Huynh

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

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATE** Anne Mason **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **MARKETING MANAGER** Kess Knight **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer **SALES COORDINATOR** Josh Haverlock

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

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren

---

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

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

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

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

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

MEDIA CONTACTS
scipak@aaas.org

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

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services

science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers

advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org

support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

EDITORIAL

Managing Colorado River risk

In the 1920s, E. C. LaRue, a hydrologist at the United States Geological Survey, did an analysis of the Colorado River Basin that revealed the river could not reliably meet future water demands. No one heeded his warning. One hundred years later, water flow through the Colorado River is down by 20% and the basin's Lake Powell and Lake Mead—the nation's two largest reservoirs—are projected to be only 29% full by 2023. This river system, upon which 40 million North Americans in the United States and Mexico depend, is in trouble. But there is an opportunity to manage this crisis. Water allocation agreements from 2007 and 2019, designed to deal with a shrinking river, will be renegotiated over the next 4 years. Will decision-makers and politicians follow the science?

It has been said that climate change is water change. Globally, the effects on rivers vary widely, from increased risk of flooding in some places, to short-run increases in river flows in others as glaciers melt and catastrophes ensue once the glaciers are gone. The only constant is change, and our inability to rely on the way rivers used to flow. Like many snowmelt-fed rivers, for the Colorado this translates into less water for cities, farms, and the environment.

> "Will decision-makers and politicians follow the science?"

Research published over the past 5 years makes the threat clear. Runoff efficiency—the percentage of rain and snow that ends up as river water—is down, with half the decline since 2000 attributed to greenhouse-driven warming. For every 1°C of warming, researchers expect another 9% decline in the Colorado's flow. This year's snowpack was 80% of average but is delivering less than 30% of average river flows. Hot, dry summers bake soils, reducing flows the following year. The Colorado is not unusual. Researchers have identified similar patterns in other North American rivers, as well as in Europe, Asia, Africa, and Australia.

Colorado River water management has a long and uneasy relationship with science. LaRue's analysis of the early 20th century was brushed aside in favor of larger, more aspirational estimates of the river's flow made by bureaucrats who wanted to build dams. Scientists who agreed with LaRue—there were many—were ignored. This left the river overallocated and put the basin at risk.

Fortunately, there has since been progress in forging water management plans on the basis of science. For example, the US Bureau of Reclamation has been incorporating climate change into its analyses for more than a decade. Admirably, it overcame some of the political and technical challenges of incorporating the effects of climate change in the water allocation rules adopted in 2019. Models used to support decision-making were adapted to incorporate the 21st-century's declining flows. Computer simulations showing emptying reservoirs were enough to convince decision-makers of the need to cut back. But have the modelers gone far enough?

The scientific challenges are formidable. Although the direction of change—a shift toward less river water—is clear, the details can be murky. This is a challenge for the handoff from science to the world of policy and politics. But we cannot allow that murkiness to stand in the way of taking seriously what the climate science is telling us.

As the basin's water management community prepares for a new round of negotiations over the water allocation rules, how bad of a "worst case scenario" should be considered and who will get less water as a result? It is tempting to use today's 20% flow decline as the new baseline—that is, modeling future reductions on the basis of what has already been observed. But only by planning for even greater declines can we manage the real economic, social, and environmental risks of running low on a critical resource upon which 40 million North Americans depend.

The United States and Mexico—not just America's West and Southwest—can't afford to get this wrong. There are still political challenges that harken back to the struggles of E. C. LaRue a century ago—namely, as political boosters chose overoptimistic estimates of the river's flows to make their jobs easier. Climate science indicates that there will likely be less water in the Colorado River than many had hoped. This is inconvenient for 21st-century decision-makers, and overcoming their resistance may be the hardest challenge of all.

–John Fleck and Brad Udall

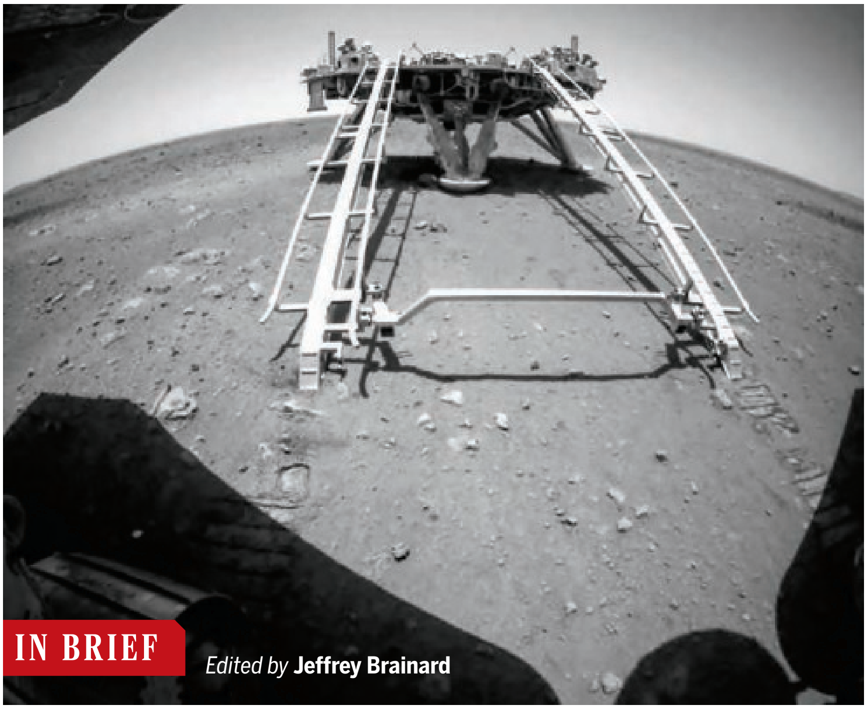
John Fleck is a Professor of Practice in Water Policy and Governance in the Department of Economics and director of the Water Resources Program at the University of New Mexico, Albuquerque, NM, USA. fleckj@unm.edu

Brad Udall Brad Udall is a Senior Water and Climate Research Scientist and Scholar in the Colorado Water Center at Colorado State University, Fort Collins, CO, USA. bradley.udall@colostate.edu

10.1126/science.abj5498

NEWS

> “I frequently vomit before going to the lab.”
>
> **Anonymous scientist,** in a survey conducted by the antibullying Academic Parity Movement and posted as a preprint. Many of the 2000 self-selected respondents said they had been bullied but didn’t report it to their institution, fearing retaliation. Most who did said they found the process unfair.



IN BRIEF

Edited by **Jeffrey Brainard**

PLANETARY SCIENCE

China’s rover rolls out on Red Planet

China’s Zhurong rover rolled off the Tianwen-1 Mars mission’s lander last week to start its explorations, leaving tracks in the Red Planet’s dust. Zhurong’s sensors will provide an up-close look at Mars’s subsurface strata, minerals, and weather and atmosphere. Researchers expect 90 days of operations but hope for more. Meanwhile, the mission’s orbiter will continue to gather data on Mars’s topography and ionosphere. The flawless 15 May landing and subsequent smooth deployment of the rover (which photographed the lander in the photo above) may encourage China in its ambitions for planetary exploration. After bringing rocks back from the Moon in December 2020, China is considering a landing site for a planned, second sample return mission to the Moon’s far side in 2023 or 2024.

Optogenetics eases blindness

GENE THERAPY | A blind man who received a gene for a light-sensing algal protein in one eye can now see objects with the help of special goggles, researchers report this week in *Nature Medicine*. It is the first published case of using optogenetics, a method of controlling neurons, to treat a disease in people. The 58-year-old French man was a participant in a clinical trial of the technique. He has an inherited disease called retinitis pigmentosa that destroys the eye’s light-sensing photoreceptor cells; he could sense light but not discern shapes. Researchers used a virus to insert the algal gene into the man’s retinal ganglion cells, which carried signals from the light-sensing protein to the brain. Months later, while wearing goggles that focused light on his retina, he could find and touch a notebook and count glass tumblers. If the treatment helps others, it may offer advantages over alternative technologies such as retinal implants.

U.S. science test scores drop

SCIENCE EDUCATION | The latest results from a quadrennial national test have disappointed U.S. science educators. The scores of fourth grade students in science showed a significant drop of three points on the National Assessment of Educational Progress (NAEP) between 2015 and 2019, while the scores of eighth and 12th grade students stayed flat. Key metrics for measuring how science is taught are also discouraging. For example, just 30% of fourth grade students engage in inquiry-based activities—a teaching method that studies have validated as more effective than others—only once or twice a year, and only 18% as often as twice a month. “Far too many elementary teachers have told us that science is not a priority in their schools and is perceived as less important than math and English language arts,” says Erika Shugart, executive director of the National Science Teaching Association. Until that attitude changes, Shugart says, “we can anticipate that lackluster NAEP scores will continue.”

U.S. global change chief returns

CLIMATE POLICY | President Joe Biden’s administration last week reinstated the past director of the U.S. Global Change Research Program (USGCRP), which coordinates climate science across 13 federal agencies and oversees a periodic and influential review, the National Climate Assessment. In November 2020, the Trump administration reassigned Michael Kuperberg, a climate scientist who had run USGCRP for 6 years, to the Department of Energy and replaced him with a climate change denier. Trump subordinates had criticized the 2018

PHOTO: CNSA VIA AP

installment of the climate assessment, which, like previous ones, predicted calamitous, costly effects from climate change. In restoring Kuperberg to his old post, the White House also directed USGCRP to accelerate its work on two fronts: advancing climate science on socially relevant topics and ensuring that knowledge is more easily accessible to the public. The next climate assessment is now due by the end of 2023.

Fusion plant takes the heat out

ENERGY | The United Kingdom's rebooted fusion reactor, MAST-Upgrade, has successfully demonstrated a novel exhaust system for superhot waste gases, key to making future commercial devices smaller and cheaper, researchers announced this week. Such reactors generate energy by fusing hydrogen isotopes in gas heated to more than 100 million degrees Celsius and confined with powerful magnets. As waste gases are expelled, they must touch a reactor surface, and not many materials can stand the heat for long. MAST-Upgrade was built with extra chambers and magnets, known as a super-X diverter, to lead the waste gases on a winding 20-meter path, during which they have time to cool. In tests since the reactor was fired up in October, researchers showed that this reduced the heat load 10-fold at the final contact surface. "This will make a big change in the amount of downtime in a future power plant," says lead scientist Andrew Kirk of the Culham Centre for Fusion Energy.

India endures black fungus

COVID-19 | India is suffering an epidemic within the pandemic. Nearly 9000 COVID-19 patients have also contracted mucormycosis, a rare disease also called black fungus, for its discolored lesions on the nose and inside the mouth. Spread by spores in the environment, mucormycosis has a mortality rate of more than 50%. Healthy people easily stave off infection, but those with weakened immune systems are vulnerable. In India, the disease is appearing mostly in COVID-19 patients given steroids to suppress an overactive immune response and in those who also suffer from diabetes. The surge in mucormycosis is causing a shortage of amphotericin B, the drug used to treat the disease. Last week, India's health ministry urged the country's 36 states and territories to declare mucormycosis epidemic, a step that leads to closer tracking of cases.

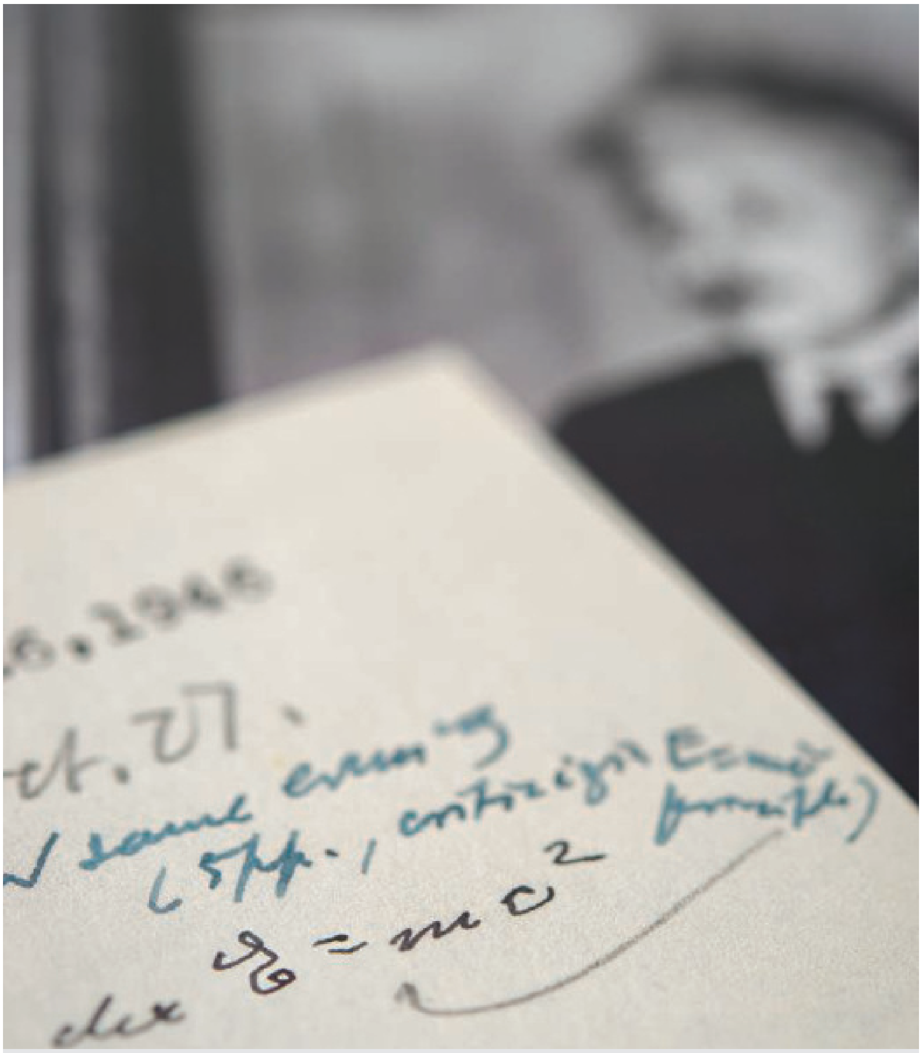


PHOTO: RR AUCTION

HISTORY OF SCIENCE

Hand-written 'E=mc^2' fetches high auction price

A picture may be worth 1000 words, but a letter containing a single equation written in Albert Einstein's shaky hand sold last week for a whopping $1.2 million, Boston-based RR Auction reported. Penned on 26 October 1946, when the famous theorist was 67, the letter contains one of four extant instances of Einstein's famous equation E=mc^2 in his own hand. The equation implies that a small amount of mass (m) equals a huge amount of energy (E) because the speed of light (c) is enormous. The letter's recipient, Polish-American physicist Ludwik Silberstein, wrote one of the first English-language textbooks on relativity—a theory that encompasses both the special theory of relativity that Einstein published in 1905 and the general theory, published in 1915, which explains the origins of gravity. Silberstein had doubts about general relativity and engaged Einstein in public debate—which Silberstein lost as the theory became a cornerstone of modern physics.

Threat to scientist spurs manhunt

COVID-19 | Belgian authorities have hunted for more than a week for a heavily armed former soldier who they said threatened a prominent virologist, Marc Van Ranst of KU Leuven, over his support of COVID-19 lockdowns. Police took Van Ranst, a member of two expert panels advising the government, and his family to a safe house on 18 May. The fugitive, far-right former military shooting instructor Jürgen Conings, is on the Belgian list of terrorism suspects and "very dangerous," Belgian Justice Minister Vincent Van Quickenborne said in a 21 May TV interview. He added that Conings spent 2 hours on 17 May near a "target," identified as Van Ranst by Belgian media. Van Ranst had been receiving police protection since July 2020 because he received pandemic-related threats regularly. Tweeting from his hideout last week, he said the threats "don't impress me at all."

Peer review spurs mixed feelings

PUBLISHING | Who voices more anxiety about peer reviews: researchers whose manuscripts have been accepted or rejected? The answer: It's a tie. That's one of the counterintuitive findings of a study of researchers' emotional and cognitive reactions to peer reviews, as revealed in more than 3600 comments posted by researchers on SciRev.org. The website allows authors to rate the quality of reviews at each of more than 3500 journals by name. The 16 May paper in *Scientometrics* unpacked the SciRev.org comments using language-analysis software. Although authors of rejected papers were not more likely to report anxiety than authors of accepted papers, they were more prone to say they were saddened by the decision. Also surprising, authors who waited longer for review decisions were no more likely to make negative comments about peer review. "Possibly, slow peer review processes have become so prevalent in academia [that] most authors did not bother to criticize it," wrote the study's author, Shan Jiang of the University of Massachusetts, Boston. Further research about psychological reactions to peer review could help improve the process, the paper suggests.

BY THE NUMBERS

8023

Depth in meters where the deepest ocean bed core was drilled. The sample, for earthquake research, was taken this month in the Japan Trench, near the epicenter of the 2011 quake that caused a tsunami and knocked out the Fukushima nuclear power plant.

Ancient DNA shows Asia upheaval

GENETICS | After the last ice age, the population of modern humans in northern East Asia may have undergone a major turnover, a study this week in *Cell* suggests. Researchers analyzed DNA from across the genomes of 25 ancient hunter-gatherers. It shows that the earliest known modern humans in the north China Plain, which stretches from Mongolia to the Amur Peninsula of Russia, who lived there 33,000 to 40,000 years ago, belonged to one widespread population. But by the end of the Last Glacial Maximum, about 19,000 years ago, they had been replaced by another population of people related to living East Asians and ancient Siberians. The first group may have died out during the ice age, the research team writes, noting that frigid temperatures in Europe may have driven a similar ancient population turnover.

Lunar GPS could guide missions

SPACE SCIENCE | Future robotic and crewed missions to the Moon could find their way on its surface more easily under a plan proposed by the European Space Agency (ESA) to establish a fleet of navigation satellites—a lunar version of GPS. The agency last week announced €2 million contracts to each of two industrial consortia to devise plans. Space agencies and companies hope to dispatch dozens of lunar probes this decade, but they must carry heavy radio gear to stay in contact with large dishes on Earth that guide them. Instead, ESA proposes that three or four satellites in lunar orbit and surface beacons could provide GPS-like signals so future missions could make do with a simple, less costly, lightweight receiver. The system would improve navigational accuracy, fixing position to within 100 meters compared with the current 500 meters at best.

Extinct plants with seed coats called cupules, including one called *Caytonia*, may have been forerunners of flowering plants.

EVOLUTION

Exquisite fossils in China illuminate the origin of flowering plants

With more than 350,000 species, flowering plants feed, fuel, and adorn the world. Now, researchers have taken a big step toward understanding the origin of traits that distinguish them from an older group of plants, the gymnosperms, which today includes pine trees and ginkgoes. Among the differences, the flowering plants or "angiosperms," which evolved about 125 million years ago, produce more sophisticated seeds, with two outer protective coats instead of just one. In 2017 in an open-pit coal mine in Inner Mongolia, palaeobotanists discovered evidence of an evolutionary link between these two major groups of plants: a treasure trove of exquisitely preserved, extinct gymnosperms with double-coated seeds. The outer coat or "cupule" most likely gave rise to the outer coat, the integument, of angiosperm seeds, palaeobotanist Gongle Shi of the Chinese Academy of Sciences's Nanjing Institute of Geology and Palaeontology and colleagues write this week in *Nature*. Modern gymnosperms lack cupules. The ancient plants, which no longer exist, also had specialized leaves or other tissues that may have been the forerunners of the female angiosperm reproductive structures called carpels.

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

PHOTO: G. SHI *ET AL.*, *NATURE* (2021) 10.1038/S41586-021-03598-W

Asmeret Asefaw Berhe is known for her climate-related research and efforts to increase diversity in science.

PHOTO: TED CONFERENCE/FLICKR/CC BY-NC-ND

BIDEN ADMINISTRATION

DOE science pick signals new focus on climate

Soil scientist Asmeret Berhe would take the lead at agency known for funding physics

By **Adrian Cho**

Sometimes a new presidential administration signals where it's headed through whom it selects to lead a federal research agency. That appears to be the case with President Joe Biden's choice to lead the Department of Energy's (DOE's) basic research wing, the Office of Science. Last month Biden tapped Asmeret Asefaw Berhe, a soil scientist at the University of California (UC), Merced, to lead the office, which has a $7 billion annual budget and is best known for funding physics, running national laboratories, and building atom smashers and other scientific megamachines.

The nomination of Berhe, 46, suggests the office will increasingly emphasize research related to climate change, scientists say. Berhe currently studies how factors such as erosion, fire, and temperature affect whether soil soaks up carbon dioxide or releases more of it into the air. She was born and raised in Eritrea and, if confirmed by the U.S. Senate, would be the first person of color to direct the office. (As usual for nominees awaiting confirmation, Berhe declined to be interviewed.)

Announced on 22 April, Berhe's nomination delighted many environmental researchers. "She's as star scientist as star scientists get," says soil ecologist Bala Chaudhary of DePaul University. Ecologist John Harte of UC Berkeley, who was Berhe's doctoral adviser, hopes her nomination marks a shift in DOE science from esoteric conceptual problems to addressing the climate crisis. "There will be, I hope, more emphasis on science that relates to the sustainability of the human enterprise as opposed to the mere sustainability of a scientific endeavor," he says.

Berhe has also long worked for greater diversity in the sciences, says geochemist Peggy O'Day of UC Merced. "She's been a real leader, both on our campus as well as nationally and internationally, in advocating for people of color in science," O'Day says. Last year, Berhe and Chaudhary published a paper in *PLOS Computational Biology* entitled, "Ten simple rules for building an anti-racist lab."

But some physicists worry Berhe may have trouble guiding the often-fractious agency, citing her scant experience managing large organizations and her unusual scientific background for a position often held by physicists. According to her CV, Berhe has held one DOE grant for $200,000 and has served as interim associate dean of UC Merced's graduate division.

As the nation's single largest funder of the physical sciences, the Office of Science supports six research programs, including fusion energy sciences, high energy physics, and nuclear physics. Its basic energy sciences program funds chemistry, materials science, and condensed matter physics, and its advanced scientific computing program provides supercomputing for myriad studies. Biological and environmental research get 10.7% of its budget. The office owns 10 of DOE's 17 national labs and builds big scientific facilities—the newest is a $730 million particle accelerator at Michigan State University.

The director's job is to set priorities among the competing research programs and coordinate billion-dollar construction projects so that as one nears completion the next is ready to go, says Bill Madia, a nuclear physicist and former director of two national labs. "It's one of the most important management jobs in science in the world," he says. "You're comparing priorities from bioenergy centers to neutrino experiments to exascale computers."

Given that much of the office's money goes to physics, Michael Lubell, a physicist at City College of New York and former head of public affairs for the American Physical Society, wonders how, as a biogeochemist, Berhe will approach those decisions. "There's nothing in her background to suggest that she knows anything about fusion, or particle physics, or nuclear physics, or atomic physics," he says.

Most past office directors have had a mixture of training in physics, experience running large organizations, and work history with DOE. But that background is not a prerequisite for success, says Raymond Orbach, a theoretical physicist and former chancellor of UC Irvine who directed the office from 2002 to 2009. Orbach won plaudits for, among other things, developing a 20-year to-do list of major projects that DOE has largely followed. But he notes that he, too, was a newcomer to DOE. "One never knows how someone with no prior formal government service (e.g. me) will turn out," he wrote in an email. The office's most recent director, Christopher Fall, has a doctorate in neuroscience and had prior management experience at DOE and the Office of Naval Research.

Some directors with traditional credentials have struggled with the job. William Brinkman, a theoretical physicist who led the office from 2009 to 2013 under former President Barack Obama, came to DOE with 14 years of experience as a director at the storied private Bell Labs. But the scholarly and cerebral Brinkman found it difficult to communicate with Congress, Lubell says. During one hearing, a legislator pressed Brinkman for a plan to deal with a particular issue. To lawmakers' dismay, Lubell recalls, "Brinkman pointed to his head and said, 'It's in here.'"

No director has to do it all on her own, notes physicist Cherry Murray of the University of Arizona, who was director from 2015 to 2017. DOE has a corps of staffers who are "incredibly competent" and can help keep the agency humming, she says. "I'm not worried at all about physics research dropping by the wayside" under Berhe, she says. "That will continue, just as under me biology research continued." Murray says she is curious to see where Berhe will head in setting policy.

If Berhe is confirmed, her success will largely rest with budgetmakers in Congress. For example, even though former President Donald Trump repeatedly tried to slash the office's budget, Congress increased it by 31% over 4 years. That boost spared Fall from having to make unpopular cuts. If the budget keeps growing, Berhe may enjoy a long honeymoon with DOE-sponsored researchers.

Should budgets tighten, she could face the challenge of retaining the support of the community while picking winners and losers. Berhe has the leadership skills to meet that potential challenge, Harte says. "I would call her steadfast with good humor and an extraordinary thoughtfulness," he says. "She will gather the respect of others because of her intense intelligence." ■

MATERIALS SCIENCE

Zinc aims to beat lithium batteries at storing energy

Rechargeable batteries based on zinc promise to be cheaper and safer for grid storage

By **Robert F. Service**

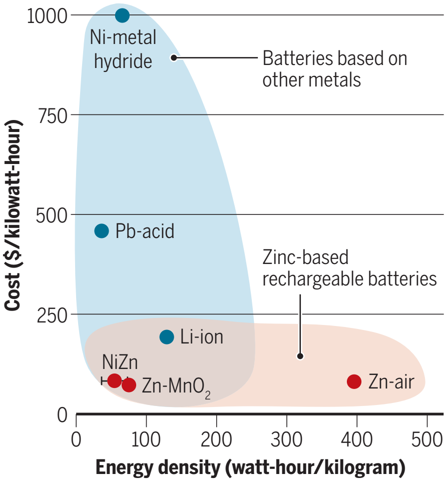
If necessity is the mother of invention, potential profit has to be the father. Both incentives are driving an effort to transform zinc batteries from small, throwaway cells often used in hearing aids into rechargeable behemoths that could be attached to the power grid, storing solar or wind power for nighttime or when the wind is calm. With startups proliferating and lab studies coming thick and fast, "Zinc batteries are a very hot field," says Chunsheng Wang, a battery expert at the University of Maryland, College Park.

Lithium-ion batteries—giant versions of those found in electric vehicles—are the current front-runners for storing renewable energy, but their components can be expensive. Zinc batteries are easier on the wallet and the planet—and lab experiments are now pointing to ways around their primary drawback: They can't be recharged over and over for decades.

The need for grid-scale battery storage is growing as increasing amounts of solar, wind, and other renewable energy come online. This year, President Joe Biden committed to making the U.S. electricity grid carbon free by 2035. To even out dips in supply, much of that renewable power will have to be stored for hours or days, and then fed back into the grid. In California alone, the public utilities commission envisions deploying more than 8800 megawatts of rechargeable batteries by 2026, and last week, California Governor Gavin Newsom proposed $350 million in state funding to develop long-duration energy storage technologies. "That trend will not go down. It will only continue to grow," says Mark Baggio, vice president for business development at Zinc8 Energy Solutions, a zinc battery producer.

A better battery

Zinc is cheaper than many battery metals and could store more energy.



GRAPHIC: B. HOPKINS *ET AL.*, *SUSTAINABLE ENERGY & FUELS*, 4, 3363 (2020), ADAPTED BY C. BICKEL/*SCIENCE*

For power storage, "Lithium-ion is the 800-pound gorilla," says Michael Burz, CEO of EnZinc, a zinc battery startup. But lithium, a relatively rare metal that's only mined in a handful of countries, is too scarce and expensive to back up the world's utility grids. (It's also in demand from automakers for electric vehicles.) Lithium-ion batteries also typically use a flammable liquid electrolyte. That means megawatt-scale batteries must have pricey cooling and fire-suppression technology. "We need an alternative to lithium," says Debra Rolison, who heads advanced electrochemical materials research at the Naval Research Laboratory (NRL).

Enter zinc, a silvery, nontoxic, cheap, abundant metal. Nonrechargeable zinc batteries have been on the market for decades. More recently, some zinc rechargeables have also been commercialized, but they tend to have limited energy storage capacity. Another technology—zinc flow cell batteries—is also making strides. But it requires more complex valves, pumps, and tanks to operate. So, researchers are now working to improve another variety, zinc-air cells.

In these batteries, a water-based electrolyte spiked with potassium hydroxide or another alkaline material separates a zinc anode and a cathode made of other conductive materials, often porous carbon. During discharge, oxygen from the air reacts with water at the cathode to form hydroxide ions, which migrate to the anode, where they react with zinc to eventually produce zinc oxide. The reaction releases electrons that flow from anode to cathode through an external circuit. Recharging the batteries

This energy storage facility under construction in southeast England uses lithium-ion batteries.

PHOTO: CHRIS RATCLIFFE/BLOOMBERG/GETTY IMAGES

means reversing the flow of current, causing zinc metal to re-form on the anode.

But zinc batteries don't like to run in reverse. Irregularities across the anode's surface cause the electric field to intensify at certain spots, which causes zinc to deposit there, further enhancing the electric field. As the cycle repeats, tiny spikes called dendrites grow, eventually perforating and shorting out the battery. Equally troublesome, water in the electrolyte can react at the anode, splitting into oxygen and hydrogen gas, which can burst the cells apart.

Researchers have begun to deal with these downsides, churning out nearly 1000 papers per year. In 2017, for example, Rolison and colleagues reported in *Science* that they reengineered the anode as a 3D network of zinc metal pocked with tiny voids. The electrode's vast surface area reduced the local electric field, which prevented the buildup of dendrites and reduced the likelihood of splitting water molecules. NRL licensed the technology to EnZinc.

This month, Wang and his colleagues reported in *Nature Nanotechnology* that when they added a fluorine-containing salt to their electrolyte, it reacted with zinc to form a solid zinc fluoride barrier around the anode. Ions could still wriggle through during charging and discharging. But the barrier prevented dendrites from growing and repelled water molecules, blocking them from reaching the anode.

"It's a great development," says Wei Wang, who directs the Energy Storage Materials Initiative at the Pacific Northwest National Laboratory. Still, Chunsheng Wang notes his device is somewhat slow to discharge. To improve that, his team wants to add catalysts at the cathode to speed up the reaction between oxygen and water.

The same strategy features in work by researchers led by Jung-Ho Lee from Hanyang University. In *Nature Energy* on 12 April, they reported creating a fibrous and conductive cathode from a mix of copper, phosphorus, and sulfur that also served as a catalyst, dramatically speeding up oxygen's reaction with water. That and other advances produced batteries that could be charged and discharged quickly and had high capacity, 460 watt-hours per kilogram (compared with about 75 Wh/kg for standard zinc cells with manganese oxide cathodes and 120 Wh/kg for scaled-up lithium-ion systems). The batteries were stable for thousands of cycles of charge and discharge. The result "looks like another important step," Chunsheng Wang says.

Such advances are injecting new hope that rechargeable zinc-air batteries will one day be able to take on lithium. Because of the low cost of their materials, grid-scale zinc-air batteries could cost $100 per kilowatt-hour, less than half the cost of today's cheapest lithium-ion versions. "There is a lot of promise here," Burz says. But researchers still need to scale up their production from small button cells and cellphone-size pouches to shipping container–size systems, all while maintaining their performance, a process that will likely take years. Burz also notes electric utilities and other companies looking to buy cheap large-scale batteries want to see years of steady operation first. "These customers need to see that it works in the real environment," he says. ■

PUBLIC HEALTH

Studies test lifestyle changes to avert dementia

"Multidomain" trials look for brain benefits from multiple, simultaneous alterations

By **Mitch Leslie**

For the past 3 years, about 6000 middle-aged and elderly Australians have pumped iron, loaded up on greens and whole grains, strived to quell stress, and challenged their wits with computer exercises, all in an effort to preserve their cognition. They're part of a clinical trial called Maintain Your Brain, one of about 30 current or planned studies that eschew pharmaceutical interventions and test whether altering multiple aspects of participants' lives improves brain health. Such multidomain studies may finally reveal whether modifying diet, exercise, and other factors can slow cognitive decline as people age—or even prevent dementia.

"There's a lot of hope for multidomain trials," says psychologist Kaarin Anstey of the University of New South Wales, Sydney, one of the principal investigators of the Maintain Your Brain trial, which will finish by the end of this year.

Although people can't escape some mental decline as they get older, lifestyle exerts a powerful influence over the risk of developing dementia—the type of severe cognitive impairment seen in conditions such as Alzheimer's disease. Last year, an international committee of scientists and psychiatrists known as the Lancet Commission on dementia prevention, intervention, and care estimated that so-called modifiable factors account for 40% of dementia risk. Their report highlighted a dozen factors, including many familiar villains—diabetes, high blood pressure, smoking, obesity, and lack of exercise.

Researchers are still probing exactly how these risk factors steal people's faculties, but they've identified some likely mechanisms. Lack of physical activity may impair cognition, for instance, because exercise stimulates formation of new neurons and soothes brain inflammation.

For decades scientists concentrated on developing drugs to treat Alzheimer's disease, but after several candidates recently failed in clinical trials, "the climate has really shifted to focus on … prevention," says neuropsychiatrist and epidemiologist Kristine Yaffe of the University of California, San Francisco. Some researchers urge governments to step up dementia prevention with measures such as public health campaigns that encourage salutary habits. "We have knowledge about some of the actions to take for society to make a difference," says psychiatrist Gill Livingston of University College London, who heads the Lancet Commission. "The time is now."

The combined effect of lifestyle factors is strong, but researchers lack conclusive evidence that modifying any of them spares the brain. "A gazillion observational studies" point to factors that influence cognitive aging, Yaffe says. "Can we say, 'Do X, Y, and Z and that will prevent Alzheimer's disease'? I don't think so."

Large, randomized, controlled trials would provide the strongest support for particular interventions. But these costly studies are rare. Multiyear trials, which have the best shot at detecting an impact on a slow-developing condition like dementia, are even rarer.

The only study to show that any lifestyle intervention cuts dementia risk was the Advanced Cognitive Training for Independent and Vital Elderly trial, launched in the late 1990s. A group of nearly 700 elderly people in the study underwent 6 weeks of cognitive training to improve their thinking speed. Ten years later, they had a 6% lower incidence of dementia than participants who received no training. But the benefits of "brain training" remain unsettled, and many commercially available games and apps lack rigorous evidence to support their claims.

Scientists are eager to nail down the value of other interventions. Observational studies suggest the Mediterranean diet, which is heavy on olive oil, fish, and whole grains but light on red meat and sugars, improves some aspects of cognition. But no large, randomized, controlled trial has tested the preventive effects of switching to the diet—or compared it with other promising regimens such as dietary approaches to stop hypertension (DASH), designed to reduce blood pressure. A 3-year study wrapping up later this year will evaluate the Mediterranean-DASH intervention for neurodegenerative delay diet, which merges the Mediterranean and DASH diets, in 600 people over age 60 who are at high risk for developing dementia. That could provide "the first evidence of whether changing diet prevents cognitive decline in older adults," says cognitive neuropsychologist Lisa Barnes of Rush University, principal investigator of the $14 million, NIH-funded study.

> **"In the next 10 years, we will see if lifestyle works."**
>
> **Laura Baker**,
> Wake Forest School of Medicine

However, most randomized trials that have focused on only one aspect of lifestyle have come up empty. Many researchers agree that multidomain trials like Maintain Your Brain offer a better chance of finding meaningful effects. Dementia results from multiple causes, the argument goes, so preventing it will require a combination of interventions.

A Finnish trial launched in 2009 was the pioneer of the genre. The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER) included 1260 people in their 60s and 70s who were susceptible to dementia because of risk factors such as hypertension. Half of them took part in an intensive program to improve their diet, heart health, mental acuity, and exercise habits. The control group received health advice from nurses but no help implementing it.

The finding that people in the intensive program improved their cognitive tests scores over the 2-year study period electrified a major Alzheimer's conference in 2014, says Heather Snyder, vice president for medical and scientific relations at the nonprofit Alzheimer's Association. "It was the first study to demonstrate on that scale and scope that these interventions in synergy could affect cognition."

Two similar studies, conducted in the Netherlands and France, also suggested a cognitive benefit in a subset of people with high dementia risk. No one knows how long the effects persist or whether the interventions prevent dementia, but ongoing follow-up of FINGER participants might provide some answers.

Now, scientists in other countries are setting up trials tailored to the diets and habits of their populations. One of Maintain Your Brain's goals is to determine whether multidomain interventions can be delivered remotely. Participants log into the trial's website to obtain coaching, access materials such as cooking demonstrations, and record their progress.

The U.S. Study to Protect Brain Health Through Lifestyle Intervention to Reduce Risk, scheduled to finish in 2024, aims to reproduce the results of the FINGER trial in the more racially and ethnically diverse U.S. population. It will test whether subjects do better when they are assigned a specific plan or when they have freedom to customize their own. "If we can create a sustainable program that's accessible for everyone … that would be a huge success," says cognitive neuroscientist Laura Baker of Wake Forest School of Medicine, who heads the trial.

Whether lifestyle changes can stall dementia or cognitive decline remains unproved for now, but Baker anticipates a mountain of new evidence. "In the next 10 years, we will see if lifestyle works." ■

Physical activity, a factor in healthy brain aging, is part of multidomain dementia prevention trials.

PHOTO: ISTOCK.COM/KALI9

COVID-19

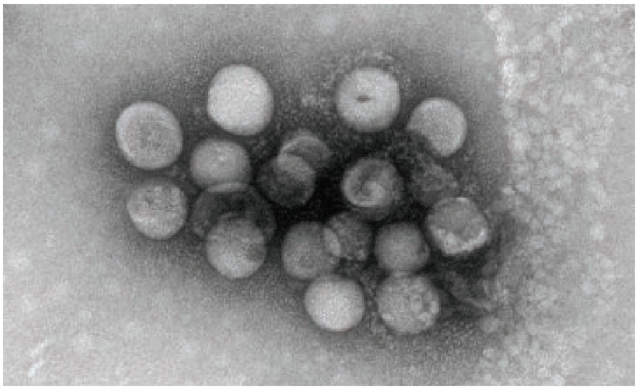
Two more coronaviruses may infect people

Concern grows about the pandemic potential of other members of the virus family

By **Anthony King**

Coronaviruses, already notorious for spilling over into people from various animals and causing new diseases—most catastrophically, COVID-19—may jump into humans even more often than researchers suspected.

Last week, an international collaboration that went looking for known or novel viruses in pneumonia patients in Malaysia reported that in eight children, they found signs of a coronavirus that may have originated in dogs. Earlier this year another group reported a coronavirus that appears to have jumped from pigs to several children in Haiti. There's no sign so far that either virus can spread from person to person—as the spark of the pandemic, SARS-CoV-2, readily does—or definitive evidence that they cause human illness. But the discoveries, which could increase the number of coronaviruses known to infect people from seven to nine, underscore the threat posed by this viral family.



An electron microscope image of a new coronavirus isolated from a child in Malaysia with pneumonia and grown in dog cells.

IMAGE: MOLECULAR AND CELLULAR IMAGING CENTER/OHIO AGRICULTURAL RESEARCH AND DEVELOPMENT CENTER/OHIO STATE UNIVERSITY

"I think the more we look, the more we will find that these coronaviruses are crossing species everywhere," says virologist Stanley Perlman of the University of Iowa.

Malaysian researchers originally partnered with a group at Duke University to study 301 adults and children hospitalized with pneumonia in 2017–18. The eight children with signs of the coronavirus were mainly living in traditional longhouses or villages on Borneo, where they likely had frequent exposure to domestic animals and jungle wildlife. Standard hospital diagnostics for pneumonia or other respiratory illness would not have detected nonhuman coronaviruses, but the Duke team, led by virologist Gregory Gray, had developed a genetic test for conserved coronavirus sequences.

The researchers screened nasopharyngeal samples—secretions and cells swabbed from the upper part of the throat in each patient—and in the children found gene sequences suggesting a novel canine coronavirus. Collaborators from Ohio State University (OSU), Wooster, then cultured virus from one of the children's samples and sequenced its whole genome. The finding, reported in *Clinical Infectious Diseases*, is the first report indicating a caninelike coronavirus can replicate in people, and further studies will need to confirm the ability.

The OSU researchers have grown the virus in dog tumor cells, but not yet in human tissue. They are also studying its features, including the spike surface protein that all coronaviruses use to bind to cells and initiate an infection.

The Malaysian children had other virus infections as well, so the coronavirus may not have been the cause of pneumonia. (They all stayed in the hospital for 4 to 6 days and fully recovered.) And each infection may have been a dead-end jump into one person from a nonhuman host, instead of from human-to-human spread. "We don't have any clear evidence that this particular [coronavirus] strain is better adapted to humans because of its spike structure," says OSU veterinary virologist Anastasia Vlasova, who led the effort to culture the virus.

Although the genome of the virus overall resembles a canine coronavirus, its spike is closely related to those of canine coronavirus type I and a pig coronavirus known as transmissible gastroenteritis virus. And one key part of the protein bears a 97% similarity to the spike of a coronavirus that infects cats.

This chimera likely arose through multiple genetic swaps between various coronaviruses coinfecting nonhuman hosts. "This is a mosaic of several different recombinations, happening over and over, when nobody's watching. And then boom, you get this monstrosity," says virologist Benjamin Neuman at Texas A&M University, College Station. The final host that transmitted the novel virus to the children could have been a cat, pig, dog, "or some wild carnivores," notes Vito Martella, a veterinary virologist at the University of Bari in Italy.

The genome marks the virus as an alpha coronavirus—the same genus as two coronaviruses that cause common colds. So far, the most dangerous human coronaviruses—those that cause COVID-19 and two other deadly illnesses, severe acute respiratory syndrome and Middle East respiratory syndrome—are in the beta genus. Researchers haven't seen alphas trigger an outbreak of serious disease in humans, Neuman says, "but that doesn't feel like much comfort in the wild world of viruses."

Coronaviruses in another genus may also pose a threat to jump into humans. In March, researchers at the University of Florida reported in a medRxiv preprint the first evidence of a delta coronavirus that infects people, finding signs of a known pig coronavirus in serum from three Haitian children who had fevers in 2014–15. Delta coronaviruses were once thought to infect only birds. Then, in 2012, one infected pigs in Hong Kong. It "appears to have jumped over from songbirds," says OSU coronavirologist Linda Saif, who grew the virus in swine cell cultures.

Porcine coronaviruses like the ones in Hong Kong and Haiti are a potential epidemic threat, say Ralph Baric of the University of North Carolina, Chapel Hill, and other virologists, because pigs are notorious "mixing vessels" in which viruses can swap genes—the same process that generates new and dangerous strains of influenza. Concern about the virus in Haiti would intensify even more if it was proven to transmit between people, Saif says.

Both studies point to the need for increased vigilance, Baric adds. "This research clearly shows that more studies are desperately needed to evaluate … the frequency of cross-species [coronavirus] transmission and potential for human-to-human spread." ■

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

Anthony King is a journalist in Dublin.

Human embryos created through in vitro fertilization are ethically sensitive research material.

DEVELOPMENTAL BIOLOGY

Door opened to more permissive research on human embryos

Updated stem cell guidelines also herald new ways to study development

By **Kelly Servick**

The world's largest stem cell society this week signaled a willingness to reconsider a long-standing restriction on laboratory efforts to grow and study human embryos. In new guidelines, the International Society for Stem Cell Research (ISSCR) also spotlights a possible alternative to using embryos that might be less ethically fraught: emerging methods to model stages of human development with stem cells.

ISSCR's influential guidelines previously put the culture of human embryos beyond 14 days postfertilization in its most restrictive category three: "prohibited research activities." The new guidelines, drafted by a task force of scientists and ethicists, omit longer embryo culture from this category and encourage a public discussion about allowing it.

The guidelines aren't legally enforceable. Laws limiting embryo research to either 14 days or formation of a structure called the primitive streak exist in several countries, including the United Kingdom, Sweden, and South Korea. But ISSCR "has an important soft power," says Annelien Bredenoord, a bioethicist at the University Medical Center Utrecht who is a member of ISSCR's ethics committee but not part of the guideline task force. "It can be the one that fuels the discussion."

Until recently the 14-day rule had little practical effect because embryos didn't survive that long in the lab. But recent advances have allowed some researchers to run experiments up to that limit.

Observing embryonic development after 14 days could help scientists better understand the origins of miscarriages and birth defects, and many scientists and bioethicists have advocated softening the restriction. In a commentary in *Science* (5 March, p. 998), researchers and ethicists including Bredenoord urged policymakers and ISSCR to "consider a cautious, stepwise approach to scientific exploration beyond the 14-day limit."

The new guidelines don't explicitly move extended embryo culture into ISSCR's less restrictive category two, which describes research permissible after scientific and ethical review. "The ISSCR has not abandoned the 14-day rule," says Amander Clark, a stem cell biologist at the University of California, Los Angeles, and a member of the guideline task force. But the guidelines call for "national academies of science, academic societies, funders, and regulators to lead public conversations" about the "societal and ethical issues raised by allowing such research."

Those groups will have to decide whether changing current restrictions is "really worth the political capital and the legal battles," says cell biologist Martin Pera of the Jackson Laboratory. The 14-day limit "has given the public a good deal of reassurance around the boundaries of human embryo research," he says, "but probably merits a careful reexamination."

Pera notes that human embryos begin to degenerate as they approach the 14-day limit. Even as scientists get better at keeping embryos alive past 14 days, only a minority will survive in a dish, he says, and the number that remain healthy will diminish as development proceeds.

To study developmental processes beyond 14 days, Pera is more optimistic about embryo models created from human stem cells, which in recent years have become increasingly complex and powerful. These lab-grown structures can be made in unlimited quantities, he notes, and enable researchers to study the role of particular mutations in development and disease. In a commentary this week in *Stem Cell Reports*, members of the guideline task force and other scientists say the need to validate stem cell–based embryo models by comparing them with natural embryos is one reason to remove the 14-day limit.

Some of these stem cell–based models carry ethical sensitivities of their own. ISSCR's updated guidelines propose a distinction between nonintegrated models—which re-create only certain features of an embryo and lack supportive "extraembryonic" cells crucial to survival in a uterus—and more complete integrated models.

The new guidelines say research on nonintegrated models can proceed without special review. But such review is necessary for integrated models such as recently reported "blastoids," which closely resemble the human blastocyst—a stage about 5 days after fertilization when an embryo implants in the wall of the uterus (*Science*, 19 March, p. 1189). Such models could eventually reveal why many pregnancies fail at this stage and could help researchers refine in vitro fertilization techniques.

But integrated models should be "maintained in culture for the minimum time necessary to achieve the scientific objective," ISSCR says, and no embryo model should be transferred into a uterus to develop further. ■

IMAGE: ZEPHYR/SCIENCE SOURCE

BIOTECHNOLOGY

Thaw coming for U.K. gene-editing regulations

Government expected to loosen rules for some biotech crops and animals

By **Erik Stokstad**

When Boris Johnson became prime minister of the United Kingdom in 2019, he pledged to "liberate the U.K.'s extraordinary bioscience sector from anti–genetic modification rules." The country had to hew to strict European biotech regulations until it finalized its divorce from the European Union in January. Next month, the government is widely expected to follow through on Johnson's promise by making it easier to test and commercialize some genetically engineered crops and livestock.

The decision, which will be announced by 17 June, applies to plants and animals whose genes have been edited with precision techniques such as CRISPR. It will put the United Kingdom in line with several countries including the United States, and U.K. biotechnologists say it will speed research and stimulate investment.

"Much as I have to swallow hard and say it through gritted teeth, Brexit has at least one dividend," says Jonathan Jones, a plant biologist at the Sainsbury Laboratory, a public crop research center. Tina Barsby, CEO of the National Institute of Agricultural Botany, says the shift may be "the most significant policy breakthrough in plant breeding for more than 2 decades."

Traditional genetic engineering endows organisms with new traits by inserting "transgenes" from other species. In contrast, gene editing alters a species' own genes without permanently adding any new genetic material. Proponents argue gene editing is merely an acceleration of classical breeding techniques, which select for traits enhanced by mutations (often created by chemicals or radiation). "We have no reason to believe that they will be any more inherently risky than crops made with traditional breeding," says Angela Karp, director of Rothamsted Research, a U.K. nonprofit agricultural research center.

Under the U.K. policy change, gene-edited plants and animals might not need detailed applications and reviews before field trials and commercial approval. In Europe, by contrast, any commercialized genetically modified organism (GMO), regardless of how it was created, faces a lengthy risk assessment by the European Food Safety Authority and must be approved by a majority of member nations before it can be planted. "It means everything just grinds to a halt," says Wendy Harwood, head of crop transformation at the John Innes Center, a U.K. public research organization. In 2018, the European Court of Justice reaffirmed that gene-edited organisms require the same regulatory scrutiny as other GMOs (*Science*, 3 August 2018, p. 435).

Only a few gene-edited crops have been commercialized anywhere. One example is a tomato called the Sicilian Rouge High GABA that makes more of an amino acid said to promote relaxation, approved for sale in Japan last year. Just two gene-edited crops have made it to U.K. field trials. One, in 2018, evaluated the performance of camelina, a mustard relative, engineered to produce an olive oil–like product. And in a recent trial, researchers tested broccoli edited for improved nutrition.

U.K. rules on gene editing are expected to be less strict than those for transgenic crops like this iron-rich wheat the John Innes Center is testing.

PHOTO: JOHN INNES CENTRE/FLICKR

Others are in the works. Rothamsted Research this month applied for a permit to field test wheat edited to contain less asparagine, an amino acid that becomes the carcinogen acrylamide when baked. The Roslin Institute, a research center at the University of Edinburgh that works on livestock, has created pigs resistant to a virus that causes porcine reproductive and respiratory syndrome, which costs U.S. and European pig farmers $2.6 billion per year. Genus PLC is commercializing the pigs in several nations.

The government decision on gene editing, which will come from the Department for Environment, Food & Rural Affairs (Defra), will not apply outside England. Other parts of the United Kingdom—Scotland, Wales, and Northern Ireland—regulate GMOs themselves and are skeptical of their value. And opponents to GM liberalization say Defra is moving too fast. They worry, for example, that animals and crops modified to resist disease could promote environmentally damaging intensive farming practices.

It's important to address such concerns, says Colin Campbell, director of the James Hutton Institute, a public research center that focuses on sustainable management of natural resources. Biotechnologists "need a license from society to operate," he says. "The commercialization can follow when you've won the trust."

Proponents also need to have realistic expectations about gene editing, says Johnathan Napier, a plant biotechnologist at Rothamsted Research. Knocking out a few genes might improve disease resistance or remove an allergen. But more complicated traits powered by many genes, such as drought tolerance, will be much more difficult to engineer without transgenic modifications, Napier warns. "This really is not a magic bullet," he says. But controls on transgenic GMOs could someday be loosened as well; Defra has requested public comments on whether reform is needed.

Even the European Union is rethinking its approach on gene editing. An April report by the European Commission finds it could make agriculture more sustainable and found "strong indications" that EU law isn't suitable for regulating it. Dirk Inzé, a molecular biologist at the Flanders Institute for Biotechnology, a Belgian research center, is heartened. But he predicts any reforms would run into problems with the European Parliament, where anti-GMO sentiment is still strong. "The debate will be very fierce," Inzé says. ■

FEATURES

DOSES OF REALITY

The Serum Institute of India aimed to be a major world supplier of COVID-19 vaccines. India's pandemic got in the way

By **Jon Cohen**, *in Pune, India*

In a world of have and have-nots, Adar Poonawalla is most decidedly a have, with both abundant personal wealth and a corner on what promises to be a massive supply of one of the world's most desperately sought commodities: COVID-19 vaccines. Affable but feisty and a fan of bespoke suits, the 40-year-old heads the Serum Institute of India, the world's largest maker of vaccines. Before the pandemic, its factories in India annually churned out 1.5 billion doses of vaccines—50% more than the next largest producer—to protect against 13 different diseases.

Poonawalla's list of personal haves is long: an office fashioned from a retired Airbus 320 jet set on Serum's 40-hectare campus. Ferraris, a Rolls Royce, and a Batmobile. A helicopter and private jets. He splits his time between a luxurious ranch house on the family's stud farm abutting the company, a 9-hectare home across town in the upscale Salisbury Park neighborhood, a former maharaja's mansion in Mumbai, and a $69,000-a-week rental in London.

But when Poonawalla met with *Science* in early April, he shied away from the airplane office. "I've had a lot of flak on my lifestyle and opulence, and I don't want that to take away from the serious work we're doing here," said Poonawalla, sitting in a new office that sports a bust of Apollo, the god of healing and epidemics. Part of that work was to supply the world this year with hundreds of millions of doses of Covishield, an inexpensive, effective COVID-19 vaccine that was the brainchild of the University of Oxford and shepherded through clinical trials by AstraZeneca.

It hasn't worked out as planned. A catastrophic fire and temporary U.S. restrictions on the export of key raw materials slowed production early this spring. The Oxford-AstraZeneca vaccine ran into difficulties, with confusing efficacy results and, more recently, concerns about a rare but serious blood clotting side effect. Most devastating to Serum's aim to become the world's premier COVID-19 vaccine supplier to lower income countries, India faces an enormous surge of infections and a dire need for an ample supply of its own. Under government pressure, the company reduced Covishield exports in March and stopped them altogether in mid-April. "I don't like to use the word 'forced,'" Poonawalla says. "We are, after all, an Indian company that supplies vaccines globally. Right now, our nation needs us."

Serum's woes—and India's—have exacerbated the global COVID-19 vaccine inequity between haves and have-nots (see p. 903). By 20 May, 124 countries and territories had only received a total of 68 million doses

PHOTO: SERUM INSTITUTE OF INDIA

The original campus of the Serum Institute of India (left) covers 20 hectares next to its equally large biotech park, which vastly expanded its vaccine production capacity. CEO Adar Poonawalla (right) has even greater ambitions.

PHOTO: RAJA SENGUPTA

of any COVID-19 vaccine, while China had 467 million and the United States 280 million. The COVID-19 Vaccines Global Access (COVAX) Facility, which coordinates distribution of doses to the world's poor, had tried to forestall the inequity by investing heavily in Covishield, paying $3 per dose—far less than most other vaccines would have cost. Covishield is also easy to store and transport.

But Serum's decision to keep everything for India prevented 90 million ordered doses from reaching low-income countries in March and April, and it recently said it may not export doses until the end of the year. COVAX will fall far short of its target of 247 million delivered vaccine doses by the end of May, and it's facing the possibility that without Serum, its projections for 2 billion shots this year could be cut in half. "Right now, it's quite depressing," says the World Health Organization's (WHO's) chief scientist and liaison to COVAX, Soumya Swaminathan, who is from India.

The wild COVID-19 vaccine ride exasperates Poonawalla. He has faced challenges from a leading Indian vaccine competitor, a slight from Prime Minister Narendra Modi, and a media trouncing for threatening to sue a Covishield recipient who alleged injury. "I've chosen to just grow a thick skin and ignore everything," Poonawalla says.

But no one can ignore that the vaccine bet placed by COVAX, on behalf of the world, hasn't paid off yet. Serum still does not know when it can fulfill its promise of producing enough doses to help stymie the pandemic, at home or abroad. "Everything that we do—good, bad, ugly—is going to be examined with a microscope, and we'll have to deal with it," Poonawalla says.

CYRUS POONAWALLA, Adar's father, told his own father 55 years ago he wanted to break from the family's lucrative racehorses. Back then, the stud farm donated retired horses to a Bombay scientific institute that made treatments for snake venoms and tetanus toxins by injecting the animals with the substances and collecting their antibody-filled serum. Now, Cyrus wanted to go into the serum business himself. With his brother Zavaray Poonawalla as a partner, Cyrus raised $12,000 selling horses, which their father matched. They set up the new company on 5 hectares the farm used as a horse graveyard. Within 2 years, it had an anti-tetanus product on the market.

From the outset, Serum wanted to undersell the competition. "When we launched the tetanus antitoxin, I could have sold it at five times the price," Cyrus says. "From day one, we had the philosophy of being philanthropic or reasonable."

Reporting for this story was supported by the Pulitzer Center.

Serum has always been a for-profit company, not a charity, but both Cyrus and Adar repeatedly refer to their work as philanthropic. "All the products right from then until today, I could have doubled my price [periodically], but I've refrained from doing so and that's still not stopped me, with the grace of God, from becoming a billionaire," Cyrus says.

The company has profited from sheer volume, building one modestly sized plant after another to increase output. It began to supply India's expanded program on immunization for children in 1978 and selling vaccines to U.N. agencies in 1993. When Gavi, the Vaccine Alliance launched in 2000 with $750 million from the Bill & Melinda Gates Foundation, Serum's vaccines helped the organization further bolster childhood immunization in developing countries. By now 65% of the world's children receive at least one product made by the company. "It's phenomenal," says Anthony Fauci, head of the U.S. National Institute of Allergy and Infectious Diseases. "They have the capability of making enough vaccine for six planets."

Today, the company has 5000 employees and an annual revenue of $850 million. "Cyrus has built an incredible franchise, and he was able to grow the capacity and modernize it and be a high-quality producer of vaccines in enormous volumes," says epidemiologist Seth Berkley, CEO of Gavi, which in 2018 gave Cyrus a "vaccine hero" award. "And he did it without being greedy."

So Serum was a natural partner for a team at Oxford developing what it hoped would become a "global vaccine" for COVID-19, one affordable to low- and middle-income countries. Sarah Gilbert at Oxford's Jenner Institute had created a candidate COVID-19 vaccine from a putatively harmless chimpanzee adenovirus, engineering it to carry the gene for the spike protein on the surface of SARS-CoV-2. Adrian Hill, who runs the Jenner Institute, had been making a candidate malaria vaccine for 3 years with Serum, and already had strong relationships with its head of R&D.

Gavi and the Coalition for Epidemic Preparedness Innovations (CEPI)—another COVAX partner—encouraged the marriage, too. Representatives from the Gates Foundation also joined the matchmaking. CEPI urged the Oxford team to bring in a large pharmaceutical company that had experience staging clinical trials and navigating regulatory approval processes. "It didn't seem ridiculous that we could have it both ways: Serum doing most of the world and a Big Pharma partner for high-income countries," Hill says. Ultimately, an agreement was sealed with AstraZeneca.

The Poonawallas were, at first, chagrined that Oxford enlisted AstraZeneca. "We were the ones aligned to them in terms of being philanthropic," says Adar, who began working at Serum in 2001 and became CEO 10 years later. "In hindsight, I don't think that was the wrong decision because of AstraZeneca's capabilities to conduct these global trials. That would have been something that we would have struggled with."

In the end, a patchwork of clinical trials in three countries showed in December 2020 that the adenovirus-based vaccine effectively protects people from COVID-19, although not as well as the messenger RNA vaccines from other companies. Serum got the rights to sell the vaccine to 92 low- and middle-income countries that have joined COVAX, while AstraZeneca can supply the rest of the world. For the time being, AstraZeneca has agreed to sell the vaccine at no profit, inking country-specific deals ranging from $5 per dose to below Serum's $3-per-dose contract with COVAX. (Serum with the government's permission recently began to sell Covishield to private hospitals in India for $8 per dose.)

As this year began, Serum's role in supplying the neediest countries with vaccine seemed assured. The Gates Foundation gave $300 million to Gavi, which directed it to Serum to both accelerate manufacturing of Covishield and another low-cost vaccine, developed by Novavax, and contribute to COVAX's initial purchase of 100 million doses of each product. COVAX has an option for another 900 million doses of the two vaccines. Serum devoted unused production lines set aside for other vaccines to the project, investing $300 million of its own money, and WHO granted Covishield an emergency use listing—a COVAX prerequisite—on 15 February. Serum shipped out 18 million doses to COVAX 10 days later. "They were unbelievable," Berkley says.

But Serum, like the maker of almost every COVID-19 vaccine, has had unforeseen delays. In January, a fire killed five construction workers building a new plant to fill and finish vaccines. Then, the U.S. government, which until recently was worried about its own COVID-19 vaccine supply, restricted exports of vaccine raw materials such as liners for bioreactors and filters. Adar predicted last year that Serum would produce 100 million doses of Covishield per month by February, but production has never topped 70 million.

"Right now, our nation needs us."

Adar Poonawalla, Serum Institute of India

In late April, with daily COVID-19 cases in India spiking to nearly 350,000, the U.S. government relented and said it would make the materials available "immediately." But by then India's burgeoning need for vaccines had largely negated Serum's ability to supply doses to COVAX and the world. To date, it has delivered about 220 million doses of Covishield. More than 155 million of those doses have stayed home, supplying 90% of the shots delivered in India. COVAX has received just 28 million.

INDIAN POLITICS have added to Serum's woes. Last summer, the government stunned scientists in India and abroad when it announced a bold plan to fast-track a COVID-19 vaccine, made not by Serum but by Bharat Biotech of Hyderabad, with help from the National Institute of Virology. The government's Indian Council of Medical Research on 2 July 2020 issued a letter that said the "indigenous" vaccine made by Bharat from inactivated virus, called Covaxin, would be ready for "public health use" 6 weeks later, on 15 August—even though it had just entered initial trials.

Some researchers believed nationalism rather than science was at work. "Why this one was special is not really clear to me," says Gagandeep "Cherry" Kang, a vaccine researcher at Christian Medical College, Vellore. "Covishield is [also] an indigenous vaccine—it's made in India." To Kang and others, Covaxin's early authorization spotlighted India's already suspect medical regulatory system. "Ultimately, we want India to be able to make drugs and vaccines for the world, which means the world has to be able to trust what we are doing," she says.

The 15 August deadline came and went without Covaxin receiving regulatory approval, but in early January, both Indian vaccines received what amounts to an emergency use authorization. In a 3 January interview on India's popular NDTV, Adar celebrated the decision for Serum, but took a swipe at Covaxin. Only AstraZeneca and Oxford, Pfizer and BioNTech, and Moderna had presented convincing vaccine efficacy data, he stressed. "Everything else has proven to be safe, just like water," he said. The comment angered Bharat's CEO, Krishna Ella, who held a press conference the next day and fired back with his own insults about Covishield.

The CEOs quickly made peace in a joint statement, and Bharat has since presented preliminary data showing 81% efficacy for Covaxin, better than the AstraZeneca-Oxford vaccine by some measures. But the spat fueled growing vaccine hesitancy in India (see p. 900), adding to doubts fed by an earlier dispute: After a participant in a small trial of Covishield in India filed suit alleging neurological harm, the company threatened to countersue for defamation, demanding $13.5 million—20 times more than the man's suit. The hubbub faded, and Cyrus is unrepentant. "I'm not Jesus Christ to give my second cheek because I was slapped on one," he says.

However much of a shine the Modi government took to Covaxin, Bharat does not have the capacity to make enough of the vaccine for the country. Yet India's first contract with Serum was only made after Covishield received authorization, and it was for a mere 11 million doses. "The government sort of twiddled its thumbs until the very last moment," Kang says.

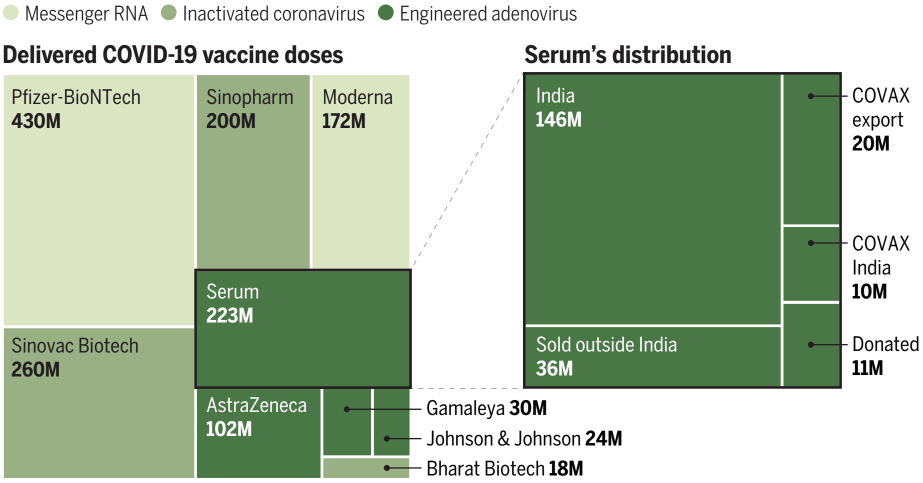
On 1 March, in an image posted on social media and now plastered on billboards across the country, Modi received his first dose of a COVID-19 vaccine. It was Covaxin.

A few weeks later, as India's second wave of COVID-19 began, the vaccine shortage became glaring. "They didn't plan and now they're in a panic situation," says an Indian researcher who has worked closely with the government but asked not to be named. "Suddenly, they want all the doses that are being produced."

Berkley has grown weary of addressing COVAX's big bet on Serum. "If I had had $12 billion, and held control over the largest biopharmaceutical companies and

Shots for the world

The Serum Institute of India has become a massive manufacturer of an inexpensive COVID-19 vaccine (production line shown below), but the surge of cases at home has forced it to stop exports. The COVID-19 Vaccines Global Access (COVAX) Facility, an effort to supply vaccine to low-income countries that relies heavily on Serum, is now 90 million doses short of what it expected by May because India received them instead. Serum might not export COVID-19 vaccines again until the year's end.


Delivered COVID-19 vaccine doses

Serum's distribution

could force them to make vaccines for me, it would have been a different outcome," he says. "Serum is a 20-plus–year reliable manufacturer with us, and they had met all of their goals. They've not had any quality control problems. But here we are, and they're not keeping any of their contracts."

Adar says the burden should not all fall on his company "I plead with other vaccine manufacturers to provide more doses to COVAX so that the pressure on me is off."

Swaminathan wants wealthy countries that have already vaccinated a large percentage of their populations to increase donations to COVAX, which is beginning to happen. And China's Sinopharm has started to negotiate with COVAX to provide hundreds of millions of doses of its vaccine. Both Moderna and Pfizer also have contracts to sell what could be hundreds of millions of doses to COVAX at discount prices.

Another huge Indian vaccine company, Biological E., could address the global need while supplying India. It has a contract with Johnson & Johnson and expects to produce up to 400 million doses this year of its vaccine, which also uses an adenovirus to produce spike but only requires a single shot. Biological E.'s managing director, Mahima Datla, insists that she will not renege on any deals with COVAX. "We believe all lives are created equal," says Datla, the representative on the Gavi board for developing country vaccine manufacturers. "I also believe until everybody's safe, nobody's safe."

Bharat, which once planned to produce only 200 million doses of Covaxin a year, is also expanding manufacturing at plants with a $200 million credit from the Indian government and hopes to more than triple output over the next year. Serum should increase output, too. India agreed on 19 April to give it a $400 million credit to help the company boost production to its originally promised 100 million doses of Covishield per month—or possibly even beyond.

ADAR POONAWALLA HAS large ambitions, which his father corralled for years. "I was a man in a hurry in my 20s and 30s, trying to prove myself, and I wanted to expand very fast, but he prevented me from doing that, and I think prevented me from making a lot of mistakes. He said, 'Look, you've got five vaccines. Why do you need any more, you're still very profitable, just grow those.'"

But Adar is already thinking ahead. In 2019, Serum opened the $400 million Poonawalla Bio-Technology Park, which doubled the size of the campus. The site is again crowded with cement trucks, cranes, and workers adding 28,000 square meters of space to vastly increase Serum's bulk manufacturing capabilities. It will include a biosafety level-3 lab that can handle all but the most dangerous of pathogens, so Serum can for the first time grow and then inactivate large batches of a virus or bacteria, an old but reliable approach to making a vaccine.

Serum has clinical trials underway of eight vaccines, for various diseases, with partners. Adar wants to expand the business to start selling vaccines and monoclonal antibodies to the United States and countries in Europe. "We are not treated in the same way that Big Pharma is treated," he says. "[Governments] don't see us as a partner they can work with even though we match the same quality, even though we have pricing that would significantly reduce their health care budgets."

In what may be Adar's loftiest enterprise, part of the latest construction includes a plant that can quickly switch to making vaccines for the next pandemic. "Imagine a dedicated facility, which can handle messenger RNA, viral vectors, anything—fill, finish, store, and dispatch—within 2 or 3 months of pushing a button," he says.

Adar predicts the pandemic plant could produce 1 billion doses per year, and he hopes governments that never again want to beg for doses and haggle over prices will commit to a kind of annual vaccine insurance payment. Serum's pandemic facility will guarantee a $3-per-dose vaccine to subscribers, with the number of doses depending on the size of a government's investment. "This is really the solution for the future," he says. "There are always going to be emerging diseases." ■

CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (PHOTO) RAJA SENGUPTA

Wajid Kadar Khan, who lives in a Mumbai slum, received a COVID-19 vaccine only because his job required it. His wife, Mumtaz (left), says she didn't need one because "I have God."

DEADLY DELAYS

Unexpected vaccine hesitancy, dose shortages, and poor government planning have compounded India's massive immunization challenge

By **Jon Cohen**, *in Mumbai, Vellore, and New Delhi, India;*
Photography by **Raja Sengupta**

This story was produced in partnership with the Pulitzer Center.

On a Sunday morning in early April, as Mumbai was in a daze from the first weeks of a surge of COVID-19 and had instituted nighttime curfews, Baliram Boomkar asked his neighbors in the city's Kaula Bandar slum whether they wanted a vaccine to protect them or had received one. Some replied that they had been vaccinated but only because their employers required it. One man said he'd get the shot if his company gave him time off to recover from side effects. "COVID is nothing," he said. "People are only spreading rumors. It's all a lie." A woman said she was afraid a clinic might test her for COVID-19, find she's positive, and then force her to quarantine—as happened last year. "I know I can't avoid the vaccine, but I want to be the last in the queue."

"Lots of people [here] don't believe that COVID exists and that God will provide if something happens," said Boomkar, who lives in the slum and works as a "barefoot researcher" for the nongovernmental organization Pukar, which conducts health-related studies and also tries to improve living conditions. "They think it's all politics." The use of masks, despite the barefoot researchers distributing them and stressing their benefits, remained sparse.

A month later, India's COVID-19 surge has become a tsunami, with hospitals overwhelmed and funeral pyres burning throughout the nights. Yet the country's vaccination campaign is languishing, with only 3% of Indians fully vaccinated as of 20 May. Widespread shortages of the shots have forced some vaccination clinics to shutter; at others, lines often form hours before they open. Some states are limiting doses to people older than 45, and to extend supplies, the government has recommended stretching the intervals between shots of the country's most heavily used vaccine, Covishield, a version of the AstraZeneca–University of Oxford vaccine produced by the Serum Institute of India. But supply is only half of the dilemma.

Delivering vaccines to India's nearly 1.4 billion people means reaching remote, difficult-to-access regions and tackling the profound divides between the lower and upper classes. And like almost everywhere in the world, India has the perplexing challenge of vaccine hesitancy. It's widespread in Indian society, far from limited to the slums that Pukar helps, but it is a new problem here. "India never had vaccine hesitancy" until COVID-19, says virologist Shahid Jameel, who directs the Trivedi School of Biosciences at Ashoka University.

Past mass vaccination campaigns in India focused on children. Adults, even the wealthiest, do not routinely get immunized against influenza, shingles, pneumococcal disease, or anything else. "You won't have too many adults asking for a vaccine, and you won't have too many doctors prescribing it either," says Renu Swarup, who heads the government's Department of Biotechnology. "There is a lot of advocacy that we have to do to bring the public on board."

Many blame a different surge for creating India's unexpected reluctance toward COVID-19 vaccines: the rumors that spread constantly on social media. "It's not a vaccine hesitancy that is deep rooted, like in Europe or the United States," says Sai Prasad, an executive director at Bharat Biotech, which makes Covaxin, the country's other COVID-19 vaccine. "This is literally due to disinformation or misinformation." Among the false assertions in wide circulation are that the vaccines make people impotent, are worthless because some vaccinated people become infected, or even lead to death. "Adults are more finicky than children: They change their minds thanks to WhatsApp University and Twitter on a second-by-second basis," Prasad says.

INDIA BEGAN ITS VACCINATION program on 16 January, just 1 month later than the United States and the United Kingdom. But there was little sense of urgency. The nation wasn't hit as hard by COVID-19 in 2020 as many expected. By 1 March, India, which allows anyone eligible for a shot to make an appointment at a local site through a single online portal, had vaccinated just over 12 million people with a first dose.

Even health care workers, the first in line for shots, were slow to get them. At the Christian Medical College (CMC), Vellore, an esteemed training ground for doctors and nurses that has five campuses with more than 2700 hospital beds, 30% of the staff still had not received a shot 6 weeks after the vaccination campaign began. By early April, after CMC administrators decided to post their own vaccination photos on social media and emphasized that 1600 unvaccinated staff had become infected and 12 had fallen critically ill, 99% of doctors and 90% of nurses and other hospital workers had received a shot.

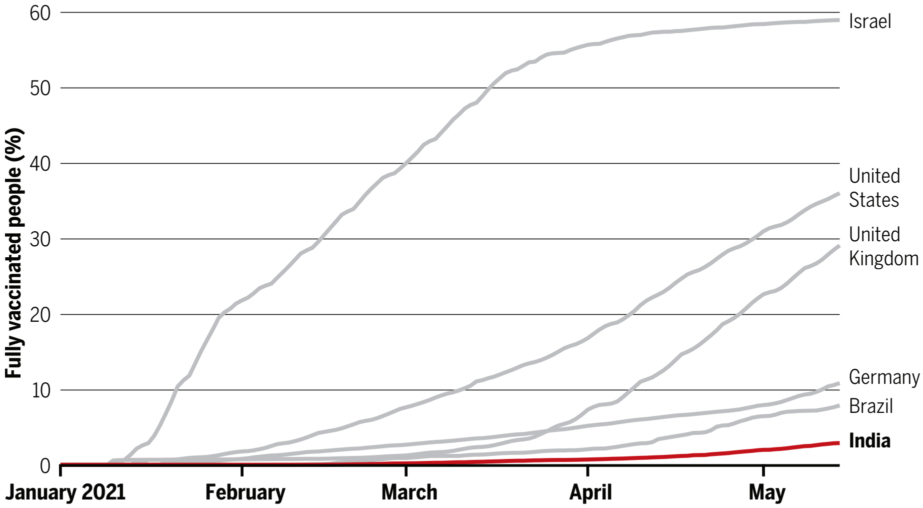
CMC is in Tamil Nadu, one of the country's most urbanized and industrialized states, and the tepid reaction toward the vaccine extended to the broader public On a morning in early April in Vellore, CMC vaccine researcher Gagandeep Kang walked downstairs from her office on the main campus to the hospital's COVID-19 vaccination clinic for her

Baliram Boomkar (top), a "barefoot researcher" who lives and work in a Mumbai slum, has had trouble persuading neighbors to wear masks or get a COVID-19 vaccine. But in some places in India, such as this rural hospital in Shikrapur, a town outside of Pune, people wait in long lines to get their shots (bottom).

PHOTOS: RAJA SENGUPTA

A slow start

A combination of vaccine hesitancy, supply issues, and the government's haphazard planning has thwarted India's efforts to protect its large population from COVID-19.



second dose. Kang paid her 250 rupees (about $3) and was vaccinated. But only a dozen other people sat in the outdoor waiting area. No one took a selfie as they got the shot or high-fived a nurse in thanks. Across town that day in the Salavanpet neighborhood government clinic where vaccine is free, only 22 people showed. The hospital had 370 doses in its refrigerator left from a batch of 500 it had received 5 days earlier.

Tamil Nadu hadn't yet been slammed by this COVID-19 surge. But even in parts of India where cases were mounting, the disease wasn't always perceived as a big threat. "You're in an environment where you see death so frequently," says CMC head J. V. Peter, a critical care specialist. "When you see people dying due to other illnesses at a higher frequency than COVID, why should people pump their fists and say, 'Hey, I've got my vaccine!' or why should they push towards getting a vaccine?"

Kang faults the government for not "preparing the ground" earlier for a massive adult immunization program. "The systems were set up for 100 people a day at immunization centers," she says. "We could scale up to five times what we're doing."

The challenges multiply in more rural areas. In Jawadhi Hills a few hours' drive away, Kang and others at her college have a project at the rural village of Vallithathankottai, helping the Malayali tribe with everything from clinical services to improved sanitation. The village's 99 houses are nestled up a steep mountain road, and a few dozen members of the tribe gathered one afternoon in their leader's house to discuss the pandemic with Kang and *Science*. Only three villagers had been vaccinated, at a clinic 5 kilometers away. Others were noncommittal. "If it's for our protection, we will all take the vaccine," said one villager, who like the others was not eligible at the time. But there was little fear of the virus. "It's not going to come to us," said one villager. Or it's simply harmless, the leader speculated. "We might have got it and it would have gone without us knowing."

IN MID-MAY, hopes rose that the devastating wave of COVID-19 was peaking in much of India, but cases continued to climb in some areas, including Tamil Nadu. And varying degrees of vaccine hesitancy remained. In wealthier, urban communities, the fact that the company making Covaxin has yet to publish its efficacy data and reports that the version of Covishield used outside of India can cause clotting problems continued to feed some reluctance. But demand for vaccine is growing, say researchers, some of who speculate that many wealthier Indians would rush to get the messenger RNA vaccines now only available abroad. "There have been few signs of hesitancy among the middle class and they are scrambling for vaccine slots," Kang now says.

A clinic at King Edward Memorial Hospital in Pune, India, stamped the arms of its COVID-19 vaccine recipients.

Neonatologist Anita Patil-Deshmukh, who founded and runs Pukar, says that in the Kaula Bandar slum, the surge has led at least "a few" pandemic doubters to change their minds, as they watched constant images of crematoria on TV and had relatives in their home villages become ill and unable to access care. But they remain exceptions. "Most people are still reluctant to take [the vaccine]. Vaccinating people who live in the slums is still a huge issue," she says.

The government needs to make it easier for the poor, she says. "Most people in the bottom of the pyramid do not possess the smartphones needed to do online registration, and those few who may possess it do not know how to navigate the system," she says, adding that Pukar soon hopes to set up registration stations in Kaula Bandar.

Kang said the Indian government should fulfill a commitment to setting up vaccination points within 2 kilometers of everyone. "We're a big country, and to reach people is challenging." She suggests some areas may need vaccinators to go door to door. "In India, in many places you have to think about outreach programs because the most vulnerable people are not going to get to vaccination centers."

Despite India's huge population, the effort could pay off quickly, some researchers argue. "Trying to vaccinate everybody is not the point," says Anurag Agrawal, a pulmonologist who heads the Institute of Genomics and Integrative Biology, a division of India's Council of Scientific and Industrial Research. India has a relatively large population of young people, who may be less vulnerable to serious symptoms. If immunization becomes widespread in those who are 45 and older, particularly in those with conditions like diabetes and obesity that can worsen COVID-19, hospital admissions and death will plummet, Agrawal contends. He calculates this population only totals about 200 million—a number India's vaccine supply should soon be able to cover.

"India does not really have a vaccine problem," he says. "It has a people's outlook problem. And this upsurge may again bring people back to reality." ■

CREDITS: (PHOTO) RAJA SENGUPTA; (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) OUR WORLD IN DATA REPOSITORY VIA JOHNS HOPKINS CENTER FOR SYSTEMS SCIENCE AND ENGINEERING

ILLUSTRATION: STEPHAN SCHMITZ/FOLIOART

FAIRER SHARES

Rich countries cornered the marketplace for COVID-19 vaccines. Here are four strategies to protect the rest of the world

By **Jon Cohen** *and* **Kai Kupferschmidt**

In January, the director-general of the World Health Organization (WHO), Tedros Adhanom Ghebreyesus, issued a blunt warning. The world was "on the brink of a catastrophic moral failure," he said. Wealthy countries were buying up available COVID-19 vaccines, leaving tiny amounts for others—a replay of what happened during the 2009 influenza pandemic. "The price of this failure will be paid with lives and livelihoods in the world's poorest countries," Tedros said.

He was right. Today, some rich countries are vaccinating children as young as 12 years old, who are at extremely low risk of developing severe COVID-19, while poorer countries don't even have enough shots for health care workers. Nearly 85% of the COVID-19 vaccine doses administered to date have gone to people in high-income and upper middle–income countries. The countries with the lowest gross domestic product per capita only have 0.3%.

Tedros lambasted the "scandalous inequity" again in his opening speech at the World Health Assembly on 24 May. By September, at least 10% of the population in every country should be vaccinated, he said.

Disparities in global health are nothing new. Lifesaving therapies such as monoclonal antibodies are unavailable in large parts of the world. Even vaccines and drugs that cost pennies to make don't reach millions of people who need them. But the COVID-19 crisis has exposed the inequities in a distinct, acute way. As normality is returning to vaccine front-runners such as Israel, the United Kingdom, and the United States, India's health system is buckling under soaring

case numbers—and the world is still recording almost 5 million cases and more than 80,000 deaths every week.

The moral argument aside, there's a very practical reason to try to distribute vaccines more equitably: No part of the world can feel safe if the pandemic rages on elsewhere, posing the risk of reintroduction and spawning potentially more dangerous viral mutants.

Things might get better. The COVID-19 Vaccines Global Access (COVAX) Facility, a nonprofit, is purchasing shots in bulk at a discount and distributing them to the world's most resource-strapped countries. Although it's in major trouble just now, in part because its main supplier, based in India, is reneging on its promises, there is reason to be hopeful: Many rich countries will soon have big stashes of superfluous vaccines—1 billion doses 4 months from now, by one projection—that they could donate to countries in need.

Production capacity is increasing rapidly, too. Some projections suggest enough vaccine could be available this year to give every person in the world at least one dose, although getting them to every country and into every arm would still be a major challenge. For the longer term, plans are now afoot to build vaccine production facilities in more regions of the world to pump up supplies, both for this pandemic and future ones.

Here's a look at all of these efforts—and when they might offer relief.

WEEKS TO MONTHS

COVAX

COVAX was formed in April 2020 to avoid exactly the scenario playing out now. Jointly run by WHO, the Coalition for Epidemic Preparedness Innovations (CEPI), and Gavi, the Vaccine Alliance, the organization aimed to bring together countries to invest in several vaccine candidates that it would then distribute equitably among participants. High-income countries, companies, and philanthropic organizations would foot the bill for the 92 poorest countries.

At first it appeared to be working: Almost every country in the world has signed up and COVAX began to deliver its first vaccine doses on 24 February, just 2 months after vaccinations started in Europe. But COVAX lacked the money to compete with rich countries that cornered the market early on by striking purchase deals with vaccinemakers. "It was slower than anyone would have wanted in making deals," says Nicole Lurie, U.S. director of CEPI. "That's been a real frustration."

Another major blow came in March, when skyrocketing COVID-19 cases at home led the Serum Institute of India—which COVAX banked on as its main supplier—to halt exports of its vaccine, made in collaboration with AstraZeneca and the University of Oxford. Exports may not resume until the end of the year (see p. 896). "We're in a bit of a crisis," says Seth Berkley, who heads GAVI. COVAX will be 190 million doses short by the end of June. "We're now trying to fill that hole," he says. As of 23 May, 125 countries had received just 68 million vaccine doses from COVAX.

Serum could resume shipments earlier than December, Berkley says, but if it doesn't, COVAX could be short 1 billion doses by the end of this year. "What we thought COVAX would prevent is happening," acknowledges WHO Chief Scientist Soumya Swaminathan.

Vaccine donations from rich countries could help. Because it was unclear in 2020 which vaccines might work, many countries made major deals with multiple companies; the United States and the European Union advance purchased enough to vaccinate their populations three times over, for example. Although not all of those doses have arrived, the United States already has a surplus, and Europe should have enough vaccine for all its residents this summer.

Wealthy countries should donate 1 billion doses to COVAX by 1 September and another billion by mid-2022, contends a report published in early May by the Independent Panel for Pandemic Preparedness and Response (IPPPR). Mark Dybul, former head of the U.S. government's President's Emergency Plan for AIDS Relief and a member of IPPPR, says this is urgent. "We need to really press and accelerate and get all the excess surplus out as rapidly as possible," he says.

Many rich countries, concerned about viral variants and new outbreaks, are hesitant to give up vaccines they might still need, Berkley says: "There's a lot of unknowns at the moment that make people nervous." But donations have already begun. Norway and New Zealand have both given the vaccine they were eligible to receive through COVAX back to the facility. Some European countries

ILLUSTRATION: STEPHAN SCHMITZ/FOLIOART

are ready to donate stocks of AstraZeneca-Oxford shots, whose use there was recently restricted to older age groups because they can cause a rare but serious clotting disorder. The biggest pledge so far comes from the Biden administration, which aims to donate 80 million doses by the end of June, including 60 million doses of the AstraZeneca-Oxford vaccine.

There's also hope that COVAX may add more vaccines to its arsenal. It is relying heavily on AstraZeneca and its partners because their adenovirus-based vaccine is cheap and, unlike messenger RNA (mRNA) vaccines, can be transported in regular refrigerators. Still, Pfizer and its partner BioNTech in January agreed to sell COVAX up to 40 million doses of its mRNA vaccine this year. On 3 May, Moderna followed with a deal to sell 34 million doses of its mRNA shot this year and 444 million in 2022. Both companies said they will give COVAX a discount, but didn't reveal their price.

At the World Health Assembly, Tedros made a bold call for all manfacturers to offer any new product to COVAX before putting it on the market, or to commit 50% of their doses to the facility.

MONTHS

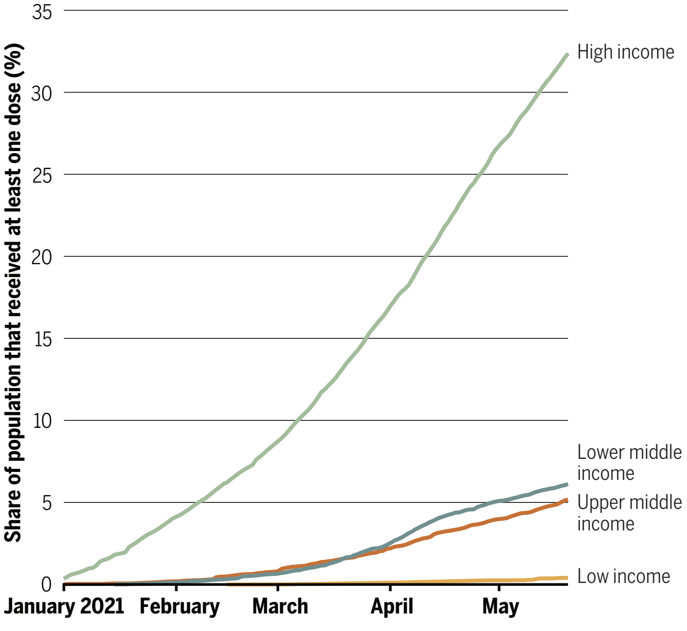
Expanding production

To date, all manufacturers combined have distributed fewer than 2 billion doses of COVID-19 vaccines, most of which require two shots. But much more vaccine is on its way. Many of the 14 manufacturers of authorized products continue to build new plants, contract with other manufacturers, and iron out production glitches.

Pfizer and BioNTech hope to produce about 3 billion doses by year's end, one-third of which they plan to offer to COVAX or directly to low- and middle-income countries. Moderna has ramped up to 1 billion. Three billion more doses could come from the conglomerate organized by AstraZeneca and Oxford, which includes Serum and several other vaccinemakers. Three China-based companies say they can collectively pump out 3 billion doses of their vaccines this year, and Johnson & Johnson hopes to add another 1 billion doses of its single-shot vaccine. Some 850 million doses of the Russian Sputnik V vaccine could come from the Gamaleya Research Institute of Epidemiology and Microbiology and its contract manufacturers. And companies such as Novavax, CureVac, and Clover Biopharmaceuticals all have vaccines in efficacy trials and are hoping for market authorizations in the months ahead.

Dose disparity

High-income countries are rapidly immunizing their populations against COVID-19. Middle-income and low-income countries have received far fewer doses to date.



CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) OUR WORLD IN DATA

All told, 14 billion doses could leave factories before the year is over, according to a document written in advance of a March summit about stepping up the production pace. That's a startling number—before the pandemic, all of the world's vaccinemakers together produced at most 5.5 billion doses annually—and the document stresses that it's the best of all scenarios. "We have a set of manufacturing challenges that make it unlikely that we can meet the 14-billion aspiration," cautions Lurie, who helped organize the meeting.

Raw materials, such as disposable bags that line bioreactors, filters, and cell-culture media, are the biggest challenge. "Raw material inputs have nowhere near kept up with the anticipated demand," Lurie says. Importing and exporting delays have exacerbated the shortages, and travel bans have made it difficult to move experts around the world to troubleshoot manufacturing snafus. Almost every vaccine producer has failed to deliver on initial promises.

New viral variants that evade vaccine-induced immunity could limit the vaccines' usefulness, as could safety problems. The AstraZeneca-Oxford shot fell out of favor in South Africa because it failed against a variant, and the European Union plans to phase out the vaccine because of the rare clotting disorder. China's Sinopharm and Sinovac Biotech vaccines have far lower efficacy in some studies than other products and may require a third shot after 6 months.

Demand for several vaccines could also drop because mRNA vaccines work so well and are easy to modify for new variants. Withholding them from large parts of the world may come to be seen as an injustice, says viral immunologist Lawrence Corey of the Fred Hutchinson Cancer Research Center. "We're going to be judged in our humanity by how we utilize the technologies," he says.

MONTHS TO YEARS

Sharing knowledge

Allowing more companies to follow the vaccine recipes developed—and sometimes fiercely protected—by a few could also boost output. AstraZeneca and Oxford provided a license for their vaccine to Serum, and then helped the company learn to manufacture it in India, an intensive process known as technology transfer. Most vaccine companies have shied away from such deals, however.

Several ideas have been floated to change that. In May 2020, WHO launched the COVID-19 Technology Access Pool (C-TAP), a system under which companies with proven products would voluntarily share their know-how and intellectual property (IP). The hope was that new manufacturers could pluck whatever they needed from C-TAP, either at a reasonable fee or for free, to start to make their own products. But C-TAP has yet to attract any participants because, critics say, companies want to hold on to monopolies and see sharing as a threat to profits. That has amplified calls for more drastic steps, including taking away the cornerstone of their IP: the patent.

In October 2020, India and South Africa asked the World Trade Organization (WTO) to issue a broad-reaching waiver for patents and other intellectual properties that pertain to "prevention, containment or treatment of COVID-19." The idea has made little headway, although proponents were heartened on 5 May when U.S. Trade Representative Katherine Tai issued a statement explicitly supporting a WTO waiver for COVID-19 vaccines. "This is a global health crisis, and the extraordinary circumstances of the COVID-19 pandemic call for extraordinary measures," Tai said.

The pharmaceutical industry decried the statement, arguing that giving away IP would remove the incentive to innovate. Inexperienced new vaccine manufacturers would also drain the limited supply of raw materials, Pfizer CEO Albert Bourla warned in a recent letter, "putting the safety and security of all at risk." Moderna CEO Stéphane Bancel said waiving patents is "the wrong question," because new companies would not be able to produce mRNA vaccines this year or next, "the most critical time of the pandemic." Although Moderna has promised not to enforce its own COVID-19–related IP rights during the pandemic, making its vaccine still requires licensing deals with other patent holders—including a critical one held by the University of Pennsylvania—and, most important, know-how.

Martin Friede, who helps coordinate vaccine research at WHO, applauds Moderna for waiving its patents but agrees that "the patent by itself is useless." New companies would still need to set up facilities, train staff, and learn how to produce the vaccines. A patent is a recipe but doesn't make you a chef, Friede says: "I've got a lot of cookbooks in my kitchen, but you wouldn't want to come to my house if I'm doing the cooking." He and his colleagues are trying to create a training center where scientists and manufacturers from low- and middle-income countries could learn how to set up the industrial process for mRNA vaccines. "You go back to your facility, and now instead of spending years setting it up, you're up and running in a matter of months," he says.

A shipment of the AstraZeneca-Oxford COVID-19 vaccine arrives in Accra, Ghana, on 7 May. COVAX, which supplied the doses, has had a slow start.

PHOTO: REUTERS/FRANCIS KOKOROKO

James Love, who runs the nongovernmental organization Knowledge Ecology International, says waiving patents would kick-start the process for many would-be vaccinemakers. "It's been illegal for a year and a half to even work on these vaccines," Love says. "Who's going to invest in developing something where it's illegal?"

Others say the threat of patent waivers can act as a cudgel, compelling big pharmaceutical companies to share their patents and expertise. IPPPR wrote in its report to WHO that IP waivers should be implemented if companies fail to negotiate voluntary licenses and technology transfers within 3 months.

That has worked in the past, Dybul says: Lifesaving antiretroviral drugs developed in the 1990s long remained too expensive for most HIV-infected people. But the threat that their patents would be steamrolled led Big Pharma to share its knowledge with generic drugmakers, and prices plummeted. To Dybul, the vaccinemakers' reaction today echoes what drug companies said back then: "Just let us produce, we know how to do it, it's too complicated, can't be done." But he acknowledges that vaccines are more difficult to produce than drugs.

Ideally, Friede says he'd like Pfizer, BioNTech, and Moderna to "play ball with us" and help set up the training facility, but he's soliciting help from other makers of mRNA vaccines who have their own patent portfolios but don't yet have regulatory authorization.

YEARS

Building plants worldwide

Freeing up IP could allow existing plants to produce vaccines that are now off-limits. But in the long term, additional manufacturing plants will be needed to serve the needs of the have-nots—not just for this pandemic, but for future ones as well. COVID-19 variants and waning immunity to the virus each could create an annual need for several billion vaccine doses. And if a new disease surfaces, the world could find itself again needing billions of doses of new vaccines.

The inequitable way COVID-19 vaccines have been rolled out underscores the limits of a system that concentrates manufacturing power in a few locations, says Leena Menghaney, who heads Doctors Without Borders's Access Campaign for South-East Asia. The bulk of COVID-19 vaccine doses so far have been made in the United States, China, India, and Europe—and all have felt pressured to vaccinate their own people first. "Once we're done with ours we'll give it to you: That argument is very much in your face and it's the subtext of the whole conversation," Menghaney says.

Jeremy Farrar, head of the Wellcome Trust, suggests establishing global enterprise zones that produce vaccines in countries with small populations, such as Senegal, Singapore, Costa Rica, or Rwanda. "They could provide vaccine for their own domestic citizens very quickly and could then be manufacturing hubs for the world," he says. And he advocates that countries get together and make "at-risk" investments—gambles that may not pay off—in these new plants.

Rwandan President Paul Kagame said at a 13 April African Union conference that his country has already discussed construction of an mRNA plant with several manufacturers. One of them, GreenLight BioSciences, has issued a "blueprint of how to vaccinate the world" with small, modular mRNA plants that it can build in Massachusetts and ship anywhere. A single production room, roughly the size of four shipping containers, could make most of its own raw materials and crank out 17 million doses per month. For about $200 million the company says it could provide a mini–vaccine plant and a "clean room" to house it. The company says one of these could be up and running as little as 1 year after an agreement is signed. "You could leapfrog our ability to fight pandemics," Dybul says.

Mark Feinberg, who heads the International Aids Vaccine Initiative, likes the idea, but says it faces many hurdles. Large vaccinemakers have economies of scale, reliable supply chains for raw materials, stringent quality control, and regular inspections, Feinberg says. "Rwanda is an amazing place in terms of public health and commitment to innovation," he says, but "it's not like you snap your fingers and all of a sudden there are highly capable people who know every detail about the manufacturing process and can ensure that it's operating as it should." It's also unclear what vaccine plants and staff would do between pandemics.

A year and a half into the pandemic, COVID-19 has shined a bright light on hugely complicated issues. "The challenge that's being taken on here is unprecedented: You want to vaccinate everyone on the planet. That has never been done before," Feinberg says. "You might say: If we do this in 3 years that's an accomplishment that is unprecedented in the history of the human race. But if you actually think about the tragedy that COVID is imposing, that seems like a really long, long time." ■

INSIGHTS

PERSPECTIVES

Healthy coral reefs without too many sea urchins and seaweeds, like this one in Papua New Guinea, suffer less mortality during ocean heatwaves.

PHOTO: TOM SHLESINGER

MARINE ECOLOGY

Local management matters for coral reefs

Impacts of marine heatwaves are worse when seaweeds and sea urchins are abundant

By **Nancy Knowlton**

The ability of corals to build reefs depends on a nutritional symbiosis between the coral animal and intracellular, single-celled microalgae. Coral bleaching is the visual manifestation of a breakdown in this relationship; it is a response to stress, including temperatures 1° to 2°C above normal maxima. Global warming has resulted in sharp increases in the frequency and magnitude of bleaching events (*1*), which have already caused enormous damage to reefs worldwide. However, the importance of other factors in aggravating the effects of high temperatures has been disputed (*2*). On page 977 of this issue, Donovan *et al.* (*3*) show that the amount of coral loss 1 year after bleaching is highly correlated globally with other aspects of reef health, specifically the abundance of macroalgae and sea urchins. This suggests that local management can help to ameliorate the impacts of marine heatwaves.

When high-temperature stress is severe, many corals die quickly even on healthy reefs far from human impacts (*4*). Mortality was catastrophic on remote northern regions of Australia's Great Barrier Reef during the major bleaching event of 2015 to 2016, and local management had little if any effect on short-term outcomes (*5*). However, the potential for environmental conditions to shape patterns of coral survivorship during heatwaves has not been extensively studied in detail.

After temperatures return to normal levels [and even sometimes before (*6*)], surviving corals can regain their symbionts, and reefs can slowly recover through the growth of these survivors and the establishment of new coral recruits (*4*). Recovery does not always occur, however, because corals weakened by the stress of bleaching may succumb to other factors, such as disease, or recruitment may fail. In St. Croix in the US Virgin Islands, for example, a major bleaching event in 2005 was followed by a disease outbreak that caused a 60% decline in live coral cover (*7*).

Given that temperatures will continue to increase for the foreseeable future, it is essential to know whether local management could improve reef prospects. Because poor water quality and overfishing are known to have killed many corals before bleaching became common (*8*, *9*), it is widely accepted that reef recovery after bleaching could be improved by facilitating recruitment and regrowth; studies of the recovery of remote or well-managed reefs after bleaching (*4*, *10*) support this idea. Unfortunately, however, according to the data from the Great Barrier Reef (*5*), the consensus has been that little could be done through management to reduce initial mortality from bleaching.

Recent data from a few locations in the Pacific (*6*, *11*) as well as an earlier assessment from the Caribbean (*8*) suggest that this consensus might be too pessimistic. At Kiritimati Atoll, corals that acquired heat-tolerant symbionts after bleaching survived at higher rates, but this only occurred where anthropogenic stress was low (*6*). In Moorea, French Polynesia, higher nitrogen concentrations were associated with a doubling of bleaching severity at low levels of temperature stress (*11*). What Donovan *et al.* have done is to greatly expand confidence in the hypothesis that local management can make a difference, by performing a global analysis of the environmental factors that increase postbleaching mortality.

Their study, based on 223 reefs from the Caribbean and Indo-Pacific, documents substantially higher coral loss in the year after bleaching on reefs with high abundances of macroalgae and sea urchins, which are typically associated with overfishing and nutrient pollution. Because 1 year is likely too short a time to detect recovery from regrowth and recruitment (*4*), the higher loss rates must largely reflect mortality either during or shortly after the bleaching event. The effects described are highly important ecologically. For example, at some levels of heat stress, reefs with more macroalgae experience a 10-fold increase in mortality. In contrast to the earlier documentation of the effect of nitrogen on bleaching severity (*11*), the strength of the negative effect of macroalgae increases with the severity of the bleaching event.

The data used in this study, from the Reef Check database, come from relatively simple reef surveys conducted by community and professional scientists and thus do not address the mechanisms underpinning these correlations. However, as the authors note, macroalgae are known to be detrimental to corals in a number of ways, and a variety of mechanisms could be responsible for the patterns observed. The relationship with sea urchin abundance is somewhat more surprising, because urchins, particularly in the Caribbean, are known to protect corals from overgrowth by macroalgae at moderate densities; hence, this probably reflects the negative impacts of extremely high amounts of grazing associated with urchin "barrens." Better understanding of the mechanisms underpinning these and other patterns reported in this study will help to refine management approaches during the coming decade, when many reefs will continue to struggle.

Despite the doom and gloom of media reports on the state of the ocean, and the enormous challenges that remain, there is growing recognition that marine conservation actions have had measurable success (*12*, *13*). Indeed, local actions can not only minimize damage from warming, but provide biodiversity and food-security benefits as well (*12*, *14*).

This does not mean that taking the appropriate steps to, for example, reduce macroalgae and sea urchin abundance is easy in practice. Genuine stakeholder engagement is essential for conservation success (*15*); this is not simply a matter of resources, because establishing the required trust among stakeholders takes time and effort. The urgent need to slow and reverse climate change to save reefs from ecological extinction is also clear. During upcoming global negotiations, governments should remember that in addition to setting ambitious targets for lowering greenhouse gas emissions, empowering local communities to manage reef (and other) marine resources is an important strategy to reduce the negative impacts of climate change. ■

REFERENCES AND NOTES

1. T. P. Hughes *et al.*, *Science* **359**, 80 (2018).
2. A. Abelson, *ICES J. Mar. Sci.* **77**, 40 (2020).
3. M. K. Donovan *et al.*, *Science* **372**, 977 (2021).
4. J. P. Gilmour *et al.*, *Science* **340**, 69 (2013).
5. T. P. Hughes *et al.*, *Nature* **543**, 373 (2017).
6. D. C. Claar *et al.*, *Nat. Commun.* **11**, 6097 (2020).
7. J. Miller *et al.*, *Coral Reefs* **28**, 925 (2009).
8. J. B. C. Jackson, M. K. Donovan, K. L. Cramer, V. V. Lam, eds., *Status and Trends of Caribbean Coral Reefs: 1970-2012* [International Union for Conservation of Nature (IUCN), 2014].
9. K. L. Cramer *et al.*, *Sci. Adv.* **6**, eaax9395 (2020).
10. R. S. Steneck *et al.*, *Front. Mar. Sci.* **6**, 265 (2019).
11. M. K. Donovan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5351 (2020).
12. N. Knowlton, *Annu. Rev. Mar. Sci.* **13**, 479 (2021).
13. C. M. Duarte *et al.*, *Nature* **580**, 39 (2020).
14. E. Sala *et al.*, *Nature* **592**, 397 (2021).
15. S. Giakoumi *et al.*, *Front. Mar. Sci.* **5**, 223 (2018).

Smithsonian Institution, Washington, DC 20560, USA.
Email: knowlton@gmail.com

10.1126/science.abi7286

BIOENGINEERING

Extracting electricity with exosuit braking

An exosuit lets wearers tense their muscles less and save energy in portions of their stride

By **Raziel Riemer**[1], **Richard W. Nuckols**[2], **Gregory S. Sawicki**[3,4,5]

Exoskeletons and exosuits are wearable devices designed to work alongside the musculoskeletal system and reduce the effort needed to walk or run. Exoskeletons can benefit users by reducing the mechanical power and metabolic energy that they need to move about on the factory floor, in the rehabilitation clinic, on the playing field, and out at the shopping mall (*1*). Portable exoskeletons can use motors to add mechanical power into movement phases [net-positive exoskeleton power (*2*, *3*)] or use springs to store and later return mechanical energy in a regenerative braking action [net-zero exoskeleton power (*4*, *5*)]. On page 957 of this issue, Shepertycky *et al.* (*6*) describe a wearable assistive device that uses a generator to extract mechanical energy from the walking cycle (net-negative power) and convert it to electricity. At the same time, the walker actually uses less metabolic energy with the exosuit, saving on the cost to operate muscles as "biological brakes."

Handgrip and pedal-powered dynamos have long been in use and can convert mechanical power to electrical power, and these devices can have efficiencies as high as 70% (*7*). More recently, "hands-free" energy harvesters have been developed that can be worn on the back (*8*) or attached with an exoskeletal structure around the lower-limb joints (*9*–*11*). A performance metric for these devices is the cost of harvesting

---

[1] Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [2] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [3] School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [4] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. [5] Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA, USA. Email: gregory.sawicki@me.gatech.edu

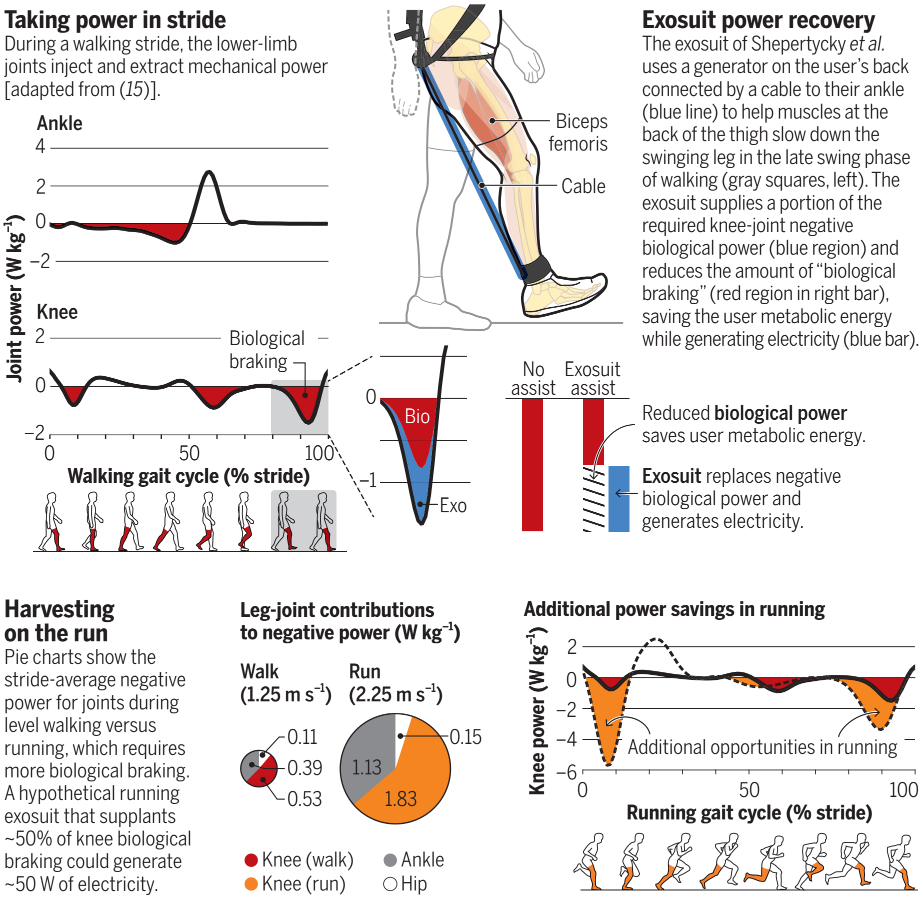
(COH), which is the ratio of the change in a user's metabolic power (measured in watts) when moving with versus without the device to the electrical power generated by the device. A positive COH means that the user must provide additional metabolic effort to generate electricity. For the examples above, the reported COH values have ranged from 4.8 for the back-mounted device (*8*) to 0.7 for a knee-joint mount (*9*).

This latter device developed by Donelan *et al.* (*9*) incorporated principles from fundamental movement biomechanics to strategically target phases of human walking where the lower-limb joints already resist motion (negative mechanical power) and behave effectively as brakes. Biomechanical analyses combining data from high-speed motion capture and instrumented force platforms with inverse-dynamics calculations reveal that the knee joint acts mostly like a brake during walking, especially at the end of the swing phase, when the foot is in the air (see the figure, top left). Muscles convert metabolic power to mechanical power with 25% efficiency when acting as motors (positive mechanical power output) and –125% efficiency when acting as brakes (negative mechanical power output) (*12*).

Donelan *et al.* designed a knee exoskeleton in which a rotary generator attached in parallel with the human knee worked to help off-load biological braking. The resistance of the generator to turning provided the braking torque. With this device, they established that by targeting phases of negative mechanical power, exoskeletons can generate electricity with minimal increase in user effort. If muscles had acted as motors to provide the 1.7 W of mechanical power needed to generate each 1 W of electricity (their device had a 60% conversion efficiency), then users would have had to expend 6.8 W more metabolic power. However, for each 1 W of generated electricity, users only expended 0.7 W of metabolic energy (COH = 0.7). Although this system still required additional user effort, the results suggested that energy can be harvested from gait while at the same time saving metabolic energy—a negative COH.

Charging ahead by braking

Shepertycky *et al*. developed an exosuit that reduces the metabolic energy needed by muscles to resist motion during gait. A generator provides the "braking" force and produces electricity.



Taking power in stride

During a walking stride, the lower-limb joints inject and extract mechanical power [adapted from (*15*)].

Exosuit power recovery

The exosuit of Shepertycky *et al.* uses a generator on the user's back connected by a cable to their ankle (blue line) to help muscles at the back of the thigh slow down the swinging leg in the late swing phase of walking (gray squares, left). The exosuit supplies a portion of the required knee-joint negative biological power (blue region) and reduces the amount of "biological braking" (red region in right bar), saving the user metabolic energy while generating electricity (blue bar).

Reduced **biological power** saves user metabolic energy.

Exosuit replaces negative biological power and generates electricity.

Harvesting on the run

Pie charts show the stride-average negative power for joints during level walking versus running, which requires more biological braking. A hypothetical running exosuit that supplants ~50% of knee biological braking could generate ~50 W of electricity.

Leg-joint contributions to negative power (W kg^{-1})

Additional power savings in running

Knee power (W kg^{-1})
2
0
−2
−4
−6
0
50
100
Additional opportunities in running
Running gait cycle (% stride)

This result highlights a key difference between skeletal muscle and engineered systems, namely, that braking is energetically cheap for machines (like a bicycle hand brake) but expensive for muscles, which have to consume metabolic energy to tense up and maintain braking force, especially when changing length (*12*). Thus, properly timed exoskeleton resistance could provide a portion of the negative muscle power that is normally lost as heat. Rather than requiring additional user effort to perform positive mechanical work on the exoskeleton generator, exoskeleton negative power would save the user the metabolic energy needed for muscle braking (*13*).

Shepertycky *et al.* designed a streamlined exosuit with a negative COH using a feedback-controlled "muscle-centric" loading profile. They specifically targeted the period during very late leg swing (just before the foot makes contact with the ground) when large braking forces are produced by actively lengthening hamstring muscles (for example, biceps femoris), rather than metabolically inactive passive elastic structures (for example, tendons and ligaments) (*14*). Their "traditional" loading profile (*10*) extracted the same total mechanical energy but resulted in a 3% metabolic penalty. The relatively subtle shift in timing and magnitude of the "muscle-centric" profile resulted in a 2.5% net metabolic benefit—a 5.5% improvement.

By strategically placing the device on the user's back, Shepertycky *et al.* were also able to reduce the carrying cost of their exosuit to just over 1%. This penalty is meager compared with the nearly 20% metabolic increase that was imposed by bulky knee-mounted exoskeletons that weighed 1.65 kg per leg (*9*). Their 1.1-kg device hardware rested at the waist near the user's center of mass. Exosuit support was supplied by tensioning cables that were routed along the posterior thigh and shank. The other ends of these cables were ultimately attached at the ankle to apply forces parallel to the hamstrings (see the figure, top middle).

Shepertycky *et al.*'s energy-extracting exosuit, which achieves a net 2.5% reduction in the metabolic cost of walking along with 0.25 W of generated electricity, may only be the first of many such devices that could achieve a negative COH. Rough calculations based on engineering specifications for generators (*7*), locomotion biomechanics data (*15*), and fundamental muscle physiology relationships (*12*) suggest many opportunities to extend the principle of "resistive assistance" (see the figure, bottom). Targets include lower-limb joints other than the knee, gait phases other than terminal swing, and locomotion tasks other than walking on

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

level ground. More intense gaits like running, where the legs cycle more positive and negative mechanical power, and tasks like walking downhill, descending staircases, or decelerating to a stop all provide increased opportunities for rigid exoskeletons or soft exosuits to assist the body's biological brakes while generating electricity.

The next-generation exosuits will begin to integrate physiological sensing systems and machine-learning algorithms to increase the versatility and impact of wearable assistive devices. During the next decade, a new challenge may be the development of an exosuit that minimizes human metabolic energy expenditure on a round-trip course spanning many kilometers over many days with access to a single onboard rechargeable battery. Optimal performance will likely require multijoint, hybrid support strategies that combine injection, extraction, and transfer of both electrical and mechanical energy to adapt continuously to locomotion-task demands and reduce metabolic energy expenditure of the user.

Such devices could have several applications, such as extending the range of on-foot search-and-rescue crews, outdoor adventurers, or soldiers on humanitarian missions. In the developing world, an exosuit could provide between 20 and 40% of the electricity needed per person on a typical day. The energy demands of portable electronics and increased recognition of the role of movement in longevity may drive exosuits toward widespread adoption. ■

REFERENCES AND NOTES

1. G. S. Sawicki, O. N. Beck, I. Kang, A. J. Young, *J. Neuroeng. Rehabil.* **17**, 25 (2020).
2. L. M. Mooney, E. J. Rouse, H. M. Herr, *J. Neuroeng. Rehabil.* **11**, 151 (2014).
3. J. Kim *et al.*, *Science* **365**, 668 (2019).
4. S. H. Collins, M. B. Wiggin, G. S. Sawicki, *Nature* **522**, 212 (2015).
5. R. Nasiri, A. Ahmadi, M. N. Ahmadabadi, *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 2026 (2018).
6. M. Shepertycky, S. Burton, A. Dickson, Y.-F. Liu, Q. Li, *Science* **372**, 957 (2021).
7. E. Schertzer, R. Riemer, *J. Neuroeng. Rehabil.* **12**, 30 (2015).
8. L. C. Rome, L. Flynn, E. M. Goldman, T. D. Yoo, *Science* **309**, 1725 (2005).
9. J. M. Donelan *et al.*, *Science* **319**, 807 (2008).
10. M. Shepertycky, Q. Li, *PLOS ONE* **10**, e0127635 (2015).
11. L. Xie, X. Li, S. Cai, G. Huang, L. Huang, *Mech. Syst. Signal Process.* **127**, 172 (2019).
12. R. Margaria, *Int. Z. Angew. Physiol. Einschl. Arbeitsphysiol.* **25**, 339 (1968).
13. P. Niu, P. Chapman, R. Riemer, X. Zhang, in *Proceedings of the 2004 IEEE 35th Annual Power Electronics Specialists Conference* (IEEE, 2004), vol. 3, pp. 2100–2106.
14. B. Whittington, A. Silder, B. Heiderscheit, D. G. Thelen, *Gait Posture* **27**, 628 (2008).
15. R. W. Nuckols *et al.*, *PLOS ONE* **15**, e0231996 (2020).

ACKNOWLEDGMENTS

R.R., R.W.N., and G.S.S. contributed equally to this work.

10.1126/science.abh4007

HYPOTHESIS

Making the hard problem of consciousness easier

Championing open science, an adversarial collaboration aims to unravel the footprints of consciousness

By **Lucia Melloni**[1,2], **Liad Mudrik**[3], **Michael Pitts**[4], **Christof Koch**[5,6]

The history of science includes numerous challenging problems, including the "hard problem" (*1*) of consciousness: Why does an assembly of neurons—no matter how complex, such as the human brain—give rise to perceptions and feelings that are consciously experienced, such as the sweetness of chocolate or the tenderness of a loving caress on one's cheek? Beyond satisfying this millennia-old existential curiosity, understanding consciousness bears substantial medical and ethical implications, from evaluating whether someone is conscious after brain injury to determining whether nonhuman animals, fetuses, cell organoids, or even advanced machines (*2*) are conscious. A comprehensive and agreed-upon theory of consciousness is necessary to answer the question of which systems—biologically evolved or artificially designed—experience anything and to define the ethical boundaries of our actions toward them. The research projects described here will hopefully point the way and indicate whether some of today's major theories hold water or not.

After prosperous decades of focused scientific investigation zeroing in on the neural correlates of consciousness (*3*), a number of candidate theories of consciousness have emerged. These have independently gained substantial empirical support (*4*–*7*), led to empirically testable predictions, and resulted in major improvements in the evaluation of consciousness at the bedside (*8*, *9*). Notwithstanding this progress, the conjectures being put forward by the different theories make diverging claims and predictions that cannot all be simultaneously true. Moreover, the theories evolve and continue to adapt as further data accumulates, with hardly any cross-talk between them. How can we then narrow down on which theory better explains conscious experience?

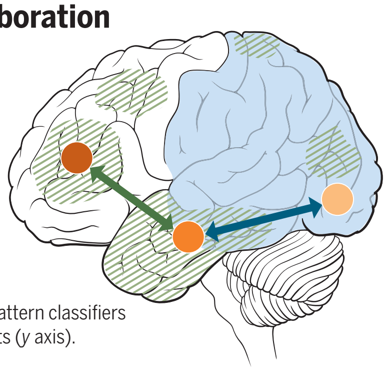
The road to a possible solution may be paved by means of a new form of cooperation among scientific adversaries. Championed by Daniel Kahneman in the field of behavioral economics (*10*) and predated by Arthur Eddington's observational study to test Einstein's theory of general relativity against Newton's theory of gravitation (*11*), adversarial collaboration rests on identifying the most diagnostic points of divergence between competing theories, reaching agreement on precisely what they predict, and then designing experiments that directly test those diverging predictions. During the past 2 years, several groups have adopted this approach, following an initiative that aims to accelerate research in consciousness. So far, several theories of consciousness are being evaluated in this manner to test competing explanations for where and when neural activity gives rise to subjective experience.

The global neuronal workspace theory (GNWT) (*4*) claims that consciousness is instantiated by the global broadcasting and amplification of information across an interconnected network of prefrontal-parietal areas and many high-level sensory cortical areas. The sensory areas carry out different functions that range from feature processing to object or word recognition. Information in those sensory areas is processed in encapsulated modules, remaining unconscious. The frontal-parietal networks support integrative and executive functions, including selective attention and working memory. According to the GNWT, a stimulus must be attended to trigger activity that helps distribute this sensory information to many parts of the brain for further processing and report. It is this global broadcasting across many modules of specialized subsystems that constitutes consciousness. Conversely, the integrated information theory (IIT) (*5*) holds that consciousness should be understood in terms of cause-effect "power" that reflects the amount of maximally irreducible integrated information generated by certain neuronal architectures. On the basis of mathematical

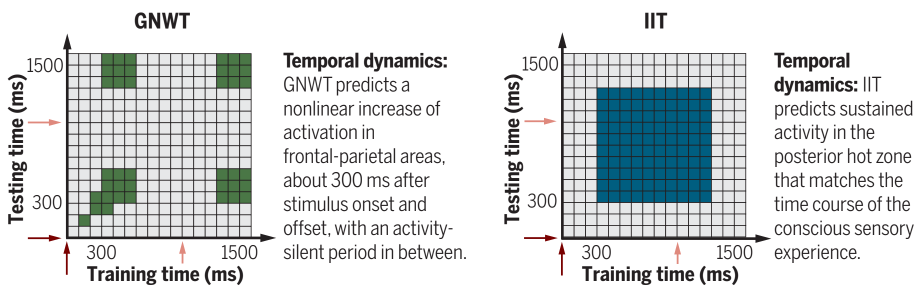
[1] Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany. [2] Department of Neurology, New York University Grossman School of Medicine, New York, NY, USA. [3] School of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel. [4] Department of Psychology, Reed College, Portland, OR, USA. [5] Allen Institute for Brain Science, Seattle, WA, USA. [6] Tiny Blue Dot Foundation, Santa Monica, CA, USA. Email christofk@alleninstitute.org

Testing hypotheses by adversarial collaboration

The neural correlates of consciousness for the global neuronal workspace theory (GNWT) and for the integrated information theory (IIT) occupy distinct and overlapping regions in the brain. Each theory predicts synchronization of activity between or within these regions.



The **time-generalized decoding matrices** depict performance of pattern classifiers trained on specific time points (*x* axis) and tested on other time points (*y* axis).



and neuroanatomical considerations, the IIT holds that the posterior cortex is ideally situated for generating a maximum of integrated information. In this theory, consciousness is not input-output information processing but the intrinsic ability or power of a neuronal network to influence itself. That is, the neuronal substrate of consciousness perpetuates itself for as long as the experience exists. The more cause-effect power a system has, the more conscious it is. For the IIT, the content of an experience is a structure of causes and effects (integrated information), whereas for the GNWT, it is a message that is broadcast globally.

Another controversy occurs between first-order (*12*, *13*) and higher-order (*6*, *14*) theories of consciousness. The former claims that reverberating activity in sensory areas suffices for consciousness, whereas the latter claims that a second, higher-order brain state must represent or "point at" these first-order sensory activations for them to be consciously experienced.

Both controversies are the types of theoretical disagreements that are currently being empirically tested by use of the adversarial collaboration approach. One of these collaborations, the COGITATE consortium (Collaboration On GNW and IIT: Testing Alternative Theories of Experience), is collecting data and has recently released a detailed preregistered report that outlines the methods, predictions, and planned analyses (https://osf.io/mbcfy). These experiments were designed by neuroscientists and philosophers who are not directly associated with the theories but are in close collaboration with advocates from each theory. The experiments are being conducted in six independent laboratories. Briefly, one of the experimental designs involves an engaging video game with seen and unseen stimuli in the background to determine whether neural correlates of the visual experience are present irrespective of the task. In another experiment, stimuli are shown for variable durations to investigate for how long the neural correlate of the visual experience exists. Neuronal activity in human subjects is measured with both invasive and noninvasive methodologies, from functional magnetic resonance imaging and simultaneous magnetoencephalography and electroencephalography to invasive electrocorticography, and is integrated across methodologies to test the theories' predictions. These focus on two key questions: Where are the anatomical footprints of consciousness in the brain: Are they located in a posterior cortical "hot zone" (*15*) advocated by the IIT, or is the prefrontal cortex necessary (*4*) as predicted by the GNWT? And, how are conscious percepts maintained over time: Is the underlying neural state maintained as long as the conscious experience lasts, in line with the IIT, or is the system initially ignited and then decays and remains silent until a new ignition marks the onset of a new percept, as the GNWT holds (see the figure)? Once the brain data are collected and analyzed, they will be made available to anyone. Relying on adversarial dialogue and collaboration, open science practices, standardized protocols, internal replication, and team science, these initiatives aim to promote empirical progress in the field of consciousness and to change the sociology of scientific practice in general.

Solving big questions may require "big science" because such questions are more likely to be solved in unison rather than through isolated, parallel, small-scale attempts. The adversarial collaboration approach builds on the success of large-scale collaborative institutes (such as the Allen Institute for Brain Science) and projects such as the Human Connectome Project or the International Brain Laboratory in neuroscience, which were preceded by initiatives in physics such as the Large Hadron Collider at the European Organization for Nuclear Research (CERN) or the Laser Interferometer Gravitational-Wave Observatory (LIGO) experiment. With this series of adversarial collaborations, neuroscientists will get closer to understanding consciousness and how it fits into the physical world while improving scientific practices along the way. As for the initial theories undergoing this approach, it may be that neither the GNWT nor the IIT are quite correct. No matter the outcome, the field can use the results to make progress in framing new thinking about consciousness and testing other potential theories in the same way. The problem of consciousness will surely remain difficult, but understanding the ancient mind-body problem will become a little bit easier. ■

REFERENCES AND NOTES

1. D. J. Chalmers, *J. Conscious. Stud.* **2**, 200 (1995).
2. T. Bayne *et al.*, *Trends Neurosci.* **43**, 6 (2020).
3. F. Crick, C. Koch, *Nat. Neurosci.* **6**, 119 (2003).
4. G. A. Mashour, P. Roelfsema, J. P. Changeux, S. Dehaene, *Neuron* **105**, 776 (2020).
5. G. Tononi, M. Boly, M. Massimini, C. Koch, *Nat. Rev. Neurosci.* **17**, 450 (2016).
6. R. Brown *et al.*, *Trends Cogn. Sci.* **23**, 754 (2019).
7. V. A. F. Lamme, *Cogn. Neurosci.* **1**, 204 (2010).
8. A. Demertzi *et al.*, *Sci. Adv.* **5**, eaat7603 (2019).
9. A. G. Casali *et al.*, *Sci. Transl. Med.* **5**, 198ra105 (2013).
10. D. Kahneman, *Am. Psychol.* **58**, 723 (2003).
11. F. W. Dyson, A. S. Eddington, C. Davidson, *Philos. Trans. R. Soc. A.* **220**, 291 (1920).
12. V. A. F. Lamme, P. R. Roelfsema, *Trends Neurosci.* **23**, 571 (2000).
13. N. Block, *Trends Cogn. Sci.* **9**, 46 (2005).
14. H. Lau, D. Rosenthal, *Trends Cogn. Sci.* **15**, 365 (2011).
15. C. Koch, M. Massimini, M. Boly, G. Tononi, *Nat. Rev. Neurosci.* **17**, 666 (2016).

ACKNOWLEDGMENTS

COGITATE is supported by a grant from the Templeton World Charity Foundation (TWCF) (www.templetonworldcharity.org/accelerating-research-consciousness-our-structured-adversarial-collaboration-projects). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of TWCF. L.M. is a Canadian Institute for Advanced Research Tanenbaum Fellow in the Brain, Mind, and Consciousness program. C.K. thanks the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support. The authors thank D. Potgieter for championing the adversarial collaboration concept and acknowledge the COGITATE consortium: K. Bentz, H. Blumenfeld, D. Chalmers, F. de Lange, S. Dehaene, S. Devore, F. Fallon, O. Ferrante, U. Gorska, R. Hirschhorn, O. Jensen, A. Khalaf, C. Koch, C. Kozma, G. Kreiman, A. Lepauvre, L. Liu, H. Luo, L. Melloni, L. Mudrik, M. Pitts, D. Richter, G. Tononi.

10.1126/science.abj3259

GRAPHIC: N. CARY/*SCIENCE*

NEUROSCIENCE

Navigating space in the mammalian brain

A study of bats reveals how hippocampal place cells code large-scale environments

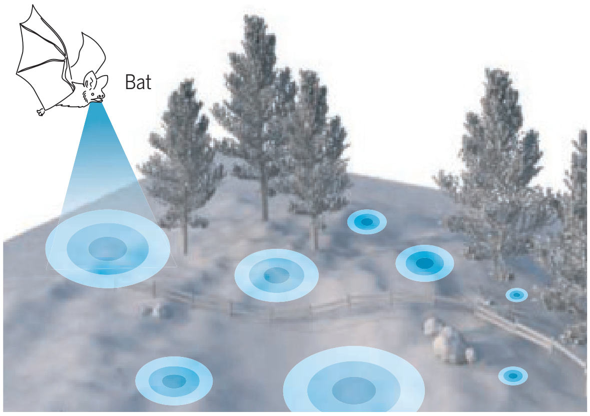
By **Emma R. Wood**[1] and **Paul A. Dudchenko**[2]

How does the brain represent the world and allow spatial navigation? One mechanism is hippocampal place cells—neurons that fire according to where an animal is in its environment. Different place cells fire according to different locations, and together they are thought to provide a cognitive map that supports spatial navigation and memory (*1*). Place cells have been described in a range of mammalian species, including mice, bats, marmosets, and humans. However, most studies have used rats in small enclosures or mazes. Thus, it is unknown how such representations might underpin larger-scale, real-world navigation. On page 933 of this issue, Eliav *et al.* (*2*) show that in bats flying in a large (200-m-long) enclosure, most place cells fire in several different locations and with varying spatial scales. Such multiscale representations are likely the most efficient way for a finite number of neurons to encode large distances.

Neurophysiological recordings in rats exploring relatively small "open-field" environments (~1 m²) or running along short tracks 1 to 2 m long have revealed that a given place cell in the hippocampus typically fires when the rat is in a single area within the apparatus (called its place field) (*1*, *3*, *4*). In the few experiments that have investigated bigger open-field environments and longer tracks, place fields are typically slightly enlarged compared with those in smaller environments (*4*–*6*), and individual place cells in CA1 (the main output region of the hippocampus) fire in multiple, irregularly spaced locations (*5*, *6*), with more place fields per cell in tracks of increasing length (*6*). Within a given environment, the different place fields of each hippocampal neuron are of a fairly uniform size, but there is an anatomical gradient, with the most dorsal hippocampal place cells having the smallest fields and ventral hippocampal cells having the largest fields (*3*, *7*). Together, these studies suggest that the hippocampus provides an ensemble place code, whereby different combinations of neurons are active in any given location, and that coding of different spatial scales is provided by different neurons across the dorsal-ventral hippocampal axis.

Navigating large, complex spaces

Eliav *et al*. found that bats exhibit multiscale place cell coding. Individual place cells in the hippocampus fire according to a range of spatial scales (place fields of a single place cell indicated by circles), allowing optimal processing of a large environment with a finite number of cells.



GRAPHIC: N. DESAI/*SCIENCE*

But how does the mammalian brain represent much larger spaces, on the spatial scale that animals would need to navigate in their natural environment? Eliav *et al.* wirelessly recorded from dorsal CA1 place cells in bats as they flew along a 200-m-long tunnel between two feeding stations. They found not only that place cells expressed multiple, irregularly spaced place fields in this very large environment but also that the size of the different place fields expressed by a given neuron varied widely: The mean ratio of the largest:smallest field was 4.4:1, but this was as high as 20:1 in some cells (see the figure). By contrast, and consistent with observations in rats, in a shorter 6-m-long tunnel, place cells expressed only one or two fields, the average field size was smaller than in the 200-m-long tunnel, and fields of the same cell were of a similar size (mean ratio <2:1).

These findings of multiscale coding by individual place cells may help answer a puzzling question: How can a finite population of place cells encode the large environments in which mammals navigate in the wild, at both large and small spatial scales? The modeling by Eliav *et al.* shows that the multiscale coding mechanism seen in the bats is a particularly efficient mechanism for coding large environments. It needs fewer neurons for accurate decoding of the current location of the bat than other ensemble coding mechanisms based on individual cells having multiple fields of the same size and other cells having fields of different sizes (as had previously been assumed).

It will be important to determine the extent to which multiscale coding by individual neurons is a general property of hippocampal coding across species and across different types and scales of environments. A preliminary study of rats following a moving robotic feeder in an 18.6-m² open-field environment reported that cells in dorsal CA1 exhibited the same type of multiscale coding as found in the tunnel-flying bats (*8*). This indicates that this type of firing may be a general principle of hippocampal coding of large-scale space across mammalian species. Moreover, perhaps in large, continuous spaces, multiscale place cell representation may be the rule.

As with many elegant studies, the work of Eliav *et al.* points to promising new avenues of research. One key question is how multiscale encoding arises. The two main inputs to CA1 (where the multiscale place cells have been described) are the CA3 and the medial entorhinal cortex (MEC). CA3 also contains place cells; indeed, the dorsal-ventral gradient of small-large place fields was described in CA3 neurons in rats (*7*). Conversely, the MEC contains a different type of spatial cell called grid cells. Each grid cell fires in multiple locations arranged in a regular hexagonal grid pattern that repeats across the environment (again with a dorsal-ventral arrangement of grid field size and spacing) (*9*, *10*). Grid cells are thought to be important for path integration, where animals use self-motion signals to estimate distances and directions

---

[1]Centre for Discovery Brain Sciences and Simons Initiative for the Developing Brain, University of Edinburgh, Edinburgh, UK. [2]Division of Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, UK.
Email: emma.wood@ed.ac.uk; p.a.dudchenko@stir.ac.uk

traveled. Eliav *et al.* suggest a feed-forward model whereby the multiscale fields in CA1 result from convergence of inputs from multiple CA3 place cells with different spatial scales onto each CA1 place cell. Predictions of this model that still need to be tested are that CA3 neurons should not show multiple fields in large environments and that either grid cells should not show multiple fields or grid cell inputs do not contribute to the firing of CA1 place fields in large environments.

A second question is whether there is a continuum of multiscale coding across environments of all sizes or whether (as suggested by Eliav *et al.*) multiscale coding occurs only in sufficiently large environments. And if the latter, what behavioral, perceptual, and neural mechanisms trigger the transition from small-scale to large-scale encoding of space?

The study of Eliav *et al.* provides a marker for the need to examine spatial coding in ethologically relevant environments. The multiscale place cell coding mechanism that they demonstrate may allow both fine-scale spatial localization and localization on a more extended scale, which would be required for navigating accurately between very distant locations hundreds of meters or kilometers apart. It will be interesting to see whether similar multiscale spatial representations occur in humans or nonhuman primates navigating (virtual or real) large, open spaces and whether multiscale coding by individual neurons occurs in other, nonspatial domains, such as the coding of time (*11*). ■

> **"...the multiscale coding mechanism seen in the bats is a particularly efficient mechanism for coding large environments."**

REFERENCES AND NOTES

1. J. O'Keefe, J. Dostrovsky, *Brain Res.* **34**, 171 (1971).
2. T. Eliav *et al.*, *Science* **372**, eabg4020 (2021).
3. M. W. Jung, S. I. Wiener, B. L. McNaughton, *J. Neurosci.* **14**, 7347 (1994).
4. J. O'Keefe, N. Burgess, *Nature* **381**, 425 (1996).
5. A. A. Fenton *et al.*, *J. Neurosci.* **28**, 11250 (2008).
6. P. D. Rich, H.-P. Liaw, A. K. Lee, *Science* **345**, 814 (2014).
7. K. B. Kjelstrup *et al.*, *Science* **321**, 140 (2008).
8. B. C. Harland *et al.*, bioRxiv 10.1101/2021.02.15.431172 (2021).
9. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, *Nature* **436**, 801 (2005).
10. V. H. Brun *et al.*, *Hippocampus* **18**, 1200 (2008).
11. C. J. MacDonald, K. Q. Lepage, U. T. Eden, H. Eichenbaum, *Neuron* **71**, 737 (2011).

10.1126/science.abi9663

MEDICINE

CRISPR diagnostics

New CRISPR enzyme activities add to the nucleic acid detection arsenal

By **Omar O. Abudayyeh** *and* **Jonathan S. Gootenberg**

Although clinical diagnostics take many forms, nucleic acid–based testing has become the gold standard for sensitive detection of many diseases, including pathogenic infections. Quantitative polymerase chain reaction (qPCR) has been widely adopted for its ability to detect only a few DNA or RNA molecules that can unambiguously specify a particular disease. However, the complexity of this technique restricts application to laboratory settings. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has underscored the need for the development and deployment of nucleic acid tests that are economical, easily scaled, and capable of being run in low-resource settings, without sacrifices in speed, sensitivity or specificity. CRISPR-based diagnostic (CRISPR-dx) tools offer a solution, and multiple CRISPR-dx products for detection of the SARS-CoV-2 RNA genome have been authorized by the US Food and Drug Administration (FDA). On page 941 of this issue, Jiao *et al.* (*1*) describe a new CRISPR-based tool to distinguish several SARS-CoV-2 variants in a single reaction.

There are multiple types of CRISPR systems comprising basic components of a single protein or protein complex, which cuts a specific DNA or RNA target programmed by a complementary guide sequence in a CRISPR-associated RNA (crRNA). The type V and VI systems and the CRISPR-associated endonucleases Cas12 (*2*, *3*) and Cas13 (*4*, *5*) bind and cut DNA or RNA, respectively. Furthermore, upon recognizing a target DNA or RNA sequence, Cas12 and Cas13 proteins exhibit "collateral activity" whereby any DNA or RNA, respectively, in the sample is cleaved regardless of its nucleic acid sequence (*4*, *6*). Thus, reporter DNAs or RNAs, which allow for visual or fluorescent detection upon cleavage, can be added to a sample to infer the presence or absence of specific DNA or RNA species (*4*–*8*).

Initial versions of CRISPR-dx utilizing Cas13 alone were sensitive to the low picomolar range, corresponding to a limit of detection of millions of molecules in a microliter sample. To improve sensitivity, preamplification methods, such as recombinase polymerase amplification (RPA), PCR, loop-mediated isothermal amplification (LAMP), or nucleic acid sequence–based amplification (NASBA), can be used with Cas12 or Cas13 to enable a limit of detection down to a single molecule (*8*). This preamplification approach, applicable to both Cas12 and Cas13 (*6*, *7*), enabled a suite of detection methods and multiplexing up to four orthogonal targets (*7*). Additional developments expanded CRISPR-dx readouts beyond fluorescence, including lateral flow (*7*), colorimetric (*9*), and electronic or material responsive readouts (*10*), allowing for instrument-free approaches. In addition, post–collateral-cleavage amplification methods, such as the use of the CRISPR-associated enzyme Csm6, have been combined with Cas13 to further increase the speed of CRISPR-dx tests (*7*). As an alternative to collateral-cleavage–based detection, type III CRISPR systems, which involve large multiprotein complexes capable of targeting both DNA and RNA, have been used for SARS-CoV-2 detection through production of colorimetric or fluorometric readouts (*11*).

FDA-authorized CRISPR-dx tests are currently only for use in centralized labs, because the most common CRISPR detection protocols require fluid handling steps and two different incubations, precluding their immediate use at the point of care. Single-step formulations have been developed to overcome this limitation, and these "one-pot" versions of CRISPR-dx are simple to run, operate at a single temperature, and run without complex equipment, producing either fluorescence or lateral flow readouts. The programmability of CRISPR makes new diagnostic tests easier to develop, and within months of the release of the SARS-CoV-2 genome, many COVID-19–specific CRISPR tests were reported and distributed around the world.

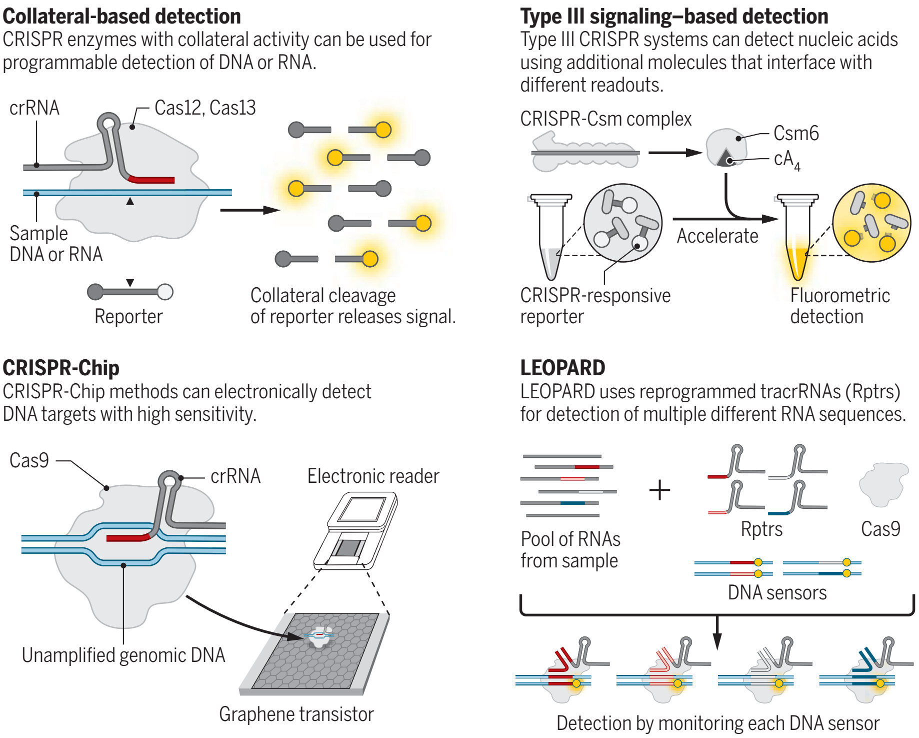
The broader capability for Cas enzyme–enhanced nucleic acid binding or cleavage has led to several other detection modalities. Cas9-based methods for cleaving nucleic acids in solution for diagnostic purposes have been combined with other detection platforms, such as destruction of undesired amplicons for preparation of next-generation sequencing libraries (*12*), or selective removal of alleles for nucleotide-specific detection

---

McGovern Institute for Brain Research at MIT, Massachusetts Institute of Technology, Cambridge, MA, USA. Email: omarabu@mit.edu; jgoot@mit.edu

(*13*). Alternatively, the programmable cleavage event from the Cas nuclease can be used to initiate an amplification reaction (*14*). Cas9-based DNA targeting has also been used for nucleotide detection in combination with solid-state electronics, promising an amplification-free platform for detection. In this platform, called CRISPR-Chip, the Cas9 protein binds nucleotide targets of interest (often in the context of the native genome) to graphene transistors, where the presence of these targets alters either current or voltage (*15*). By utilizing additional Cas9 orthologs and specific guide designs, CRISPR-Chip approaches have been tuned for single–base-pair sensitivity (*15*). Because they are integrated with electronic readers, CRISPR-Chip platforms may allow facile point-of-care detection with handheld devices.

Different classes of CRISPR diagnostics


Collateral-based detection
CRISPR enzymes with collateral activity can be used for programmable detection of DNA or RNA.

Type III signaling–based detection
Type III CRISPR systems can detect nucleic acids using additional molecules that interface with different readouts.

CRISPR-Chip
CRISPR-Chip methods can electronically detect DNA targets with high sensitivity.

LEOPARD
LEOPARD uses reprogrammed tracrRNAs (Rptrs) for detection of multiple different RNA sequences.

cA_4, cyclic tetra-adenylate; Cas, CRISPR-associated endonuclease; crRNA, CRISPR-associated RNA; LEOPARD, leveraging engineered tracrRNAs and on-target DNAs for parallel RNA detection; tracrRNA, trans-activating crRNA.

Jiao *et al.* use a distinct characteristic of type II CRISPR systems, which involve Cas9, to develop a new type of noncollateral based CRISPR detection. Unlike Cas12s and Cas13, Cas9-crRNA complex formation requires an additional RNA known as the trans-activating CRISPR RNA (tracrRNA). By sequencing RNAs bound to Cas9 from *Campylobacter jejuni* in its natural host, the authors identified unexpected crRNAs, called noncanonical crRNA (ncrRNA), that corresponded to endogenous transcripts. Upon investigation of this surprising observation, it became clear that the tracrRNA was capable of hybridizing to semi-complementary sequences from a variety of RNA sources, leading to biogenesis of ncrRNAs of various sizes. Recognizing that they could program tracrRNAs to target a transcript of interest, the authors generated a reprogrammed tracrRNA (Rptr) that could bind and cleave a desired transcript, converting a piece of that transcript into a functional guide RNA. By then creating fluorescent DNA sensors that would be cleaved by the Rptr and ncrRNAs, the sensing of RNA by Cas9 could be linked to a detectable readout. This platform, called LEOPARD (leveraging engineered tracrRNAs and on-target DNAs for parallel RNA detection), can be combined with gel-based readouts and enables multiplexed detection of several different sequences in a single reaction (see the figure).

Jiao *et al.* also combined LEOPARD with PCR in a multistep workflow to detect SARS-CoV-2 genomes from patients with COVID-19. Although more work is needed to integrate this Cas9-based detection modality into a single step with RPA or LAMP to create a portable and sensitive isothermal test, an advantage of this approach is the higher-order multiplexing that can be achieved, allowing multiple pathogens, diseases, or variants to be detected simultaneously. More work is also needed to combine this technology with extraction-free methods for better ease of use; alternative readouts to gel-based readouts, such as lateral flow and colorimetric readouts, would be beneficial for point-of-care detection.

In just 5 years, the CRISPR-dx field has rapidly expanded, growing from a set of peculiar molecular biology discoveries to multiple FDA-authorized COVID-19 tests and spanning four of the six major subtypes of CRISPR systems. Despite the tremendous promise of CRISPR-dx, substantial challenges remain to adapting these technologies for point-of-care and at-home settings. Simplification of the chemistries to operate as a single reaction in a matter of minutes would be revolutionary, especially if the reaction could be run at room temperature without any complex or expensive equipment. These improvements to CRISPR-dx assays can be achieved by identification or engineering of additional Cas enzymes with lower-temperature requirements, higher sensitivity, or faster kinetics, enabling rapid and simple amplification-free detection with single-molecule sensitivity.

Often overlooked is the necessity for a sample DNA or RNA preparation step that is simple enough to be added directly to the CRISPR reaction to maintain a simple workflow for point-of-care testing. In addition, higher-order multiplexing developments would allow for expansive testing menus and approach the possibility of testing for all known diseases. As these advancements are realized, innovative uses of CRISPR-dx will continue in areas such as surveillance, integration with biomaterials, and environmental monitoring. In future years, CRISPR-dx assays may become universal in the clinic and at home, reshaping how diseases are diagnosed. ■

REFERENCES AND NOTES

1. C. Jiao *et al.*, *Science* **372**, 941 (2021).
2. B. Zetsche *et al.*, *Cell* **163**, 759 (2015).
3. S. Shmakov *et al.*, *Mol. Cell* **60**, 385 (2015).
4. O. O. Abudayyeh *et al.*, *Science* **353**, aaf5573 (2016).
5. A. East-Seletsky *et al.*, *Nature* **538**, 270 (2016).
6. J. S. Chen *et al.*, *Science* **360**, 436 (2018).
7. J. S. Gootenberg *et al.*, *Science* **360**, 439 (2018).
8. J. S. Gootenberg *et al.*, *Science* **356**, 438 (2017).
9. W. K. Spoelstra *et al.*, *Biolphys. J.* 10.1016/j.bpj.2021.02.013 (2021).
10. M. A. English *et al.*, *Science* **365**, 780 (2019).
11. A. Santiago-Frangos *et al.*, *medRxiv* 10.1101/2020.10.14.20212670 (2020).
12. W. Gu *et al.*, *Genome Biol.* **17**, 41 (2016).
13. K. Pardee *et al.*, *Cell* **165**, 1255 (2016).
14. W. Zhou *et al.*, *Nat. Commun.* **9**, 1 (2018).
15. R. Hajian *et al.*, *Nat. Biomed. Eng.* **3**, 427 (2019).

ACKNOWLEDGMENTS

We thank B. Desimone and J. Crittenden for helpful feedback. O.O.A. and J.S.G. are coinventors on patent applications relating to CRISPR diagnostics. O.O.A. and J.S.G. are cofounders of Sherlock Biosciences, Pine Trees Health, Tome Biosciences, and Moment Biosciences.

10.1126/science.abi9335

GRAPHIC: ERIN DANIEL

MEDICINE

Bispecific antibodies

Bispecific antibodies have emerged as molecules with a multitude of talents

By **Ulrich Brinkmann**[1] and **Roland E. Kontermann**[2]

Bispecific antibodies (bsAbs) bind two different epitopes on the same or different antigens. Through this dual specificity for soluble or cell-surface antigens, bsAbs exert activities beyond those of natural antibodies, offering numerous opportunities for therapeutic applications. Although initially developed for retargeting T cells to tumors, with a first bsAb approved in 2009 (catumaxomab, withdrawn in 2017), exploring new modes of action opened the door to many additional applications beyond those of simply combining the activity of two different antibodies within one molecule. Examples include agonistic "assembly activities" that mimic the activity of natural ligands and cofactors (for example, factor VIII replacement in hemophilia A), inactivation of receptors or ligands, and delivery of payloads to cells or tissues or across biological barriers. Over the past years, the bsAb field transformed from early research to clinical applications and drugs. New developments offer a glimpse into the future promise of this exciting and rapidly progressing field.

Monoclonal antibodies (mAbs) comprise antigen-binding sites formed by the variable domains of the heavy and light chain and an Fc region that mediates immune responses. BsAbs, produced through genetic engineering, combine the antigen-binding sites of two different antibodies within one molecule, with a plethora of formats available (*1*). Conceptually, one can discriminate between bsAbs with combinatorial modes of action where the antigen-binding sites act independently from each other, and bsAbs with obligate modes of action where activity needs binding of both, either in a sequential (temporal) way or dependent on the physical (spatial) linkage of both (see the figure) (*2*). BsAbs approved as drugs are so far in the obligate dual-binding category: A T cell recruiter (blinatumomab) against cancer and a factor VIIIa mimetic to treat hemophilia A (emicizumab). Most but not all of the more than 100 bsAbs in clinical development address cancers. Some are in late stage (such as amivantamab, epcoritamab, faricimab, and KN046), but most are still in early stages (*2*). Most of these entities enable effector cell retargeting to induce target cell destruction.

An increasing number of programs also explore alternative modes of action. This includes bsAbs that target pathways involved in tumor proliferation (such as amivantamab), invasion, ocular angiogenesis (such as faricimab), or immune regulation by blocking receptors and/or ligands, mainly in a combinatorial manner. Challenges for all of these entities are potential adverse effects, toxicity in normal tissues, and overshooting and systemic immune responses, especially with T cell retargeting or immune-modulating or activating entities. Such issues need to be carefully addressed.

Most of the bispecific T cell engagers comprise a binding site for a tumor-associated antigen and CD3 [a component of the T cell receptor (TCR) activation complex] as trigger molecule on T cells. To prevent or ameliorate "on-target, off-tumor" effects of T cell recruiters, approaches currently investigated include the modulation of target affinities and mechanisms to allow conditional activation upon target cell binding. Thus, a reduced affinity for CD3 increased tolerability by reducing peripheral cytokine concentrations that are associated with nonspecific or overshooting immune reactions (*3*). Similarly, reduced affinity for the target antigen was shown to ameliorate cytokine release and damage of target-expressing tissues (*4*). Tumor selectivity can be further increased by implementing avidity effects—for example, by using 2+1 bsAb formats with two low-affinity binding sites for target antigens and monovalent binding to CD3 (*4*).

In further approaches, binders to CD3 were identified that efficiently trigger target cell destruction without inducing undesired release of cytokines, demonstrating the importance of epitope specificity to potentially uncouple efficacy from cytokine release (*5*). Complementing these T cell–recruiting principles, the nonclassical T cell subset of $\gamma9\delta2$ T cells with strong cytotoxic activity emerged as potent effectors, which can be retargeted with bsAbs binding to the $\gamma9\delta2$ TCR. Thereby, global activation of all T cells, including inhibitory regulatory T cells (T_{reg} cells), through CD3 binding, may be avoided (*6*). However, even these approaches might result in a narrow therapeutic window to treat solid tumors because of T cell activation in normal tissues.

Consequently, there are several approaches to conditionally activate T cells within tumors, including a local liberation of the CD3-binding sites or triggering local assembly of CD3-binding sites from two half-molecules. For example, CD3-binding sites have been masked by fusing antigen binding or blocking moieties—such as peptides, aptamers, or anti-idiotypic antibody fragments—to one or both variable domains. These moieties are released within the tumor by tumor-associated proteases, or through biochemical responses to hypoxia or low pH (*7*). This approach can also be applied to confer specific binding of antibody therapeutics, including bsAbs, to antigens on tumor cells (*8*).

An on-target restoration of CD3-binding sites requires application of two target-binding entities, each comprising parts of the CD3-binding site, which assemble into functional binding sites upon close binding of both half-antibodies. The feasibility of this approach was recently shown, for example, for a split T cell–engaging antibody derivative (Hemibody) that targets a cell surface antigen (*9*). Such approaches can also be applied to half-antibodies that recognize two different targets expressed on the same cell, further increasing tumor selectivity.

Regarding T cell engagers, increasing efforts are made to target not only cell-surface antigens expressed on tumor cells but also human leukocyte antigen (HLA)–presented tumor-specific peptides. This expands the target space of bsAbs toward tumor-specific intracellular antigens and can be achieved by using either recombinant TCRs or antibodies with TCR-like specificities combined with, for example, CD3-binding arms to engage T cell responses. A first TCR–anti-CD3 bispecific molecule is in phase I and II trials to treat metastatic melanoma (*10*). A challenge of this approach is the identification of TCRs or TCR-like antibodies that bind the peptide in the context of HLA with high affinity and specificity, without cross-reacting with related peptides to reduce or avoid off-target activities. Comprehensive screening tools and implementation of computational approaches are being developed to achieve this task.

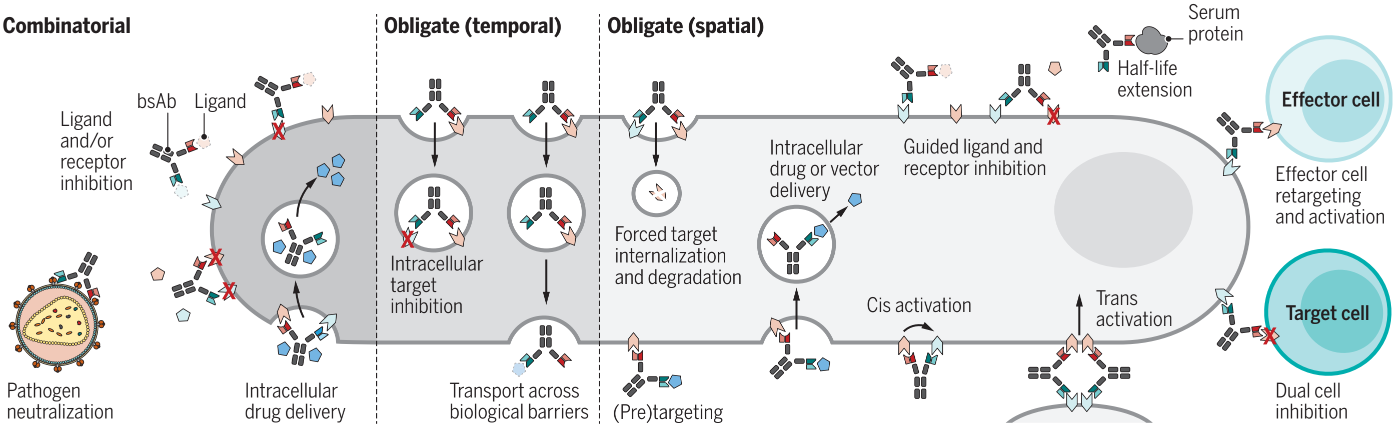
A rapidly growing area of bsAbs in cancer therapy is their use to foster antitumor immune responses. Here, they are especially applied for dual inhibition of checkpoints that prevent immune responses—for example, programmed cell death protein 1 (PD-1) and its ligand (PD-L1), cytotoxic T lymphocyte–associated antigen 4 (CTLA-4), or lymphocyte activation gene 3 (LAG-3; for example,

---

[1] Roche Pharma Research and Early Development, Large Molecule Research, Roche Innovation Center Munich, Nonnenwald 2, 82377 Penzberg, Germany. [2] Institute of Cell Biology and Immunology, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany. Email: ulrich.brinkmann@roche.com; roland.kontermann@izi.uni-stuttgart.de

Modes of action of bispecific antibodies

There are >100 bispecific antibodies (bsAbs) in clinical development. These are broadly classified as combinatorial, combining the activity of two antibodies within one molecule, or obligate, where combining both binding sites creates a temporal or spatial activity.


Combinatorial
Obligate (temporal)
Obligate (spatial)
Ligand and/or receptor inhibition
bsAb
Ligand
Pathogen neutralization
Intracellular drug delivery
Intracellular target inhibition
Transport across biological barriers
Forced target internalization and degradation
(Pre)targeting
Intracellular drug or vector delivery
Guided ligand and receptor inhibition
Cis activation
Trans activation
Serum protein
Half-life extension
Effector cell
Effector cell retargeting and activation
Target cell
Dual cell inhibition

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

KNO46). Tumor-targeted bsAbs can also target costimulatory factors such as CD28 or 4-1BB ligand (4-1BBL) to enhance T cell responses when combined with PD-1 blockade or to provide an activity-enhancing costimulatory signal in combination with CD3-based bsAbs (*11*). Furthermore, bsAbs are being developed for local effects by targeting one arm to antigens that are expressed by tumor cells or cells of the tumor microenvironment (*2*).

Clinical application of bsAbs now expands to other therapeutic areas, including chronic inflammatory, autoimmune, and neurodegenerative diseases; vascular, ocular, and hematologic disorders; and infections. In contrast to mAbs, bsAbs can inactivate the signaling of different cytokines with one molecule to treat inflammatory diseases (*12*). Simultaneous dual-target binding is not essential to elicit activity for bsAbs against combinations of proinflammatory cytokines, such as tumor necrosis factor (TNF), interleukin-1α (IL-1α), IL-1β, IL-4, IL-13, IL-17, inducible T cell costimulator ligand (ICOSL), or B cell–activating factor (BAFF). This presumably also applies to blockade of immune cell receptors, although dual targeting might confer increased efficacy due to avidity effects and increased selectivity through simultaneous binding of two different receptors.

A further application of combinatorial dual targeting is in ophthalmology. Loss of vision in wet age-related macular degeneration (AMD) results from abnormal proliferation and leakiness of blood vessels in the macula. This can be treated with antibodies that bind and inactivate factors that stimulate their proliferation (*13*). In contrast to mAbs or fragments that recognize individual factors, bsAbs bind two such factors. For example, faricimab that binds vascular endothelial growth factor A (VEGF-A) and angiopoietin-2 (ANG2), demonstrated dual efficacy in preclinical studies, and is currently in phase 3 trials.

BsAbs with obligate modes of action that mandate simultaneous dual-target binding are "assemblers" that replace the function of factors necessary to form functional protein complexes. One of these bsAbs with an assembly role (emicizumab, approved in 2018) replaces factor VIIIa in the clotting cascade. Deficiency of factor VIII causes hemophilia A, which can be overcome by substitution with recombinant factor VIII. However, a proportion of patients develop factor VIII–neutralizing immune responses and no longer respond to therapy. To overcome this, a bsAb was developed with binding sites that recognize and physically connect factors IXa and X, a process normally mediated by factor VIIIa. Extensive screening of a large set of bsAbs was required to identify those that combine suitable epitopes with optimized affinities and geometry to serve as functional factor VIIIa mimetics (*14*). This exemplifies the complexity of identifying the best bsAb for therapeutic applications.

A mode of action requiring sequential binding of two targets is the transport of bsAbs across the blood-brain barrier (BBB). This is a tight barrier of brain capillary endothelial cells that controls the transport of substances between the blood and the cerebrospinal fluid—the brain parenchyma. Passage of large molecules, including antibodies, across the BBB is thereby restricted. Some proteins, such as transferrin or insulin, pass through the BBB by way of transporters on endothelial cells. Antibodies that bind these shuttle molecules, such as the transferrin receptor (TfR), can hitchhike across the BBB. BsAbs that recognize brain targets (such as β-amyloid for Alzheimer's disease) and TfR with optimized affinities, epitopes, and formats can thereby enter the brain. Such bsAbs are currently in clinical evaluation to treat neurodegenerative diseases (*15*).

In the past years, there has been a transition from a technology-driven phase, solving hurdles to generate bsAbs with defined composition, toward exploring and extending the modes of action for new therapeutic options. The challenge of generating bsAbs is not only to identify suitable antigen pairs to be targeted in a combined manner. It is now recognized that the molecular composition has a profound impact on bsAb functionality (*13*). That more than 30 different bsAb formats are in clinical trials proves that development is now driven by a "fit for purpose" or "format defines function" rationale. Many candidates differ in their composition, affecting valency, geometry, flexibility, size, and half-life (*1*). Not all members of this "zoo of bsAb formats" qualify to become drugs. Strong emphasis is therefore on identifying candidates that exhibit drug-like properties and fulfill safety, developability, and manufacturability criteria. There is likely to be an exciting new wave of bsAb therapeutics available in the coming years. ■

REFERENCES AND NOTES

1. U. Brinkmann, R. E. Kontermann, *MAbs* **9**, 182 (2017).
2. A. F. Labrijn *et al.*, *Nat. Rev. Drug Discov.* **18**, 585 (2019).
3. K. Staflin *et al.*, *JCI Insight* **5**, e133757 (2020).
4. D. Slaga *et al.*, *Sci. Transl. Med.* **10**, eaat5775 (2018).
5. N. D. Trinklein *et al.*, *MAbs* **11**, 639 (2019).
6. I. de Weerdt *et al.*, *Cancer Immunol. Res.* **9**, 50 (2021).
7. W. W. Lin *et al.*, *J. Biomed. Sci.* **27**, 76 (2020).
8. K. A. Autio *et al.*, *Clin. Cancer Res.* **26**, 984 (2020).
9. A. Banaszek *et al.*, *Nat. Commun.* **10**, 5387 (2019).
10. M. R. Middleton *et al.*, *Clin. Cancer Res.* **26**, 5869 (2020).
11. J. C. Waite *et al.*, *Sci. Transl. Med.* **12**, 549 (2020).
12. Q. Zhao, *BioDrugs* **34**, 111 (2020).
13. S. Nie *et al.*, *Antib. Ther.* **3**, 18 (2020).
14. T. Kitazawa *et al.*, *Nat. Med.* **18**, 1570 (2012).
15. F. Weber *et al.*, *Cell Rep.* **22**, 149 (2018).

ACKNOWLEDGMENTS

U.B. is employed by and owns stocks in Roche and is a co-inventor on patent applications covering bsAbs. R.E.K. is a co-inventor on patent applications covering bsAbs and reports licensing agreements and research support from Baliopharm, BioNTech, and Oncomatryx and consultancy fees from Oncomatryx, Roche, and Immatics.

10.1126/science.abg1209

POLICY FORUM

CLIMATE FINANCE

Accounting for finance is key for climate mitigation pathways

Investors' expectations can hamper a low-carbon transition

By **Stefano Battiston**[1,2], **Irene Monasterolo**[3,4,5], **Keywan Riahi**[5,6], **Bas J. van Ruijven**[5]

The financial system—the ecosystem of investors (e.g., banks, investment funds, insurance), markets, and instruments—is often considered to play an enabling role in climate mitigation pathways to a low-carbon transition (*1*). But it can also have a hampering role, e.g., if investors' perceptions of low risk from a missed transition and low opportunities from a transition fail to trigger a reallocation of capital into low-carbon investments. This increases the chance of the transition not occurring within the time window required to stabilize the climate or occurring in a disorderly fashion. Indeed investors, who can influence the realization of climate mitigation pathways, themselves rely on estimates of climate mitigation pathways from process-based integrated assessment models (IAMs) (*2*). And IAMs do not model the financial system or investors' decisions; thus, the feedback loop between the financial system and mitigation pathways is not taken into account, neither by the IAMs nor by the finance community. This limitation to our understanding of the dynamics and the feasibility of the low-carbon transition weakens the ability of IAMs to inform policy and investment decisions. We propose a framework to capture the interdependence between investors' perception of future climate risk, depending on the credibility of climate policies, and the allocation of investments in the economy.

CLIMATE MITIGATION SCENARIOS

Process-based, large-scale IAMs are used to develop long-term emission projections and socioeconomic scenarios assessed by the Intergovernmental Panel on Climate Change (IPCC) (*1*). Scenarios are constructed to suggest how to reach given targets in terms of cumulative emissions (and thus in terms of carbon budget) by 2100, which translate into temperature targets with associated probabilities. The IAM literature, assessed by the IPCC, produced a set of archetypical climate mitigation scenarios representing the most distinct features of how the transition could happen in the next decades. This is where the notion of risk is key. Whereas investors' preferences differ in terms of risk aversion and investment strategies, they all make investment decisions based on their assessment of risk.

In 2019, the Network of Central Banks and Supervisors for Greening the Financial System (NGFS), a global platform of more than 80 financial authorities, recognized that climate change poses new risks for citizens' investments and savings. It recommended a climate risk assessment of financial portfolios using several high-level scenarios (*3*), including (i) an orderly transition, in which climate policies are introduced early and predictably and climate risks are priced in by financial markets; and (ii) a disorderly transition, in which the impact of climate policies is not (fully) anticipated by investors. In the first case, firms have time to plan ahead and investors to reallocate capital gradually. In the second case, high-carbon firms and investors face losses that can trigger market instability and costs for society as a whole. High-carbon firms would lose out in both situations, and more so in the disorderly scenario. By contrast, low-carbon firms would benefit in both situations, but not necessarily more in the disorderly scenario.

To translate these scenarios into quantitative trajectories of economic variables, the NGFS and a growing number of investors already use the output of process-based IAMs as an input for climate financial risk analysis (*4*). These scenarios, describing what the world might look like several years from now, have the power to shift markets' expectations today, because they are endorsed by the NGFS and large investors. It is thus critical to understand if these scenarios for potential tomorrows could lead, unintentionally, to insufficient investments today, owing to their not accounting for the role of financial actors themselves. Our framework addresses this challenge and allows us to derive scenarios that complement the current IPCC and NGFS scenarios, strengthening climate financial risk assessment.

Firms and consumers' responses to climate policies are endogenous to the IAMs and have long been investigated (*5*). But the ways in which investors' responses to climate policies affect the outcome of those policies have not been investigated in IAMs and are not well understood (*6*). Indeed, IAMs [including large-scale computable general equilibrium models (*7*)] consider "finance" only to the extent that firms' access to financing is assumed to be available at no cost and with no limits (*8*). They overlook that financing is provided by investors on the basis of assessed risk, resulting in financing costs and limits on funding. IAMs include no actors (e.g., banks) that can decide to grant loans to firms, or actors (e.g., insurance firms, pension funds) that can decide to invest (or not) in stock market shares of firms. This leads to the opportunity to interface IAMs with models where investors carry out a financial risk assessment.

AN ENABLING OR HAMPERING ROLE OF THE FINANCIAL SYSTEM

Consider a utility firm that seeks financing to shift its power plants from high- to low-carbon technologies. If a bank perceives the strategy as less (more) risky than staying on high-carbon technologies—because the climate policy, e.g., a carbon price, is perceived as credible (noncredible)—it will soon charge a lower (higher) interest rate on the loan, thus facilitating (delaying) the firm's technological conversion. In general, if investors start to perceive high risk from a missed transition and, conversely, high opportunities from a successful transition with credible climate policies (*9*), they adjust their expectations. They thus reallocate capital into low-carbon investments early and gradually and can even anticipate the policy impact, as described by the notion of climate sentiments (*10*). This "enabling" behavior facilitates the transition, because it leads to smoother adjustments of the economy and of prices. If, by contrast, investors' perception goes in the opposite direction, they react late and suddenly. This "hampering" behavior makes the transition more costly for society, because it can lead to abrupt reallocations of capital and price adjustments. The enabling or hampering roles of the financial system can explain how the orderly and disorderly transition emerges from the interplay of policy timing and investors' reactions.

Overall, the presence of the financial system may induce a path dependence in the complex dynamics leading to lock-in effects similar to those described in models of tech-

---

[1] Department of Banking and Finance, University of Zurich, Zurich, Switzerland. [2] Department of Economics, Ca' Foscari University of Venice, Venice, Italy. [3] Vienna University of Economics and Business, Vienna, Austria. [4] Global Development Policy Center, Boston University, Boston, MA, USA. [5] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [6] Graz University of Technology, Graz, Austria. Email: stefano.battiston@uzh.ch

nology diffusion. In particular, transition trajectories could differ from those described by IAMs' scenarios because IAMs only consider technology cost and not the financing costs—nor investors' reactions—to deploy such technologies. For instance, investors could interpret NGFS scenarios of orderly transition as a situation in which high-carbon firms are only slightly more risky than low-carbon ones, because firms can adjust their technology mix and spread over time losses arising from stranded assets, i.e., unusable high-carbon installations (*11*). Driven by this risk perception, investors could play a hampering role and only reallocate capital from high- to low-carbon firms to a limited extent. It is not guaranteed that this level of reallocation is sufficient to fund investments in low-carbon energy that this scenario assumes; a low-risk perception induced by the orderly scenario could make the scenario unfeasible in the real world.

By contrast, investors could interpret a scenario of a disorderly transition as a situation in which high-carbon firms become substantially more risky than low-carbon ones, following the introduction of stringent carbon prices. For instance, high-carbon energy firms have an incentive to delay their own conversion, but owing to increasing public and investor demand, politicians could eventually find an agreement to introduce stringent climate policies. Because of the opposing interests in the negotiation, this could happen at the last moment. Investors who want their portfolio to withstand such an outcome could play an enabling role and start to demand a higher interest rate from high-carbon firms well before the policy introduction. They would reallocate capital to low-carbon firms to hedge the risk (if they continue to have a preference for lower risk). Thus, the capital reallocation, driven by risk perception, could lead to investments in low-carbon firms that increase earlier and at higher levels than those assumed in the scenario, and hence to larger mitigation opportunities. Furthermore, the financial feedback on firms' investment decisions can also lead to cascading effects in the economy. To some extent, this is precisely what the disorderly scenario is meant for: to allow an assessment of risk by investors in order to avert the realization of the scenario itself. This should not come as a surprise, in the same way that requiring buildings' owners to consider fire scenarios is ultimately meant to avoid the adverse scenario.

Possible inconsistencies between the investment levels in the original IAMs' scenarios and those resulting from the role of the financial system motivate the need for a new framework to connect climate mitigation scenarios and financial risk assessment in a circular way.

GRAPHIC: K. FRANKLIN/*SCIENCE*

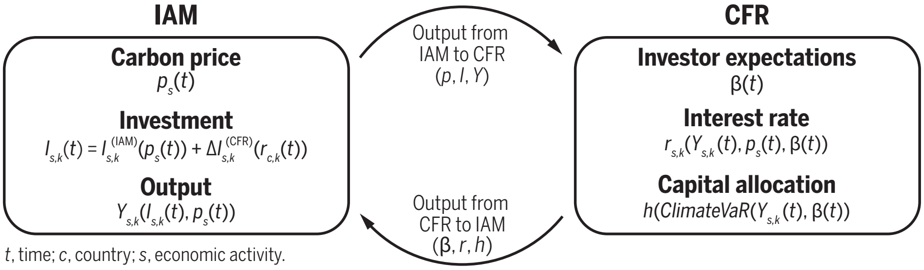
CONNECTING MITIGATION SCENARIOS AND FINANCIAL RISK ASSESSMENT

The use of IAM scenarios to assess climate-related financial risks has been introduced by academic work on the climate-stress test of the financial system (*12*), and in recent applications to central banks' data (*13*). It consists of translating IAMs' output trajectories across technologies (e.g., fossil fuels and renewables) and scenarios into financial shocks on investors' portfolios. Combined with financial network models, it also captures the amplification of shocks through financial interconnectedness and the implications for individual and systemic financial stability. We refer to this approach as the climate financial risk (CFR) model.

In IAMs, the decision of firms on how many energy plants of a given technology to build is determined by the carbon pricing. The financing costs and the fact that they vary with the risk attributed to technologies by investors is not taken into account but can be obtained from a CFR model.

Linking climate mitigation pathways and the financial system

An integrated assessment model (IAM) generates economic output trajectories under climate policy scenarios. A climate financial risk (CFR) model uses IAM output to compute interest rates for firms using different energy technologies (k). Investors' expectations and climate value-at-risk (*ClimateVaR*) determine capital allocation across technologies. The IAM updates to reflect diversity in financing costs.


IAM
Carbon price
$p_s(t)$
Investment
$I_{s,k}(t) = I_{s,k}^{(IAM)}(p_s(t)) + \Delta I_{s,k}^{(CFR)}(r_{c,k}(t))$
Output
$Y_{s,k}(I_{s,k}(t), p_s(t))$
Output from IAM to CFR
(p, I, Y)
Output from CFR to IAM
(β, r, h)
CFR
Investor expectations
$\beta(t)$
Interest rate
$r_{s,k}(Y_{s,k}(t), p_s(t), \beta(t))$
Capital allocation
$h(ClimateVaR(Y_{s,k}(t), \beta(t)))$

t, time; c, country; s, economic activity.

We propose a general framework (see the first figure) to link IAMs and CFR in a circular way, which can be embodied with different choices of specific IAMs and CFR models. The IAM module generates sets of climate mitigation scenarios, which are then used by the CFR to model how investors assess the financial risk of high- and low-carbon firms along the IAM's trajectories. If investors assess a higher risk for high-carbon firms, they demand higher interest rates on loans, and higher yields for bonds, to provide funding. They may also divest from some high-carbon firms to reinvest in low-carbon firms in order to balance their portfolio risk. The resulting trajectories of financing cost across low- and high-carbon firms are fed back to the IAMs to update the respective mitigation scenarios, thus closing the loop between the IAM and the CFR.

Such IAM-CFR scenarios can demonstrate the interplay between the role of the financial system (enabling or hampering) and the timing of the climate policy introduction (see the second figure for a stylized illustration), the latter being identified by the IPCC as a key dimension of climate mitigation. We condition the analysis to a temperature target of 2°C, in line with the Paris Agreement, but other scenarios can be analyzed. On the basis of the IPCC and NGFS scenarios, we consider two options for the timing of policies. The immediate case focuses on 2020 [based on (*1*)], but results would be similar for 2021 or 2022. The delayed case focuses on 2030, because a transition that starts later than 2040 is considered not compatible with the 2°C target.

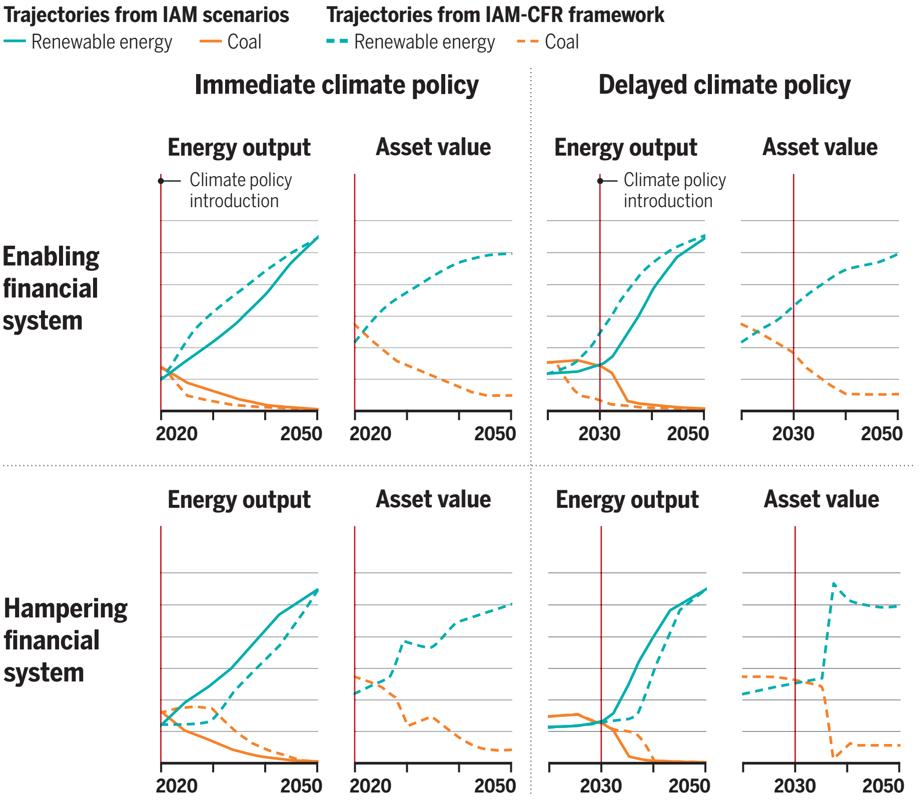
All IAMs share the general result that, in a transition scenario that achieves a 2°C target, the output (energy production) of high-carbon activities starts to decline at the introduction of the policy, and the reverse applies for low-carbon activities. Although quantitative details of output trajectories vary across IAMs, the solid curves in the second figure represent this common stylized behavior.

In the enabling cases (see the second figure, top panels), investors start to demand higher interest rates for high-carbon firms at the introduction of the policy, or even earlier (and lower interest rates for low-carbon firms). In the IAM-CFR scenarios, output of high-carbon firms must be lower than in the original IAM-only scenarios where only the carbon tax is considered (the reverse for low-carbon firms). This is due to financing costs, which are additional to the carbon tax and enhance the differences in profitability of firms across technologies. Accordingly, the values of assets of high-carbon firms decrease gradually and those of low-carbon firms increase gradually.

Investors' expectations are the key feature that sets the system to an enabling or a hampering role of the financial system. Because expectations are subject to herding dynamics whereby investors try to guess each other's next move, a stampede can occur unexpectedly (e.g., the 2008 subprime mortgage crisis). In the hampering cases (see the second figure, bottom panels), investors delay revising their expectations, but then expectations change suddenly. This results in output trajectories that increase (or decrease) at a faster pace than in the cases without CFR, as trajectories have less time to meet the same carbon budget. Financial asset values also adjust suddenly. The adjustment has to occur before 2040; otherwise the 2°C target is missed.

Interplay between the financial system and climate policy timing

Four main climate transition scenarios are shown. Solid curves are the same in top and bottom panels and represent stylized trajectories from existing integrated assessment model (IAM) scenarios of electricity production from coal or renewable energy. Dashed curves represent stylized trajectories from the IAM–climate financial risk (CFR) framework of electricity production and asset values. The difference between solid and dashed curves is the effect of accounting for the role of the financial system.



In the NGFS scenarios, the orderly versus disorderly character of scenarios is assumed, independently of the role of the financial system. Here we show two cases where the financial system can largely modify the outcome of those scenarios. First, an immediate transition to 2°C is classified in the NGFS scenarios as orderly. If, however, the financial system plays a hampering role (see the second figure, bottom left panel), the transition is delayed and there are large and sudden financial value adjustments. These features threaten financial stability and would be more consistent with a disorderly scenario. Second, a delayed transition to 2°C is classified there as disorderly. If the financial system plays an enabling role (see the second figure, top right panel), the gradual price adjustments along the trajectory would be more consistent with an orderly NGFS scenario. Finally, a disorderly transition could also lead to higher risk than described in the NGFS disorderly scenario, if the financial system plays a hampering role. Neglecting the role of the financial system could thus lead to an overestimate or underestimate of risk across NGFS scenarios.

CONCLUSION AND POLICY IMPLICATIONS

Our approach opens the way to a new understanding of risks and opportunities associated with the low-carbon transition. By conditioning the investment decision of investors to the credibility of climate policy scenarios, we consider the role of the financial system as enabling or hampering the low-carbon transition. This could reverse the ordering of costs and benefits of climate mitigation policies, which are currently distorted by not considering the financial system.

Such analyses can provide new insights on the implementation of fiscal policies, such as carbon pricing and the phasing out (in) of fossil fuel (renewable energy) subsidies. Neglecting the role of finance implies that a projected carbon price schedule could miss the emissions target because the mitigation scenario does not necessarily imply a risk perception by the financial system that leads to the investment reallocation assumed by the scenario. Similarly, plans for phasing out carbon subsidies have an impact on the perception of risk by the financial system with regard to high-carbon technologies. Thus, our framework could help the IPCC community to revise their carbon price projections obtained from climate mitigation models to make them more consistent with the role that the financial system plays.

Our framework could also support financial authorities in encouraging investors' assessment of climate-related financial risk. The IAM-CFR scenarios would limit the underestimation of financial risk in climate stress-test exercises. Accounting for the role of the financial system also has implications for criteria used by central banks to identify eligible assets in their collateral frameworks and purchasing programs. Furthermore, our results illustrate the need for financial authorities to monitor and tame the possible moral hazard of the financial system in the dynamics of the low-carbon transition. ■

REFERENCES AND NOTES

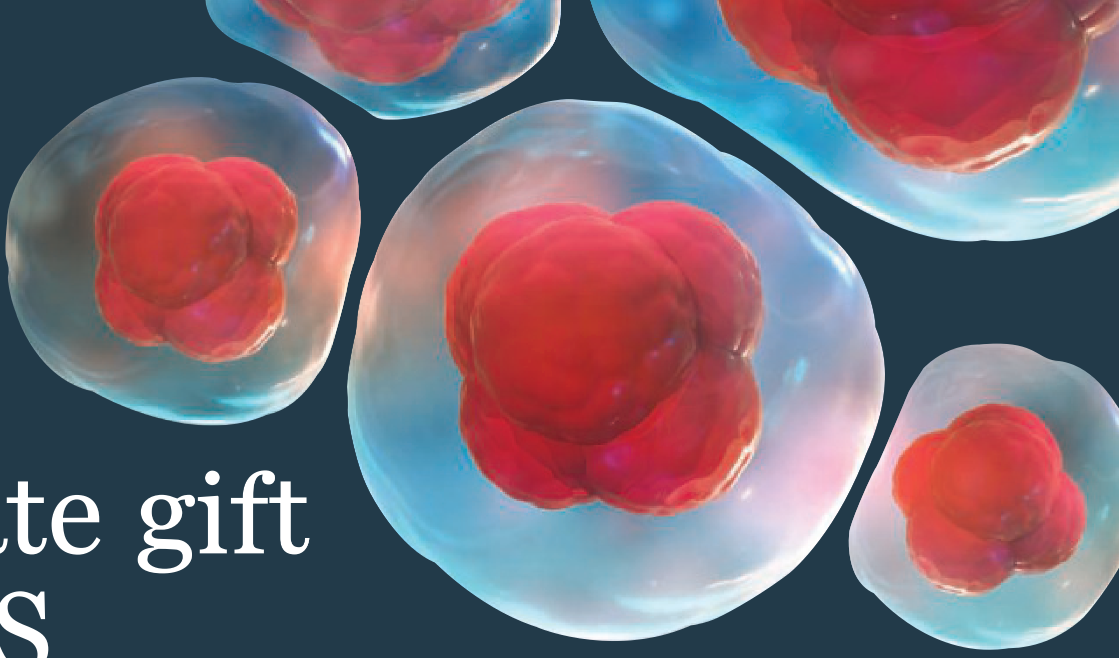
1. IPCC, "Global warming of 1.5°C" (2018); www.ipcc.ch/sr15/.
2. J. Weyant, *Rev. Environ. Econ. Policy* **11**, 115 (2017).
3. Network for Greening the Financial System (NGFS), "Guide for Supervisors Integrating climate-related and environmental risks into prudential supervision" (2020); www.ngfs.net/en/guide-supervisors-integrating-climate-related-and-environmental-risks-prudential-supervision.
4. UNEP-FI, "Decarbonisation and Disruption: Understanding the financial risks of a disorderly transition using climate scenarios" (2021); www.unepfi.org/publications/banking-publications/decarbonisation-and-disruption/.
5. D. McCollum *et al.*, *Nat. Energy* **3**, 589 (2018).
6. I. Keppo *et al*., *Environ. Res. Lett.* **16**, 053006 (2021).
7. Y. Chen *et al.*, "The MIT Economic Projection and Policy Analysis (EPPA) model: Version 5" (Joint Program Technical Note TN #16, March 2017); http://globalchange.mit.edu/publication/16620.
8. H. Pollitt, J.-F. Mercure, *Clim. Policy* **18**, 184 (2018).
9. K. S. Rogge, E. Dütschke, *Environ. Sci. Policy* **87**, 74 (2018).
10. N. Dunz, A. Naqvi, I. Monasterolo, *J. Financ. Stab*. **54**, 100872 (2021).
11. C. McGlade, P. Ekins, *Nature* **517**, 187 (2015).
12. S. Battiston, A. Mandel, I. Monasterolo, F. Schütze, G. Visentin, *Nat. Clim. Chang*. **7**, 283 (2017).
13. A. Roncoroni, S. Battiston, L. O. L. Escobar-Farfán, S. Martinez-Jaramillo, *J. Financ. Stab*. **54**, 100870 (2021).

ACKNOWLEDGMENTS

S.B. received funding from the foundation G+B Schwyzer-Winiker-Stiftung. I.M. received funding from the 11th ACRP call project GreenFin (grant agreement no. KR18ACOK14634) and from the European Union's Horizon 2020 research and innovation programme under grant 821010 (CASCADES). K.R. and B.J.v.R. received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 821471 (ENGAGE).

10.1126/science.abf3877

GRAPHIC: K. FRANKLIN/SCIENCE



BOOKS *et al.*

SCIENCE LIVES

Getting to the root of forest symbioses

An ecologist's memoir emphasizes the interdependencies of life

By **Jeremy B. Yoder**

Microbe-mediated mutual aid is key to a healthy forest, argues Suzanne Simard.

The role of mutual aid in the history of life has been a persistent countertheme to the survival of the fittest—from the earliest studies of animal sociality to the discovery that key eukaryotic organelles descend from intracellular symbionts. In *Finding the Mother Tree*, forest ecologist Suzanne Simard recounts a career spent seeking out the practical implications of life's interdependencies.

Plants' symbioses with soil microbes are textbook examples of mutual aid. Plants may have first colonized land with the help of mycorrhizal fungi, which infiltrate roots to trade water or nutrients for the sugars produced by photosynthesis. Ectomycorrhizal fungi, a subset of mycorrhizae that establish commerce without penetrating root cells, are particularly important partners for a wide diversity of tree species. For a fragile seedling, linking into the web of fungal hyphae that lace through the richest layer of soil provides a sort of surrogate root system.

Finding the Mother Tree: Discovering the Wisdom of the Forest
Suzanne Simard
Knopf, 2021. 368 pp.

Over decades of research, Simard has built a body of evidence that mycorrhizae do not simply support trees but actually connect trees to one another. Using radioactive isotopes, she has tracked the movement of nutrients through fungal linkages, both between trees of different species and between older "mother trees" and nearby seedlings of the same species. Simard argues that fungus-mediated flows of mutual aid and information let forests respond to their environments—communication and intelligence, if not as we know them.

Simard's interaction with fungal soil networks stretches back to a childhood spent exploring the woods of British Columbia as a fourth-generation descendent of homesteaders and loggers. She recalls having a nonmetaphorical taste for dirt—particularly rich humus, flavored with fallen leaves and shot through with industrious fungi. Some of the book's best passages are loving descriptions of time in the woods, drinking in the vanilla scent of ponderosa pine, tromping through thick-growing wildflowers, or burrowing into the soil to find a fungal body and trace its hyphal links from one tree to another.

Simard gives similar attention to the design and execution of key experiments, from logistical setbacks to the thrill of early results. Readers should come away with a firm grasp of the evidence underlying Simard's vision of the interconnected forest, and scientists will take inspiration from her wry descriptions of the hurdles of fieldwork.

An early job with a logging company paved the way for her to join what had become a family vocation, but Simard rapidly became disillusioned with industrial forestry. British Columbia licenses forests to loggers on the condition that they replant and ensure that replanted seedlings persist until big enough to be "free to grow" without serious competition from shrubs and trees that would otherwise spring up around them. Weeding and herbicide were prescribed to speed replanted tracts, but Simard recalls comparing sickly seedlings in a failing replanting to healthy young trees in nearby undisturbed forest: the former, barely any better rooted than when they were first put into the ground; the latter, deeply tied into the mycorrhizal web. Pushing for forestry that better parallels natural succession became her driving motivation.

That drive can sometimes give *Finding the Mother Tree* a feeling of tunnel vision. Simard's family history spans fascinating transitions in Canadian society, from the arrival of European settlers to the development of industrial forestry to an economy in which the descendants of loggers work as university professors. Yet, apart from brief references to First Nations peoples displaced by white settlers, this history goes largely unexamined. Similarly, the book's discussion of scientific responses to Simard's early discoveries mentions in passing that some objections were raised in the context of a larger scientific debate about the relative roles of competition and mutual aid in living communities, but that debate is left largely unexamined.

Nevertheless, science is as much about burrowing into a topic to trace the ramifications of a single question as it is about positioning that question within a global context. If Simard could be accused of missing the forest for the trees, her response would likely be that a forest can only be understood as it grows, one tree joining another. ■

10.1126/science.abh3999

The reviewer is at the Department of Biology, California State University Northridge, Northridge, CA 91330, USA. Email: jeremy.yoder@csun.edu

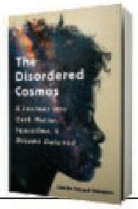
PHOTO: © DIANA MARKOSIAN

SCIENCE LIVES

Moving physics forward

Weaving together her own story with reflections on the field, a physicist calls for progress

The Disordered Cosmos: A Journey into Dark Matter, Spacetime, and Dreams Deferred
Chanda Prescod-Weinstein
Bold Type Books, 2021.
336 pp.



By **Marco Muzio**

In *The Disordered Cosmos: A Journey into Dark Matter, Spacetime, and Dreams Deferred*, physicist Chanda Prescod-Weinstein balances on a knife's edge, inspiring both awe at the elegant laws governing our Universe and fury at the field that has discovered them at great social cost. Readers will discover the fantastical realm of dark matter, quantum field theory, and curved spacetimes that modern physics has revealed, while also confronting uncomfortable truths about the social dynamics that have led to these discoveries. In a field often thought of as having a "culture of no culture," Prescod-Weinstein emerges as a salient and uncompromising voice of progress too long delayed.

From her childhood home in majority-Latinx East Los Angeles, Prescod-Weinstein would spend the 3-hour round-trip bus rides to high school regaling her peers with tales of the quarks and leptons that make up the world. Her mother, Margaret Prescod, a community organizer and activist, made sure to nurture Chanda's passion for science, taking her daughter comet hunting in Joshua Tree National Park and to see *A Brief History of Time* at age 10½.

Prescod-Weinstein left East Los Angeles for Harvard College, where she studied physics, astronomy, and astrophysics. She went on to earn a master's degree in astronomy and astrophysics at the University of California–Santa Cruz and a doctorate in physics at the University of Waterloo and the Perimeter Institute for Theoretical Physics. In her current position at the University of New Hampshire, her research focuses on cosmology, neutron stars, and dark matter. In *The Disordered Cosmos*, she undertakes a wholesale accounting of modern physics, describing the standard model of particle physics, dark matter, general relativity, and cosmology.

As a Black, Jewish, queer, agender woman, Prescod-Weinstein—who also holds a core faculty position in women's and gender studies—is uniquely suited to articulate the counterproductive, exclusionary, and often toxic aspects of academia, and readers will find phrases such as "loop quantum gravity" alongside "white supremacist ableist heterocispatriarchy." But the book spends as much time exposing readers to the realities of Prescod-Weinstein's existence in a field that never anticipated her presence as it does scrutinizing how these realities came to be.

Prescod-Weinstein explores how American and European histories have been framed and how these framings influence who receives credit for scientific progress. She also considers the implications of continuing the scientific enterprise in this mold. Why is it that we learn about so few Black and female scientists, Prescod-Weinstein wonders, for example. Is it because they are a modern creation or a historical afterthought?

The book interrogates the ways in which colonialism and the ideas of colonized peoples have benefited both science and scientists themselves throughout history. Prescod-Weinstein asks readers to reconsider, for example, the credit given to white scientists for "discoveries" gleaned from the wisdom of Indigenous communities and exposes how scientists have routinely prioritized their quest for progress over the needs of people.

Prescod-Weinstein also offers an insightful and incisive exploration into the way academic science exploits the labor of its least powerful: the underpaid graduate students who carry out the bulk of a lab's research, the minority professors who spend their nights answering emails from marginalized students looking for hope and guidance, and the custodial staff who support the scientific endeavor at the most basic level. Her own journey, however, suggests that there is little relief, even at the top. As an assistant professor, she reports that she is "tired of the disjointed feelings of liking the ideas but finding it hard to breathe in the community in which [she has] to share them."

In the end, *The Disordered Cosmos* calls for a reimagining of physics that not only realizes diversity in science and physics faculties but also creates a future where Black children can gaze at the naked stars, free of smog and city lights. The book, which is challenging and, at times, upsetting, is nonetheless a worthwhile and rewarding read that is certain to earn its place on reading lists for activists and science enthusiasts. But its intended audience—physicists themselves—may prove to be the most difficult to reach. ■

Protesters oppose the construction of the Thirty Meter Telescope on lands sacred to native Hawaiians in 2019.

PHOTO: BRUCE ASATO/*HONOLULU STAR-ADVERTISER*/AP IMAGES

The reviewer is at the Department of Physics, New York University, New York, NY 10003, USA.
Email: msm659@nyu.edu

10.1126/science.abj3436

Barbary sheep (*Ammotragus lervia*) are rare in their native region but threaten ecosystems in places where they have been introduced.

LETTERS

Edited by **Jennifer Sills**

Assisted colonization risk assessment

In their Policy Forum "Global policy for assisted colonization of species" (30 April, p. 456), J. F. Brodie and colleagues fail to acknowledge the difficulties in predicting the impacts of invasions, particularly the environmental context-dependencies that can render such forecasts unreliable (*1*), and they suggest the term "neonative" without addressing its potential misuse. Before endorsing guidance for assisted colonization, Brodie *et al.* should take these problems into account.

The challenge of accurately predicting the consequences of species translocations cannot be understated. It is doubtful that any current risk assessment technique could predict, more often than not, whether a species threatened in its native range will become invasive when introduced to another region. For example, Monterey pine (*Pinus radiata*) are endangered in California and Mexico but invasive in both Australia and New Zealand; the banteng cattle (*Bos javanicus*) are endangered in Bali but invasive in Australia; Barbary sheep (*Ammotragus lervia*) and mouflon (*Ovis orientalis*) are endangered in their native ranges but invasive on the Canary Islands (*2*).

Brodie and colleagues recommend revising the Convention on Biological Diversity policy to grant special status to populations translocated to regions beyond their native range through assisted colonization. They suggest categorizing these alien species as "neonatives," a term recently proposed for species colonizing new areas on their own in response to anthropogenic environmental change (*3*). The practical and scientific value of the term has been vigorously debated (*4*). Further conflating an already problematic term and entrenching it into international biodiversity policy would create unnecessary confusion that weakens protection against invasive alien species. It would contribute to a long, slippery slope marked by attempts to define as "native," for the purposes of policy, an alien species that someone sees as satisfying imprecise ad hoc criteria. For example, it must "naturalize ecologically" in its invaded region (*5*), develop "mutual dependence and controls" with some subset of native species (*6*), or be recognized as dangerous by native prey species (*7*). Finally, the policy would inevitably create conflict with conservation efforts to protect native biodiversity in the regions targeted for translocations—correctly described by Brodie and colleagues as "acts of invasion." This proposal seems just as likely to hinder as to enhance global conservation efforts.

Anthony Ricciardi[1]* and Daniel Simberloff[2]
[1]Redpath Museum, McGill University, Montreal, Quebec H3A 0C4, Canada. [2]University of Tennessee, Department of Ecology and Evolutionary Biology, Knoxville, TN 37996, USA.
*Corresponding author.
Email: tony.ricciardi@mcgill.ca

REFERENCES AND NOTES

1. A. Ricciardi *et al.*, *Environ. Rev.* **29**, 10.1139/er-2020-0088 (2021).
2. M. P. Marchetti, T. Engstrom, *Conserv. Biol.* **30**, 434 (2016).
3. F. Essl *et al.*, *BioScience* **69**, 908 (2019).
4. J. R. U. Wilson, *BioScience* **70**, 110 (2020).
5. N. Hettinger, *Environ. Values* **10**, 193 (2001).
6. J. Rodman, in *The Nature of Things: Language, Politics, and the Environment*, J. Bennett, W. Chaloupka, Eds. (University of Minnesota Press, 1993), pp. 139–153.
7. A. J. R. Carthey, P. B. Banks, *PLOS One* **7**, e31804 (2012).

10.1126/science.abj2682

Response

Ricciardi and Simberloff highlight the dangers of some species introductions (which are real) but do not address the concurrent (and also very real) dangers of inaction when it comes to ameliorating the impacts of climate change on biodiversity. We are not proposing what the guidance about assisted colonization should be or endorsing assisted colonization in all cases. We propose just the opposite—that a task force of experts develop guidance and best practice. Assisted colonization is going to be used and has been used (*1*). We agree that there are cases when it would not be appropriate, but maintaining a policy void will not stop people from taking action. Well-constructed policy guidance would assist conservationists and governments in making informed decisions. There are 196 governments that are Parties to the Convention on Biological Diversity, and it is far preferable to develop guidance on translocation decision-making that all of

PHOTO: STEVEN KAZLOWSKI/MINDEN PICTURES

these governments could consult than to have 196 separate approaches.

We agree with Ricciardi and Simberloff that predicting when an introduced species will become invasive is difficult. But that is all the more reason to have a structured decision-making protocol for assisted colonization—to ensure that risk is actually considered before action is taken and to take advantage of tools that conservation scientists have available when dealing with uncertain outcomes (*2*). The alternatives to regulated assisted colonization are either more unregulated translocations, which are already being conducted by interest groups in the absence of national or international policies (*1*), or no translocations at all, which locks in species extinction. Not using a conservation tool like assisted colonization is itself an active decision—a choice to do nothing (*3*)—that will inevitably have negative consequences in light of the rapid deterioration of Earth's ecosystems and novel threats facing its species. In our view, both alternatives will be more detrimental to conservation outcomes.

Ricciardi and Simberloff point to several examples of species that were threatened in their native range becoming invasive when introduced to remote islands or Australia. Notably, all of the releases they mention were made without any motivation or planning for conservation purposes, and all provide excellent examples of why a framework guiding assisted colonization is so important. It is exceedingly unlikely that a formal structured decision-making protocol for assisted colonization would have approved any of these translocations; we know of no serious proposals that would transgress major biogeographical barriers such as oceans or continents, let alone that would bring organisms to remote archipelagos already known to be highly susceptible to invasions.

Ricciardi and Simberloff make a fair point that certain uses of the term "neonative" in other contexts have been problematic. But the binary distinction of "native" versus "alien" is arguably even more problematic, as it does not account for the inherently and fundamentally dynamic nature of species distributions and community composition—dynamism that is only increasing in the face of rapid climate change.

Jedediah F. Brodie[1,2]*, Susan Lieberman[3], Axel Moehrenschlager[4,5,6,7], Kent H. Redford[8,9,10], Jon Paul Rodríguez[11,12], Mark Schwartz[13], Philip J. Seddon[14], James E. M. Watson[15,16]

[1]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA. [2]Wildlife Biology Program, University of Montana, Missoula, MT 59812, USA. [3]Wildlife Conservation Society, Bronx, NY 10460 USA. [4]International Union for Conservation of Nature (IUCN) Species Survival Commission Conservation Translocation Specialist Group, Calgary, AB T2E 7V6, Canada. [5]Centre for Conservation Research, Calgary Zoological Society, Calgary, AB T2E 7V6, Canada. [6]Department of Forestry and Environmental Conservation, Clemson University, Clemson, SC 29634, USA. [7]Wildlife Conservation Research Unit, Oxford University, Oxford, UK. [8]Archipelago Consulting, Portland, ME 04112, USA. [9]Department of Environmental Studies, University of New England, Biddeford, ME 04005, USA. [10]Griffith University, Nathan, QLD 4111, Australia. [11]IUCN Species Survival Commission, Caracas, Venezuela. [12]Venezuelan Institute for Scientific Investigation and Provita, Caracas, Venezuela. [13]Department of Environmental Science & Policy, University of California, Davis, CA 95616, USA. [14]Department of Zoology, University of Otago, Dunedin 9054, New Zealand. [15]School of Earth and Environmental Sciences, The University of Queensland, St. Lucia, QLD 4072, Australia. [16]Centre for Biodiversity and Conservation Science, The University of Queensland, St. Lucia, QLD 4071, Australia.
*Corresponding author.
Email: jedediah.brodie@umontana.edu

REFERENCES AND NOTES

1. M. W. Schwartz, T. G. Martin, *Ann. N.Y. Acad. Sci.* **1286**, 15 (2013).
2. M. A. McCarthy, *Ann. N.Y. Acad. Sci.* **1322**, 77 (2014).
3. R. Phelan *et al.*, *Conserv. Sci. Pract.* **3**, e371 (2021).

10.1126/science.abj3134

Inclusion and equity through STEM training

Structural racism, negative stereotypes, and lack of mentorship are pervasive in science, technology, engineering, and mathematics (STEM) disciplines (*1–3*), with negative effects on the recruitment, retention, and professional development of scholars from racial and ethnic minorities (*2*). STEM practitioners would benefit from addressing inequity in their workplaces. To create a culture of inclusion, STEM fields should reexamine the scope of scientific training.

Scientific undergraduate and graduate programs should raise students' awareness about the importance of diversity, equity, and inclusion in science labs and departments by incorporating the national and global discourse on racial justice into their training. Currently, STEM training provides few incentives (and sometimes disincentives) for investing time in such efforts. STEM students are less likely than their peers in non-STEM disciplines to vote and engage in political processes (*4*). Requiring coursework in the humanities, philosophy, ethics, and social sciences would better prepare STEM students to appreciate the complexity of issues surrounding structural racism and colonialism in academia (*5*, *6*) and empower them to support more inclusive policies in their own institutions (*7*, *8*).

A broad curriculum would also help scientists and engineers produce more effective science (*7*, *8*). Meeting scientific goals requires an exploration of the root causes of the problems under investigation and an understanding of the broader implications of potential results. Improving the poor health outcomes of COVID-19 among Black and Hispanic or Latino communities (*9*), for example, requires an understanding of how the health system continues to fail those who are already suffering and vulnerable. Ethical research extends beyond issues of informed consent to the risks of inequity and unethical use of results. A well-rounded scientific education would lay the groundwork for students who become professional scientists or administrators to grapple with the fact that research funded by public money may lead to therapeutics that only the very rich can afford or that software systems designed by university scholars may be used by authoritarian regimes to target an ethnic minority.

Recent programs have started to integrate artificial intelligence and computer science education with philosophy and ethics [e.g., (*10*, *11*)]. These initiatives should be expanded to all STEM fields. Scientists, engineers, and mathematicians deserve access to the tools and resources required to create a more equitable scientific enterprise and, in turn, better and more ethical science.

Muhammad Hamid Zaman
Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.
Email: zaman@bu.edu

REFERENCES AND NOTES

1. E. O. McGee, W. H. Robinson, Eds., *Diversifying STEM: Multidisciplinary Perspectives on Race and Gender* (Rutgers University Press, 2020).
2. E. O. McGee, *Black, Brown, Bruised: How Racialized STEM Education Stifles Innovation* (Harvard Education Press, 2020).
3. N. Forrester, *Nature* **585**, S65 (2020).
4. N. Thomas *et al.*, "Democracy counts: A report on U.S. college and university student voting" (Institute for Democracy & Higher Education, Jonathan M. Tisch College of Civic Life, Tufts University, 2017).
5. R. D. Roy, "Science still bears the fingerprints of colonialism," *Smithsonian Magazine* (2018).
6. K. Norton, "Where science and social justice meet," *PBS* (2020).
7. J. Horgan, "Why STEM students need humanities courses," *Scientific American* (2018).
8. D. Albert, "Ten important reasons to include the humanities in your preparation for a scientific career," *Science* (2011); https://blogs.sciencemag.org/sciencecareers/2011/05/ten-important-r.html.
9. W. F. Marshall, "Coronavirus infection by race: What's behind the health disparities?", *Mayo Clinic* (2020).
10. Georgetown University, "Ethics to become key part of Georgetown's computer science curriculum" (2019).
11. P. Karoff, "Embedding ethics in computer science curriculum," *The Harvard Gazette* (2019).

10.1126/science.abj1410

AAAS makes science relatable through diverse efforts

Policy, education, and outreach projects make research relevant

By **Becky Ham**

When the constitution of the American Association for the Advancement of Science was revised in 1946, its statement of objectives contained new language: "…to increase public understanding and appreciation of the importance and promise of the methods of science in human progress."

The association has since fulfilled that charge in diverse sectors, including policy, education, and public engagement, to make science more relatable and relevant to the public.

Making science relatable also requires a variety of engagement strategies, including facilitating in-depth discussions with local policy leaders, translating technical language into digestible summaries for the classroom, and promoting science role models.

In the case of the AAAS Center for Scientific Evidence in Public Issues or EPI Center, for instance, a successful part of bringing clear and actionable scientific advice to policy-makers has been encouraging discussions among a broad group of experts and policy peers.

During meetings organized by the EPI Center this year, city council members, mayors, water engineers, and local utility managers joined scientists to discuss per- and polyfluoroalkyl substances or PFAS, synthetic chemicals found in drinking water systems. At least two PFAS have been associated with increased rates of some cancers and thyroid disease.

The EPI Center provides nontechnical syntheses of topics for policy-makers, "but one thing we have seen is that examples from their peers that have implemented and used the scientific evidence are much more valuable and easier to understand," said Kathryn McGrath, communications director for the center.

Teachers at a 2018 Science in the Classroom workshop.

Whether the focus is clean water or voting technology or hydraulic fracturing, the EPI Center strives to make the science of these topics relatable by talking with the public and policy-makers to find out exactly what information would be helpful for them. The discussions allow city council members, for instance, "to ask the science experts what they need to know to go back to their communities and regions and take action on some of these issues," McGrath said.

AAAS's Local Science Engagement Network, a grassroots platform that nurtures local and state science advocates for climate and energy policy, has also found success with local partnerships. In Colorado, Missouri, and Georgia, LSENs work with organizations in each state that "have a good sense of policy landscapes as well as the cultural and scientific landscapes in those areas," said Daniel Barry, local and state advocacy director and head of LSEN at AAAS.

LSENs offer an avenue for engagement and advocacy that AAAS members have been asking for, by connecting scientists with their own elected representatives on the local, state, and federal levels. As both constituents and neutral, honest brokers of scientific information, LSEN participants can be a key resource when legislatures grapple with the more local implications of climate change, such as modernizing the state power grid, said Barry. "They can step up and say, 'Science, that's what I do, and I live here in this community. I know how to get you the science you need.'"

LSEN members also condense technical research into locally relevant analyses in plain English for business leaders and citizens. So far in 2021, Missouri LSEN partner MOST Policy Initiative has produced more than 80 such "science notes" about pending state legislation.

Among AAAS's numerous education efforts to make science more relevant is Science in the Classroom, an initiative that annotates and provides additional resources to accompany research papers from the *Science* family of journals. The goal is to make scientific papers more accessible to high school, community college, and undergraduate students, while putting a face on the papers' authors in communities with little exposure to working scientists, said program director Suzanne Thurston.

The popular resource had more than 1 million page views in the past 3 years, and the hunger for accessible scientific content during a pandemic year led to a 50% increase in total site visits in 2020 compared to 2019. The program also offers professional development workshops to educators, researchers, and annotators. By showcasing a range of authors and annotators, Science in the Classroom helps "to expose students to diversity within STEM and demonstrates what 'actual living scientists' look like," said Thurston, who serves as a program director in AAAS's Inclusive STEM Ecosystems for Equity and Diversity (ISEED).

The IF/THEN Ambassador program, led by AAAS's Center for Public Engagement with Science and Technology, was another recent effort to show off the diverse faces of science, by highlighting 125 women in STEM as role models for middle school girls.

Lyda Hill Philanthropies, which funds the IF/THEN initiative, wanted to work with AAAS on the ambassador program after the association's success with other public engagement initiatives such as the AAAS Mass Media Science & Engineering Fellowship and the Leshner Leadership Institute for Public Engagement with Science, said Emily Therese Cloyd, director of the AAAS Center for Public Engagement with Science and Technology.

The ambassador program was distinguished by its emphasis on increasing visibility for women in STEM who demonstrate how science is involved in everyday careers beyond the traditional lab, said Cloyd. "We're moving beyond scientists who work at an academic institution and thinking about the ways that a video game designer or a fashion designer might be using STEM every day."

AAAS is committed to making science relatable and relevant for everyone from policy-makers to educators to students. It is at the core of the organization's mission and will continue to be a top priority for years to come.

PHOTO: AAAS

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

ECOSYSTEM SENTINELS

Sampling seabirds

The vastness of the worlds' oceans makes them difficult to monitor. Seabirds that forage and breed across oceans globally have been recognized as sentinels of ocean health. Sydeman *et al.* looked across seabird species of both the Northern and Southern Hemispheres and found varying patterns. Northern Hemisphere species exhibited greater signs of stress and reduced breeding success, indicative of low fish resources. Southern Hemisphere species showed less impact on reproductive output, suggesting that the fish populations there have thus far been less disturbed. The differences across hemispheres indicate different strategies for conservation, with active recovery needed in the north and enhanced protection in the south. —SNV *Science*, abf1772, this issue p. 980

Climate change and human activities are having greater effects on seabird species in the Northern Hemisphere, such as crested auklets along the northern Pacific coast.

PHOTO: DESIGN PICS INC./ALAMY STOCK PHOTO

IMMUNOLOGY

Ligand-patterned immune synapses

To execute virus-infected and tumor cells, natural killer (NK) cells form immune synapses with target cells through which they release cytotoxic granules, processes that require the receptors LFA-1 and CD16. Verron *et al*. generated artificial immune synapses containing different patterns of ligands for LFA-1 and CD16 (see the Focus by Veillette). Their results suggested that a uniform distribution of LFA-1 and CD16 ligands promoted immune synapse stability and that degranulation occurred only when both the LFA-1 and CD16 ligands were present. —AMV

Sci. Signal. **14**, eabe2740, eabi8525 (2021).

QUANTUM CRITICALITY

Pervasive criticality

Iron-based superconductors are believed to host a quantum critical point (QCP), a zero-temperature phase transition, beneath the "dome" delineating the superconducting phase. Elucidating the nature of this QCP is, however, tricky. Worasaran *et al*. set out to do just that in a prototypical iron-based superconductor, barium iron arsenide. By applying strain to their samples, the researchers found power-law behaviors that are characteristic of nematic quantum criticality. The associated quantum fluctuations were present over a large portion of the phase diagram. This method may be useful in studying quantum criticality in other material systems. —JS

Science, abb9280, this issue p. 973

GENOMIC EVOLUTION

Organismal evolution of the 3D genome

The conformation of chromosomes within the nucleus can reflect a cell's type or state. However, studies of the conservation and evolutionary history of the mechanisms regulating genome structure across species are lacking. Hoencamp *et al*. mapped three-dimensional (3D) genome organization in 24 eukaryote species, including animals, fungi, and plants. At interphase, species' telomeres and centromeres either clustered across chromosomes or oriented in a polarized state maintaining individual chromosomal territories within the cell, a difference attributed to condensin II. An experimental loss of condensin II in human cells promotes the formation of centromere clusters but has no effect on loop or compartment formation. Whether the structure of the 3D genome varies across species may thus depend on whether they carry a functional condensin II gene. —LMZ

Science, abe2218, this issue p. 984

CELLULAR MICROBIOLOGY

Bacterial effectors manipulate membranes

Many pathogenic bacteria use molecular syringes to translocate proteins called effectors into the host cell to hijack the cellular machinery for their proliferation. *Legionella pneumophila*, the causative bacteria of Legionnaires' disease, uses a large effector arsenal and harnesses the host membrane system to establish a specialized vacuole where it replicates. Hsieh *et al*. show that, within this effector arsenal, the phospholipid kinase MavQ and the phosphatase SidP work together and self-organize on the intracellular membrane network of its eukaryotic host to promote membrane remodeling. The interactions between MavQ and SidP constitute positive and negative feedback loops, respectively, that orchestrate their spatiotemporal oscillation during infection. —SMH

Science, aay8118, this issue p. 935

BIOMATERIALS

Topography controls the T cell response

Biomaterials are regularly implanted throughout the body, and biomaterial structural properties can alter the associated tissue-healing response around the implant. To better understand this process, Hu *et al.* created biomaterial membranes with varied surface topography using electrospinning. When they examined the microenvironment around each membrane at the single-cell level in rodents, the T cell response occurred earlier in the most aligned scaffolds, and the T cells appeared to modulate the overall healing response. Although further assessment of these materials is needed to better understand the signaling pathways involved, the single-cell analysis approach used in this study could inform the design of future biomaterials to improve wound healing and tissue integration. —JST

Sci. Adv. 10.1126/sciadv.abf0787 (2021).

QUANTUM COMPUTING

Simulating quantum walkers

Quantum walks are the quantum mechanical analogs of classical random walks, describing the propagation of a quantum walker across a lattice, and find application in developing algorithms for simulating quantum many-body systems. Gong *et al*. used an 8-by-8 two-dimensional (2D) superconducting qubit square lattice containing 62 functional qubits to show how multiple (two) walkers traverse a 2D qubit array, interfering as they go. The authors were also able to program the paths that the walkers follow, demonstrating a Mach-Zehnder interferometer in which a single or multiple quantum walkers coherently traverse two paths before interfering and exiting at a single port. The results illustrate the potential for superconducting-based quantum processors in simulating large-scale quantum systems. —ISO

Science, abg7812, this issue p. 948

FERROELECTRICS

Imprinting oxides

Dislocations can be problematic for the properties of functional oxides and are often avoided as a result. Höfling *et al*. found that introducing a network of dislocations to barium titanate actually enhanced the dielectric and piezoelectric properties. The authors introduced the dislocation network with uniaxial compression, which forced the material to have a domain structure that enhanced the piezoelectric coefficient by a factor of 19. This strategy should be a useful tool for optimizing properties of other functional oxides. —BG

Science, abe3810, this issue p. 961

IN OTHER JOURNALS

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

EDUCATION

Bringing cultural awareness to mentoring

Mentoring is a critical component for retaining diverse talent in the STEM field. However, despite their best intentions, mentors may be unaware of how racial or ethnic differences may influence their mentees' research experience. To further investigate this potential disparity, Byars-Winston and Butz developed a scale to assess cultural diversity awareness related to race and ethnicity in research-mentoring relationships informed by multicultural counseling theory and social cognitive theory. Instrument validity yielded a three-factor mentor scale assessing attitudes, behavior, and confidence and a two-factor mentee scale assessing attitudes and behavior. This scale can be used as a self-assessment to prompt mentors' reflection on their mentoring practices and to encourage new ways of acknowledging and appreciating cultural diversity in their research-mentoring relationships. —MMc

CBE Life Sci. Educ. 10.1187/cbe.19-06-0127 (2021).

CLIMATE CHANGE

Tusk records

Understanding how climate change may affect Arctic species is challenging given that change has been gradual. The impact of environmental transition is reflected in dietary shifts of species at upper trophic levels. Narwhals provide a unique opportunity to track wider ecological change in the Arctic because modifications in their diets can be detected using isotope analysis of the dentine deposited over their lifetimes in their elongated tusks. Dietz *et al.* measured isotope ratios in the tusks of 10 male narwhals collected between 1962 and 2010. The data revealed patterns consistent with dietary shifts from ice-associated (sympagic) to open-water (pelagic) food species over that time. Further,

PHOTO: NOAA PHOTO LIBRARY/FLICKR/CC BY

MARINE ECOLOGY

Seagrass offsets acidification

Gradual acidification of the world's oceans is driven by the uptake of carbon dioxide. This feature of contemporary climate change has potential consequences for marine life and the functioning of marine ecosystems. Ricart *et al.* show that acidification can be locally slowed or ameliorated by seagrass meadows, where uptake of carbon dioxide by the plants exceeds that produced by respiration. Along 1000 kilometers of Californian coastal waters and measured over 6 years, pH was elevated in most of the seagrass sites examined compared with adjacent sites. These findings add to the potentially beneficial suite of effects of the presence of seagrasses and macroalgae in coastal waters and indicate possible routes for the management of acidification in these systems. —AMS

Glob. Change Biol. **27**, 2580 (2021).

Seagrass beds consume carbon dioxide and thereby reduce local seawater acidification.

mercury levels were found to increase with the trophic level of prey, as might be expected. However, in recent decades, mercury levels in narwhals' tissues rose sharply, possibly reflecting an environmental source–related change. —SNV

Curr. Biol. **31**, 2012 (2021).

CLINICAL PSYCHOLOGY

Culture and posttraumatic stress

It is unclear whether posttraumatic stress disorder (PTSD) is a universal response to violence found everywhere or if it is culturally specific to certain parts of the world. Zefferman and Mathew interviewed warriors among the Turkana, a population of subsistence pastoralists living in Kenya. Compared with a sample of American military servicemembers who had been treated for PTSD, Turkana were equally likely to experience reactive symptoms such as hypervigilance, which may be more sensitive to experiences of danger, but they were less likely to experience depressive symptoms such as detachment and loss of interest, which may be related to feelings of moral violation. These findings suggest that symptoms of PTSD directly tied to dangers of combat may be universal, whereas the symptoms tied to the morality of combat may be more culturally variable. —TSR

Proc. Natl. Acad. Sci. U.S.A. **118**, e2020430118 (2021).

A warrior from the Turkana tribe of northern Kenya

PHOTO: CHRISTOPHER FURLONG/STAFF/GETTY IMAGES

MICROBIOLOGY

Actin to trap bacteria

Salmonella bacteria cause tens of millions of cases of food-related illnesses every year. On infecting host cells, the bacterium injects virulence factors that induce rearrangements of the cytoskeleton to allow internalization inside a *Salmonella*-containing vacuole (SCV). Here, the bacteria replicate protected from the host immune system. Hahn *et al.* show that although the actin network is hijacked to allow bacterial entry, it is also part of a cellular defense system. Through proteomics analysis, the authors identified the kinase SIK2 as a central player in coordinating actin defenses. During bacterial infection, SIK2 is recruited to the SCV together with proteins involved in actin polymerization. SIK2 depletion allows the escape of bacteria from the SCV, which results in collapse of cortical actin structures. Thus, SIK2 coordinates an actin network that limits intracellular proliferation of *Salmonella*. —VV

Proc. Natl. Acad. Sci. U.S.A. **118**, e2024144118 (2021).

AGING

Live fast, die young

Aging sees many physiological changes. Studies are revealing the dynamics of the cells and molecular factors that contribute to the aging process. For example, chronological age can be estimated by comparing DNA methylation status. Anderson *et al.* measured DNA methylation in the baboons of Kenya's Amboseli National Park to find out whether epigenetic changes associated with aging are affected by an animal's social environment in the wild. Early adversity and social bonds did not apparently affect the epigenetic clock, but there was a signal from male social status and competitiveness. By measuring the accumulation of epigenetic markers, high-ranking dominant males appeared older than their chronological ages. —BAP

eLife **10**, e66128 (2021).

POLYMER CHEMISTRY

Light frees a reactive thiol

Postsynthetic modification of polymers can be aided by the release of a caged reactive group after synthesis. Rodrigues *et al*. report on the photorelease of a thiol group that can undergo Michael addition. Visible light causes *o*-thiopyrinidylbenzaldehyde to undergo a ring opening that exposes a reactive aromatic thiol group that can undergo Michael reaction with electron-deficient alkenes and alkynes in a variety of solvents. They show that a polyethylene glycol bearing a terminal alkyne group could be ligated through an esterification reaction. —PDS

J. Am. Chem. Soc. 10.1021/jacs.1c03213 (2021).

REVIEW SUMMARY

ADDITIVE MANUFACTURE

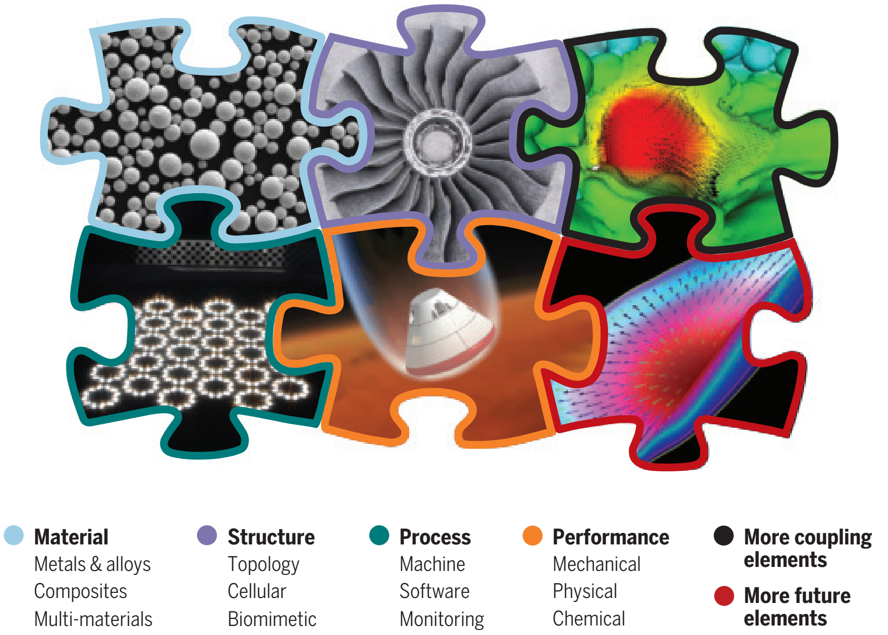
Material-structure-performance integrated laser-metal additive manufacturing

Dongdong Gu*, Xinyu Shi, Reinhart Poprawe, David L. Bourell, Rossitza Setchi, Jihong Zhu

BACKGROUND: Metallic components are the cornerstone of modern industries such as aviation, aerospace, automobile manufacturing, and energy production. The stringent requirements for high-performance metallic components impede the optimization of materials selection and manufacturing. Laser-based additive manufacturing (AM) is a key strategic technology for technological innovation and industrial sustainability. As the number of applications increases, so do the scientific and technological challenges. Because laser AM has domain-by-domain (e.g., point-by-point, line-by-line, and layer-by-layer) localized forming characteristics, the requisite for printing process and performance control encompasses more than six orders of magnitude, from the microstructure (nanometer- to micrometer-scale) to macroscale structure and performance of components (millimeter- to meter-scale). The traditional route of laser-metal AM follows a typical "series mode" from design to build, resulting in a cumbersome trial-and-error methodology that creates challenges for obtaining high-performance goals.

ADVANCES: We propose a holistic concept of material-structure-performance integrated additive manufacturing (MSPI-AM) to cope with the extensive challenges of AM. We define MSPI-AM as a one-step AM production of an integral metallic component by integrating multimaterial layout and innovative structures, with an aim to proactively achieve the designed high performance and multifunctionality. Driven by the performance or function to be realized, the MSPI-AM methodology enables the design of multiple materials, new structures, and corresponding printing processes in parallel and emphasizes their mutual compatibility, providing a systematic solution to the existing challenges for laser-metal AM. MSPI-AM is defined by two methodological ideas: "the right materials printed in the right positions" and "unique structures printed for unique functions." The increasingly creative methods for engineering both micro- and macrostructures within single printed components have led to the use of AM to produce more complicated structures with multimaterials. It is now feasible to design and print multimaterial components with spatially varying microstructures and properties (e.g., nanocomposites, in situ composites, and gradient materials), further enabling the integration of functional structures with electronics within the volume of a laser-printed monolithic part. These complicated structures (e.g., integral topology optimization structures, biomimetic structures learned from nature, and multiscale hierarchical lattice or cellular structures) have led to breakthroughs in both mechanical performance and physical/chemical functionality. Proactive realization of high performance and multifunctionality requires cross-scale coordination mechanisms (i.e., from the nano/microscale to the macroscale).

OUTLOOK: Our MSPI-AM continues to develop into a practical methodology that contributes to the high performance and multifunctionality goals of AM. Many opportunities exist to enhance MSPI-AM. MSPI-AM relies on a more digitized material and structure development and printing, which could be accomplished by considering different paradigms for AM materials discovery with the Materials Genome Initiative, standardization of formats for digitizing materials and structures to accelerate data aggregation, and a systematic printability database to enhance autonomous decision-making of printers. MSPI-oriented AM becomes more intelligent in processes and production, with the integration of intelligent detection, sensing and monitoring, big-data statistics and analytics, machine learning, and digital twins. MSPI-AM further calls for more hybrid approaches to yield the final high-performance/multifunctional achievements, with more versatile materials selection and more comprehensive integration of virtual manufacturing and real production to navigate more complex printing. We hope that MSPI-AM can become a key strategy for the sustainable development of AM technologies. ■



Material-structure-performance integrated additive manufacturing (MSPI-AM). Versatile designed materials and innovative structures are simultaneously printed within an integral metallic component to yield high performance and multifunctionality, integrating in parallel the core elements of material, structure, process, and performance and a large number of related coupling elements and future potential elements to enhance the multifunctionality of printed components and the maturity and sustainability of laser AM technologies.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dongdonggu@nuaa.edu.cn

Cite this article as D. Gu *et al.*, *Science* 372, eabg1487 (2021). DOI: 10.1126/science.abg1487

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg1487

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Multiscale representation of very large environments in the hippocampus of flying bats

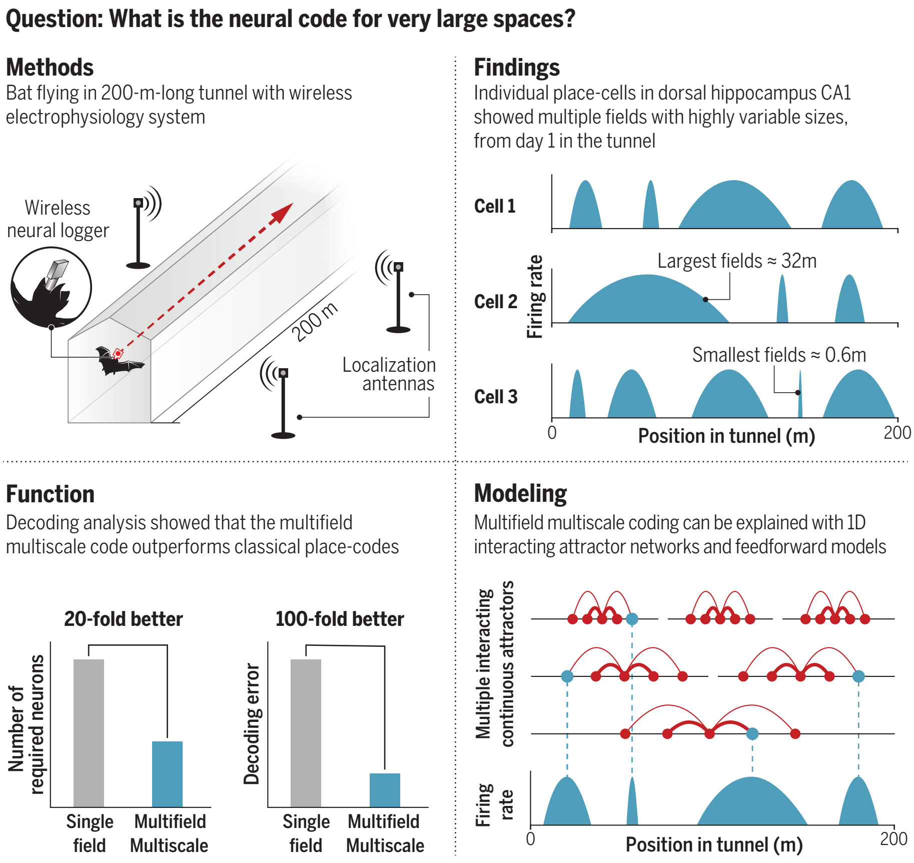
Tamir Eliav†, Shir R. Maimon†, Johnatan Aljadeff, Misha Tsodyks, Gily Ginosar, Liora Las, Nachum Ulanovsky*

INTRODUCTION: Place cells are neurons in the hippocampus that represent the animal's position in space and are important for supporting navigation behaviors. These cells increase their spiking activity when the animal passes through a specific region of space, called the neuron's "place field." Since the discovery of place cells half a century ago, nearly all the research on spatial representations in the mammalian brain has focused on rats and mice as animal models and used small laboratory environments as experimental setups—usually small boxes or short linear tracks ~1 to 2 m in size. In such small environments, individual place cells typically have one place field, with a small field size. However, outdoor navigation of all mammals occurs in natural environments that span much larger spatial scales, of hundreds of meters or kilometers, and nothing is known about the neural codes for such large spatial scales.

RATIONALE: We reasoned that in very large environments, the hippocampus must exhibit a different coding scheme than seen in small environments because large environments cannot be tiled fully by the limited number of hippocampal neurons. We set out to discover this alternative coding scheme and thus to close the longstanding gap between the neurobiology of navigation as studied in the laboratory and natural large-scale navigation. To this end, we studied bats flying in a 200-m-long tunnel while we recorded the activity of hippocampal dorsal CA1 neurons using a custom wireless-electrophysiology system.

RESULTS: We found that place cells recorded in the large environment exhibited a multifield, multiscale representation of space: Individual neurons exhibited multiple place fields of diverse sizes, ranging from <1 m to more than 30 m, and the fields of the same neuron could differ up to 20-fold in size. This multifield, multiscale code was observed already from the first day in the environment and was similar between wild-born and laboratory-born bats that were never exposed to large environments. By contrast, recordings from a small-scale 6-m environment did not reveal such a multiscale code but rather classical single fields. Theoretical decoding analysis showed major advantages of the multiscale code over classical single-field codes, both in the number of required neurons and in the decoding errors. Thus, the multiscale code provides an efficient population code with a high capacity for representing very large environments. We conducted neural-network modeling, which suggested that the multiscale code may arise from interacting attractor networks with multiple scales or from feedforward networks, which yielded experimentally testable predictions for the inputs into CA1.

CONCLUSION: Using this experimental setup, our study uncovered a new coding scheme for large spaces, which was never observed before in small spaces: a multiscale code for space. This coding scheme existed from day 1 in the environment and was observed in both wild-born and laboratory-born bats, suggesting that it does not require previous experience. These findings provide a new notion for how the hippocampus represents space. The large naturalistic scale of our experimental environment was crucial for revealing this type of code. More generally, this study demonstrates the power of studying brain circuits under naturalistic conditions. ■


Question: What is the neural code for very large spaces?
Methods
Bat flying in 200-m-long tunnel with wireless electrophysiology system
Wireless neural logger
200 m
Localization antennas
Findings
Individual place-cells in dorsal hippocampus CA1 showed multiple fields with highly variable sizes, from day 1 in the tunnel
Cell 1
Cell 2
Cell 3
Firing rate
Largest fields ≈ 32m
Smallest fields ≈ 0.6m
0
200
Position in tunnel (m)
Function
Decoding analysis showed that the multifield multiscale code outperforms classical place-codes
20-fold better
100-fold better
Number of required neurons
Decoding error
Single field
Multifield Multiscale
Modeling
Multifield multiscale coding can be explained with 1D interacting attractor networks and feedforward models
Multiple interacting continuous attractors
Firing rate
0
200
Position in tunnel (m)

Multiscale hippocampal spatial code for very large environments. (Methods) We wirelessly recorded neural activity from hippocampal neurons of bats flying in a 200-m tunnel. (Findings) Single neurons exhibited multiple place fields with highly heterogeneous field sizes for the same neuron. (Function) This multiscale neural code for space strongly outperforms classical single-field place codes. (Modeling) Modeling by using interacting attractor networks and feedforward models recapitulated the multiscale coding.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: nachum.ulanovsky@weizmann.ac.il
†These authors contributed equally to this work.

Cite this article as T. Eliav *et al.*, *Science* **372**, eabg4020 (2021). DOI: 10.1126/science.abg4020

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg4020

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile

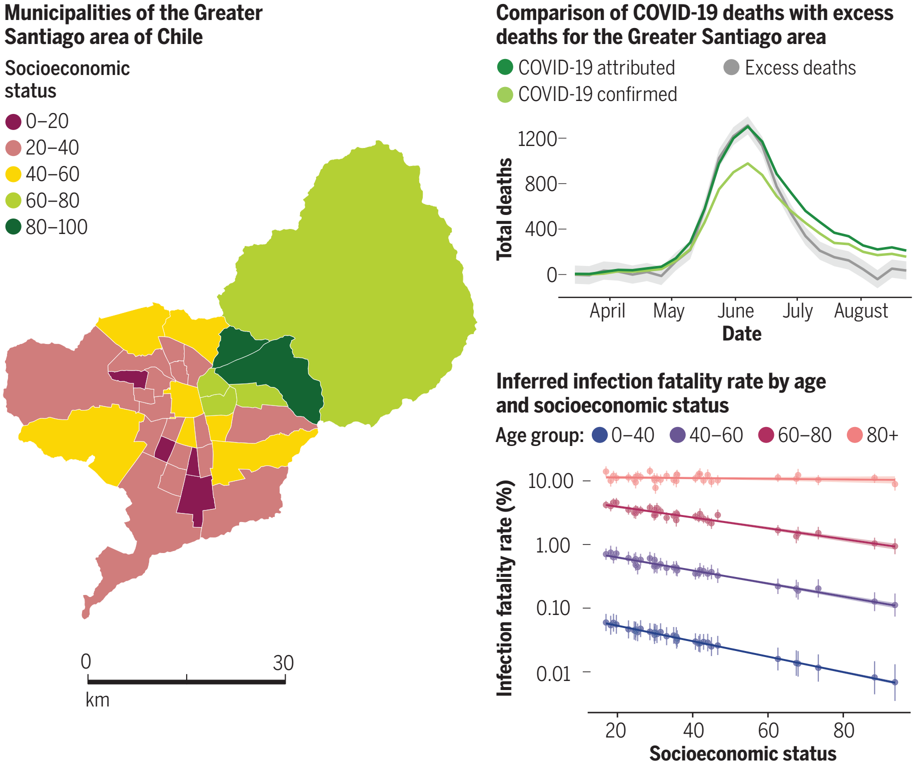
Gonzalo E. Mena†*, Pamela P. Martinez†*, Ayesha S. Mahmud, Pablo A. Marquet, Caroline O. Buckee‡, Mauricio Santillana‡

INTRODUCTION: The COVID-19 crisis has exposed major inequalities between communities. Understanding the societal risk factors that make some groups particularly vulnerable is essential to ensure more effective interventions for this and future pandemics. Here, we focus on socioeconomic status as a risk factor. Although it is broadly understood that social and economic inequality has a negative impact on health outcomes, the mechanisms by which socioeconomic status affects disease outcomes remain unclear. These mechanisms can be mediated by a range of systemic structural factors, such as access to health care and economic safety nets. We address this gap by providing an in-depth characterization of disease incidence and mortality and their dependence on demographic and socioeconomic strata in Santiago, a highly segregated city and the capital of Chile.

RATIONALE: Combining publicly available data sources, we conducted a comprehensive analysis of case incidence and mortality during the first wave of the pandemic. We correlated COVID-19 outcomes with behavioral and health care system factors while studying their interaction with age and socioeconomic status. To overcome the intrinsic biases of incomplete case count data, we used detailed mortality data. We developed a parsimonious Gaussian process model to study excess deaths and their uncertainty and reconstructed true incidence from the time series of deaths with a new regularized maximum likelihood deconvolution method. To estimate infection fatality rates by age and socioeconomic status, we implemented a hierarchical Bayesian model that adjusts for reporting biases while accounting for incompleteness in case information.

RESULTS: We find robust associations between COVID-19 outcomes and socioeconomic status, based on health and behavioral indicators. Specifically, we show in lower–socioeconomic status municipalities that testing was almost absent early in the pandemic and that human mobility was not reduced by lockdowns as much as it was in more affluent locations. Test positivity and testing delays were much higher in these locations, indicating an impaired capacity of the health care system to contain the spread of the epidemic. We also find that 73% more deaths than in a normal year were observed between May and July 2020, and municipalities at the lower end of the socioeconomic spectrum were hit the hardest, both in relation to COVID-19–attributed deaths and excess deaths. Finally, the socioeconomic gradient of the infection fatality rate appeared particularly steep for younger age groups, reflecting worse baseline health status and limited access to health care in municipalities with low socioeconomic status.

CONCLUSION: Together, these findings highlight the substantial consequences of socioeconomic and health care disparities in a highly segregated city and provide practical methodological approaches useful for characterizing the COVID-19 burden and mortality in other urban centers based on public data, even if reports are incomplete and biased. ■



Effect of socioeconomic inequalities on COVID-19 outcomes. The map on the left shows the municipalities that were included in this study, colored by their socioeconomic status score. For the comparison between COVID-19 deaths and excess deaths (top right), COVID-19–confirmed deaths are shown in light green and COVID-19–attributed deaths in dark green. Excess deaths, shown in gray, correspond to the difference between observed and predicted deaths. Predicted deaths were estimated using a Gaussian process model. The shading indicates 95% credible intervals for the excess deaths. The infection fatality rates (bottom right) were inferred by implementing a hierarchical Bayesian model, with vertical lines representing credible intervals by age and socioeconomic status.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: gonzalo.mena@stats.ox.ac.uk (G.E.M.); pamelapm@illinois.edu (P.P.M.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

Cite this article as G. E. Mena *et al.*, *Science* 372, eabg5298 (2021). DOI: 10.1126/science.abg5298

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg5298

RESEARCH ARTICLES

CELLULAR MICROBIOLOGY

Dynamic remodeling of host membranes by self-organizing bacterial effectors

Ting-Sung Hsieh[1]†, Victor A. Lopez[1]†, Miles H. Black[1]†, Adam Osinski[1], Krzysztof Pawłowski[1,2], Diana R. Tomchick[3,4], Jen Liou[5], Vincent S. Tagliabracci[1,6,7]*

During infection, intracellular bacterial pathogens translocate a variety of effectors into host cells that modify host membrane trafficking for their benefit. We found a self-organizing system consisting of a bacterial phosphoinositide kinase and its opposing phosphatase that formed spatiotemporal patterns, including traveling waves, to remodel host cellular membranes. The *Legionella* effector MavQ, a phosphatidylinositol (PI) 3-kinase, was targeted to the endoplasmic reticulum (ER). MavQ and the *Legionella* PI 3-phosphatase SidP, even in the absence of other bacterial components, drove rapid PI 3-phosphate turnover on the ER and spontaneously formed traveling waves that spread along ER subdomains inducing vesicle and tubule budding. Thus, bacteria can exploit a self-organizing membrane-targeting mechanism to hijack host cellular structures for survival.

Intracellular bacterial pathogens use a variety of strategies for survival and proliferation inside the host cell, and many manipulate the host endomembrane system (*1*). Because phosphoinositides function as identifiers for membrane-bound organelles and play critical roles in vesicle trafficking, they are frequently targeted by these pathogens, including *Legionella pneumophila*, the causative agent of Legionnaires' disease. Two hallmarks of host cell subversion by *L. pneumophila* are reprogramming of membrane trafficking via phosphoinositide conversion and hijacking of the host endoplasmic reticulum (ER) membrane to establish a replicative niche known as the *Legionella*-containing vacuole (LCV) (*1*). *L. pneumophila* translocates more than 300 effector proteins via the type 4 secretion system (T4SS) into the host cell to manipulate cellular machinery, and a few *Legionella* effectors acting as phosphoinositide kinases or phosphatases have been identified (*2*–*5*). However, none of these *Legionella* phosphoinositide kinases and phosphatases are known to target the ER.

Here, we describe a two-component module—consisting of the *Legionella* effector MavQ and the *Legionella* phosphatidylinositol (PI) 3-phosphatase SidP—that dynamically self-organizes to remodel the host ER.

The *Legionella* effector MavQ is an atypical PI 3-kinase

We analyzed the *Legionella* T4SS effectors and identified MavQ (Lpg2975) as a distant homolog of the protein kinase superfamily (fig. S1A). To gain insight into the molecular function of MavQ, we solved the crystal structure of MavQ$^{1-580}$ (residues 1 to 580), which resolved an atypical kinase domain (KD) followed by a four-helical bundle (Fig. 1, A and B). The KD consists of a β strand–rich N-lobe and an α helix–rich C-lobe. The catalytic pocket of the KD adopts an active conformation with adenosine diphosphate (ADP) and Mg^{2+} bound in the active site (Fig. 1C).

Searches using the DALI server (*6*) identified several phosphoinositide kinases as the closest structural homologs of MavQ (fig. S1B). A systems biology screen also suggests that MavQ may be a phosphoinositide kinase (*7*). Indeed, MavQ$^{1-853}$, but not its predicted catalytically inactive mutant MavQ$^{1-853,\ \text{D160A}}$, phosphorylated PI and PI 5-phosphate (PI5P) (Fig. 1D and fig. S1C). Ala substitutions of active-site residues decreased MavQ's kinase activity (fig. S1, D and E). MavQ$^{1-853}$ produced mono- and diphosphorylated PI species when PI and PI5P were used as substrates, respectively, and treatment with the *Legionella* PI 3-phosphatase SidP (*3*) removed the phosphates introduced by MavQ (Fig. 1E and fig. S1F). Furthermore, the MavQ$^{1-853}$ reaction products with PI-containing liposomes as substrates were detected by the PI 3-phosphate (PI3P)–specific probe PI3P Grip (fig. S1G). Thus, MavQ is a PI 3-kinase that phosphorylates PI and PI5P in vitro.

MavQ generates PI3P on the ER and atypical vesicular-tubular structures

Fluorescence and immunogold electron microscopy revealed that MavQ fused to green fluorescent protein (GFP-MavQ) localized to the ER or atypical vesicular-tubular membrane structures in HeLa cells and RAW264.7 macrophages (Fig. 2, A and B, fig. S2, A to D, and movie S1). MavQ, but not MavQ$^{\text{D160A}}$, redistributed the PI3P probes 2xFYVE$_{\text{Hrs}}$ and 2xFYVE$_{\text{EEA1}}$ from early endosomes—the major PI3P-positive membrane compartment—to the ER or vesicular-tubular structures (Fig. 2C and fig. S2, E to G). Similar results were obtained using a doxycycline-inducible expression system (fig. S2H) and when recombinant GFP-MavQ$^{1-853}$ was microinjected into the cell (fig. S2I). Notably, MavQ-positive vesicular-tubular structures rarely colocalized with the early endosomal marker RAB5B (fig. S2G), indicating that these structures were unlikely to be of endosomal origin. Moreover, catalytically inactive MavQ$^{\text{D160A}}$ was mostly cytosolic (Fig. 2C and fig. S2E), which suggests that MavQ's membrane association depended on its kinase activity. The PI 3,5-bisphosphate [PI(3,5)P$_2$] probe mCherry-2xML1N$_{\text{TRPML1}}$ was not redistributed to MavQ-positive compartments (fig. S2J), likely because of the absence of PI5P on MavQ-targeted organelles or the accuracy of the probe (*8*).

MavQ exhibits spatiotemporal oscillatory behavior and remodels the host ER membrane

The ER constitutes the major intracellular PI pool (*9*) and targeting of the host PI 3-kinase VPS34 to the ER produces PI3P to drive autophagosome formation (*10*). By analogy, we wondered whether MavQ could target the ER and drive the formation of MavQ-positive vesicular-tubular structures. Sequence analysis predicted that MavQ contains noncanonical, ER-targeting FFAT motifs (fig. S3A), and mutations in three of them decreased MavQ's ER targeting (fig. S3B). We used live cell imaging to track the dynamics of MavQ and observed budding of MavQ-enriched vesicles and tubules from the ER (Fig. 3A, fig. S3C, and movie S2). These MavQ-enriched vesicles and tubules were capable of fusing with the ER (Fig. 3B and movie S3). Remarkably, MavQ was able to oscillate back and forth between the ER and vesicles/tubules at intervals on the order of minutes, and this oscillation propagated in a wave-like manner on the order of micrometers (Fig. 3C, fig. S3D, and movies S4 and S5). The spatial switching of MavQ and the propagation of switching events as waves suggested a feedback mechanism and possibly cooperativity underlying its ER association.

---

[1] Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2] Department of Biochemistry and Microbiology, Institute of Biology, Warsaw University of Life Sciences, Warsaw 02-776, Poland. [3] Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [4] Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [5] Department of Physiology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [6] Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA [7] Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

***Corresponding author. Email: vincent.tagliabracci@utsouthwestern.edu**

†These authors contributed equally to this work.

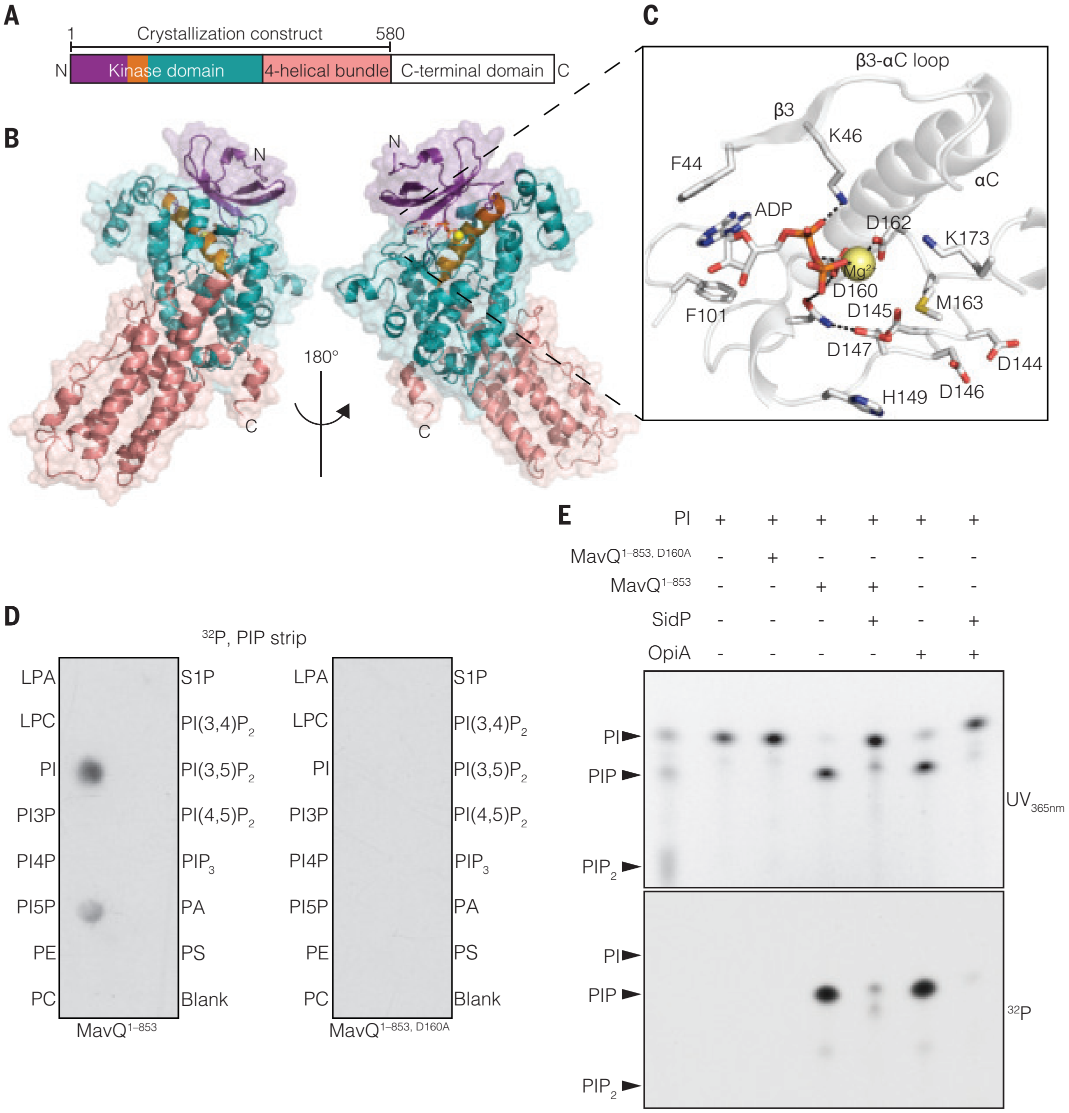


Fig. 1. The *Legionella* effector MavQ is an atypical PI 3-kinase. (**A** and **B**) Domain architecture of MavQ (A) and overall structure of MavQ$^{1-580}$ (B) depicting the N lobe (violet), the αC helix (orange), the C lobe (teal), the four-helical bundle (salmon), and the C-terminal domain (white). (**C**) Magnified view of MavQ$^{1-580}$ kinase active site depicting the interactions (dashed lines) involved in nucleotide binding. ADP is shown as sticks; the Mg^{2+} ion is a yellow sphere. (**D**) Autoradiograph depicting the transfer of γ-^{32}P from [γ-^{32}P]ATP onto PI and PI5P on a PIP strip by MavQ$^{1-853}$ but not by the predicted catalytically inactive mutant MavQ$^{1-853,\ D160A}$. (**E**) Chromatogram depicting the incorporation of γ-^{32}P from [γ-^{32}P]ATP by MavQ$^{1-853}$ or the inactive mutant, using BODIPY-labeled PI as a substrate. Reactions were subsequently treated with or without the PI 3-phosphatase SidP. OpiA, a *Francisella tularensis* PI 3-kinase (*5*), was used as a positive control. Reaction products were separated by thin-layer chromatography and visualized by UV (365 nm) fluorescence and autoradiography. Amino acid abbreviations: A, Ala; D, Asp; F, Phe; H, His; K, Lys; M, Met.

The C-terminal domain of MavQ recognizes MavQ-positive compartments, which leads to self-enhancement of MavQ's membrane association

We hypothesized that MavQ may promote its own membrane association because positive feedback is generally involved in biological oscillators (*11*). Whereas MavQ^{D160A} was mostly cytosolic (Fig. 2C and figs. S2E and S4A), expression of wild-type MavQ redistributed MavQ^{D160A} to the ER or vesicular-tubular structures (fig. S4A). Furthermore, the C-terminal domain of MavQ (MavQ^{CTD}, residues 581 to 871), but not the KD (MavQ^{KD}, residues 1 to 345), was also redistributed to the ER or vesicular-tubular structures by wild-type MavQ (fig. S4B). Thus, the MavQ^{CTD} recognizes MavQ-positive compartments, likely by interacting with membrane-bound MavQ or with its reaction product PI3P. Indeed, MavQ$^{1-853}$ predominantly co-sedimented with liposomes containing PI3P or PI(3,5)P_2 (Fig. 4A and fig. S4C), and the MavQ^{CTD} was necessary and sufficient for PI3P binding (fig. S4D). Thus, MavQ promotes its own membrane association in part by interaction with its reaction product PI3P.

The *Legionella* PI 3-phosphatase SidP facilitates spatiotemporal oscillatory behavior of MavQ

For a biological system to oscillate, negative feedback is required (*11*). We reasoned that a counteracting PI 3-phosphatase could be the negative regulator underlying the spatiotemporal oscillation of MavQ. Host PI 3-phosphatases may fulfill the role to some extent, as some type of oscillatory behavior occurred when only MavQ was expressed in the cell (Fig. 3C, fig. S3D, and movies S4 and S5). However, the *Legionella* PI 3-phosphatase SidP is functionally linked to MavQ (*7*). Cytosolic mCherry-SidP was redistributed to MavQ-positive compartments when coexpressed with GFP-MavQ (fig. S5A) and caused the sequestration of GFP-MavQ from the main ER network into clumps that retained ER identity (Fig. 4, B and C, and fig. S5B). Live cell imaging revealed that SidP facilitated spatiotemporal oscillation of MavQ (Fig. 4B and fig. S5C). MavQ- and SidP-enriched ER clumps or subdomains were able to generate MavQ and SidP protein waves, which propagated on the ER membrane and drove membrane remodeling during propagation (Fig. 4D, fig. S5D, and movies S6 and S7). Such wave propagation correlated with PI3P turnover on the ER (fig. S5E and movie S8), which suggests that the concerted PI 3-kinase and phosphatase activities of MavQ and SidP respectively constitute positive and negative feedback loops to drive MavQ protein dynamics and MavQ-induced membrane dynamics.

To further test the role of PI3P turnover in MavQ dynamics, we recruited the human PI 3-phosphatase MTM1 to the ER by means of the inducible FKBP-FRB system (*12*). GFP-MavQ was rapidly released from the ER, and

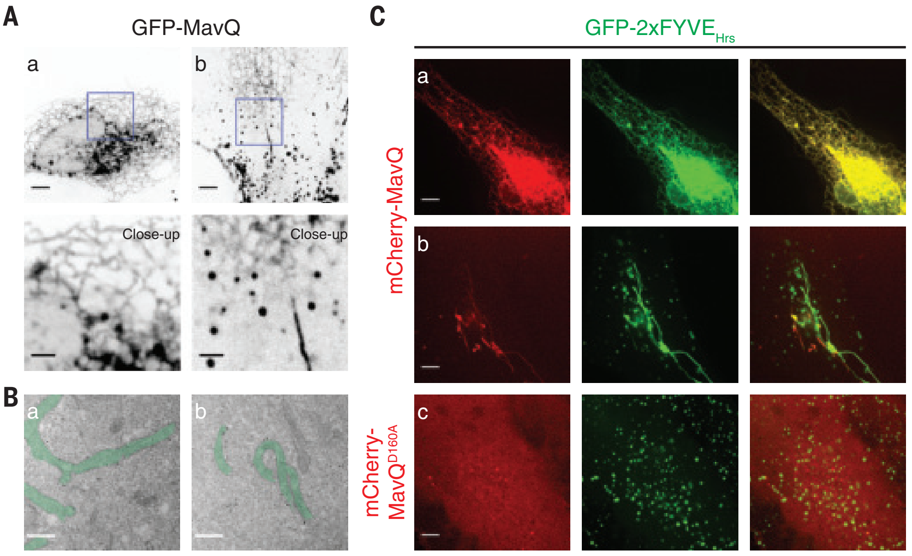


Fig. 2. MavQ generates PI3P on the ER and atypical vesicular-tubular structures. (**A**) Confocal images of HeLa cells expressing GFP-MavQ. In (a), GFP-MavQ localizes to an ER-like compartment. In (b), GFP-MavQ concentrates on vesicular-tubular structures, in addition to the weak ER-like localization. Scale bars, 5 μm; close-ups, 2 μm. (**B**) Transmission electron micrographs depicting the ultrastructural localization of GFP-MavQ in HeLa cells. GFP-MavQ was visualized by immunogold labeling with antibody to GFP. Membrane structures positive for GFP-MavQ are highlighted in green. The micrographs were taken with ultrathin sections from a cell where GFP-MavQ localizes to an ER-like compartment (a) or from a cell where GFP-MavQ localizes to vesicular-tubular structures (b). See also fig. S2, A and B. Scale bars, 200 nm. (**C**) Confocal images of HeLa cells expressing the PI3P probe GFP-2xFYVE$_{\text{Hrs}}$, along with mCherry-MavQ (a and b) or mCherry-MavQ$^{\text{D160A}}$ (c). In addition to its normal endosomal localization still visible in (b), GFP-2xFYVE$_{\text{Hrs}}$ distributes to MavQ-positive ER (a) or vesicular-tubular structures (b). MavQ$^{\text{D160A}}$ is mostly cytosolic and does not affect normal endosomal localization of GFP-2xFYVE$_{\text{Hrs}}$ (c). Scale bars, 5 μm.

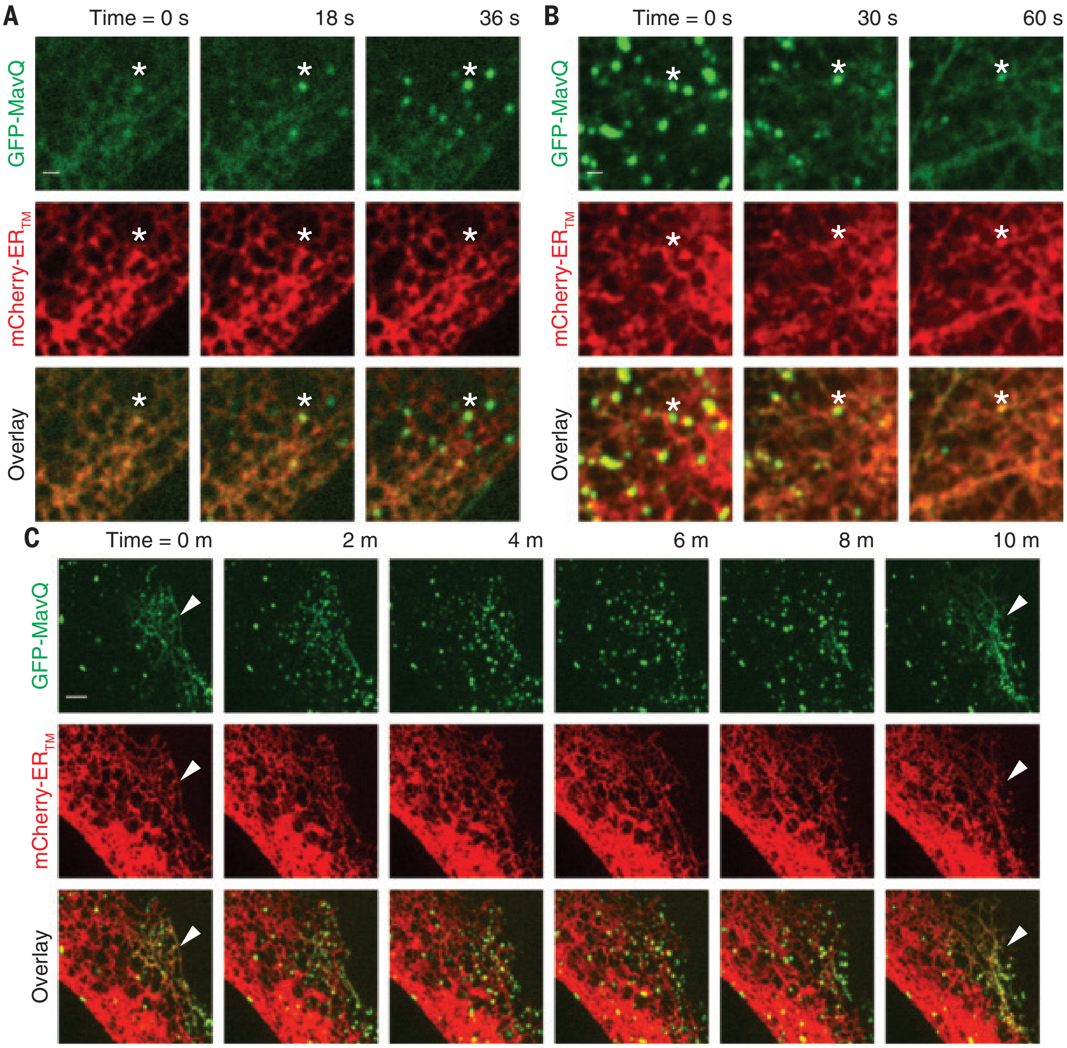


Fig. 3. MavQ exhibits spatio-temporal oscillatory behavior and remodels the host ER membrane. (**A** to **C**) Localization of GFP-MavQ relative to the ER membrane probe mCherry-ER$_{\text{TM}}$ in live HeLa cells by time-lapse confocal imaging. (A) GFP-MavQ initially distributes mostly on the ER, but several MavQ-enriched vesicles (an instance marked by asterisks) then emerge within 36 s; see also movie S2. Note the time stamps here are –54 s from the ones shown in movie S2. (B) GFP-MavQ initially distributes mostly on vesicles, but these MavQ-enriched vesicles (an instance marked by asterisks) then fuse with the ER within 60 s. Notice the increase in intensity of GFP-MavQ on the ER over time while MavQ-enriched vesicles vanish; see also movie S3. (C) At 0 min, MavQ localizes on the ER subdomain in the right corner (marked by arrowheads) and on vesicles in the rest of the cell. ER-localizing MavQ then dissipates while more MavQ-positive vesicles form in the region. At 10 min, MavQ redistributes to the ER subdomain in the right corner again (marked by arrowheads). See also movie S4. Scale bars, 1 μm [(A) and (B)], 3 μm (C).

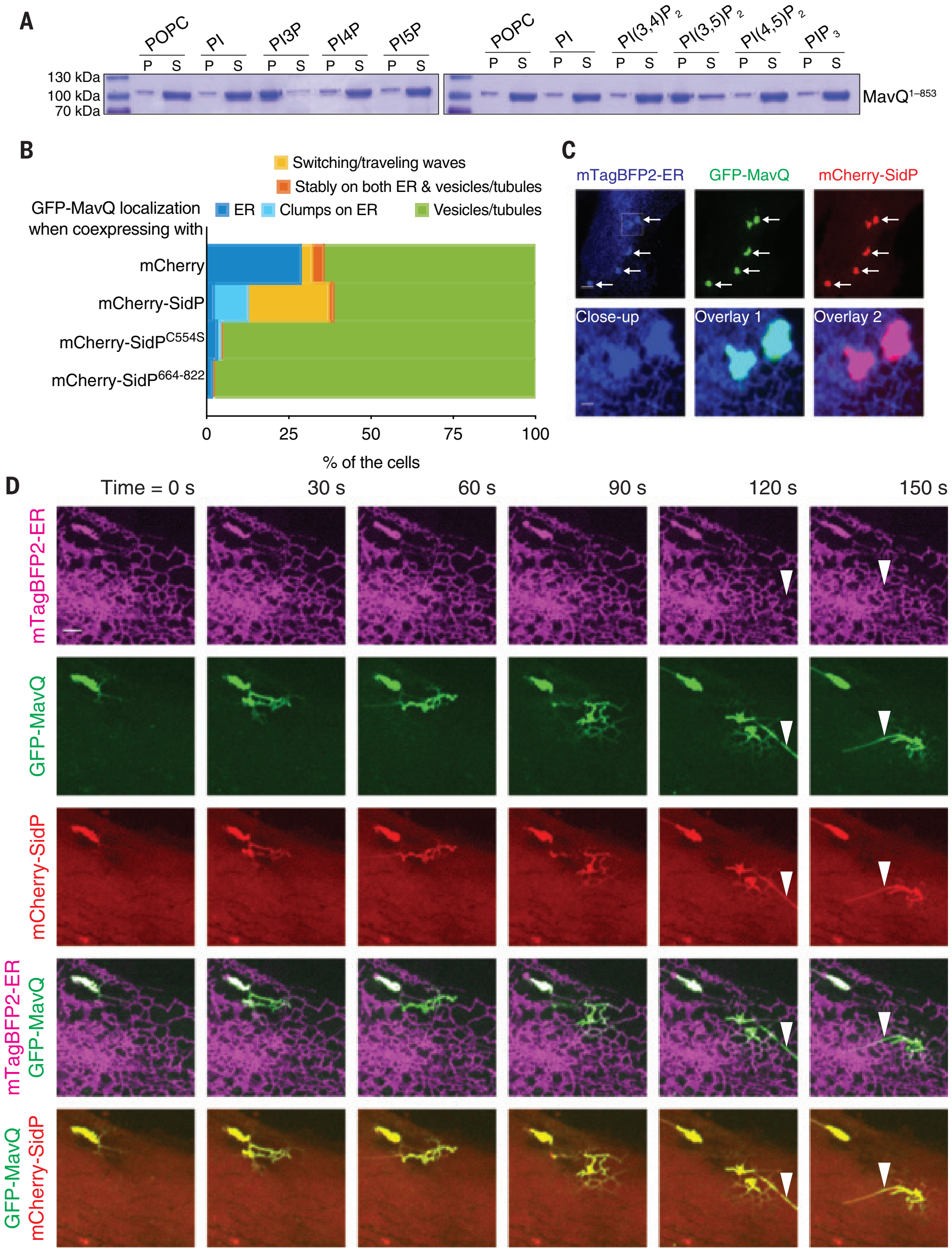


Fig. 4. Positive and negative feedback loops facilitate MavQ protein dynamics and MavQ-induced membrane dynamics. (**A**) Sedimentation of MavQ$^{1-853}$ with 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) or phosphoinositide-containing liposomes. Pellet (P) and supernatant (S) fractions were resolved by SDS–polyacrylamide gel electrophoresis (PAGE) and visualized by Coomassie staining. Liposome-bound proteins are present in the pellet. See fig. S4C for quantification. (**B**) Frequency bar plot showing the proportions of cells with different types of MavQ localization when coexpressed with mCherry, mCherry-SidP, mCherry-SidP^{C554S}, or mCherry-SidP$^{664-822}$ (200 cells per condition, pooled from four independent experiments). See also fig. S5, B and C. (**C** and **D**) Localization of GFP-MavQ and mCherry-SidP relative to the ER in live HeLa cells by confocal imaging. (C) mCherry-SidP sequesters GFP-MavQ from the main ER network into clumps that still retain ER identity, as indicated by arrows. The close-up shows two such clumps. Scale bars, 5 μm; close-up, 1 μm. (D) Initially, GFP-MavQ and mCherry-SidP concentrate on an ER clump. A wave of ER-bound GFP-MavQ and mCherry-SidP then emits and propagates on the ER network. MavQ- and SidP-positive tubules, which are negative for the ER luminal probe mTagBFP2-ER, form during the process (instances marked by arrowheads). See also movie S6. Scale bar, 3 μm.

MavQ-positive vesicular-tubular structures formed upon rapamycin-induced ER recruitment of wild-type MTM1 but not of the catalytically inactive MTM1^{C375S} mutant (fig. S6, A and B, and movies S9 and S10). Thus, PI3P is necessary to sustain the ER localization of MavQ, and PI3P depletion is sufficient to trigger the spatial switching of MavQ.

SidP contains an N-terminal phosphatase domain, an insertion domain, and an α-helical C-terminal domain (fig. S6C) (*3*). We reasoned that SidP may modulate the MavQ-ER interaction not only through its phosphatase activity but also in another manner. Indeed, catalytically inactive SidP^{C554S} as well as the SidP^{CTD} (residues 664 to 822) colocalized with MavQ on vesicular-tubular structures (fig. S6D), and both were sufficient to reduce the association of MavQ to the ER (Fig. 4B and fig. S5B). Thus, although the phosphatase activity of SidP was required for facilitating MavQ dynamics (Fig. 4B and fig. S5C), SidP also modulates MavQ's ER association via its CTD. SidP^{C554S} co-sedimented with MavQ$^{1-853}$ when using PI3P-containing liposomes (fig. S6, E and F). Furthermore, both MavQ^{CTD} and SidP^{CTD} were necessary for this interaction (fig. S6, E and F). Thus, SidP interacts with membrane-bound MavQ via its CTD. Collectively, PI3P-dependent MavQ-SidP interaction and SidP's PI 3-phosphatase activity constitute the negative feedback to facilitate the spatiotemporal oscillation of MavQ.

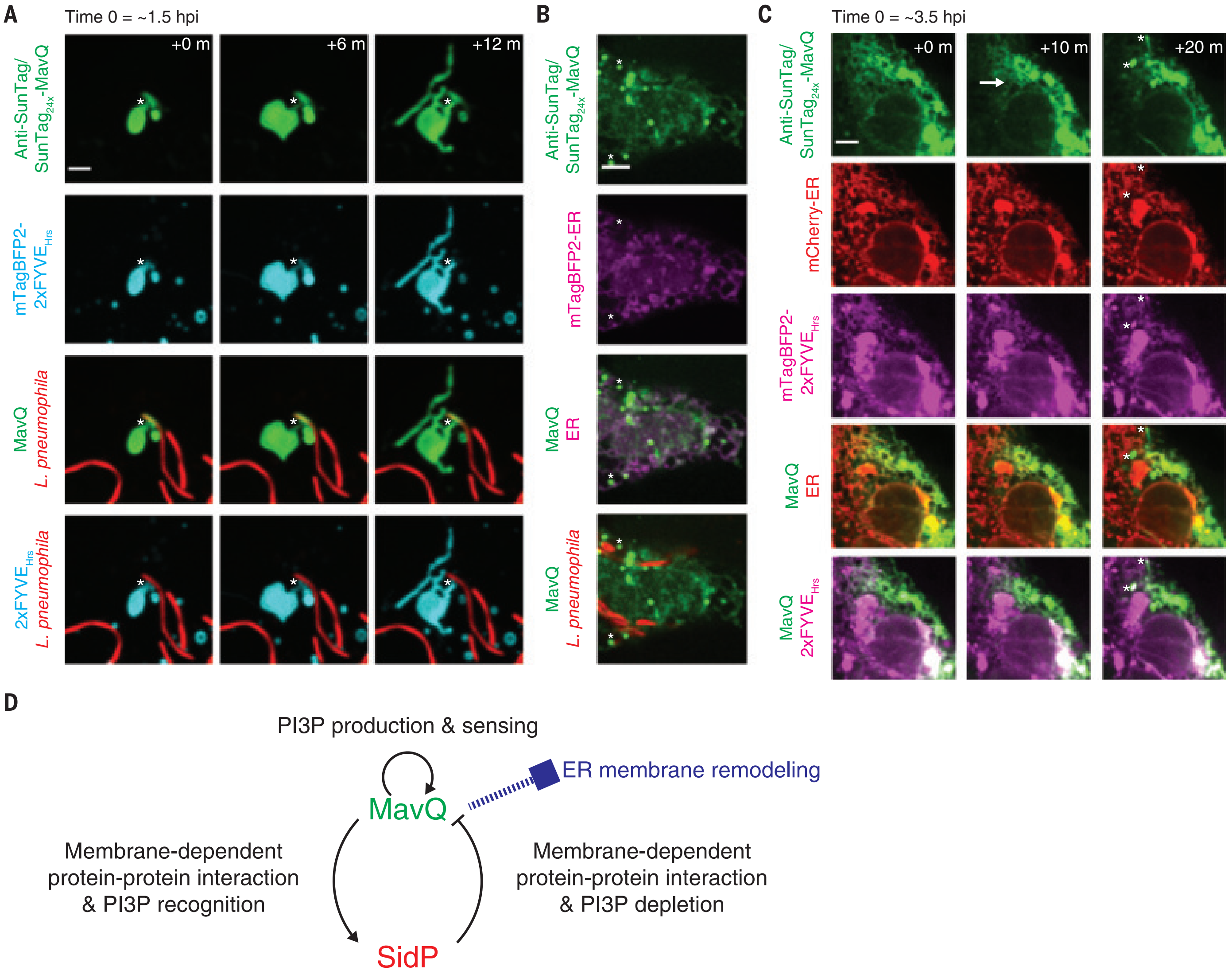


Fig. 5. MavQ exhibits spatiotemporal oscillatory behavior and remodels the host ER membrane during infection. (**A** and **B**) Confocal images of live COS-7 cells expressing anti-SunTag together with the PI3P probe mTagBFP2-2xFYVE$_{\text{Hrs}}$ (A) or the ER luminal probe mTagBFP2-ER (B) challenged with opsonized, mCherry-labeled *L. pneumophila* expressing SunTag$_{24x}$-MavQ. (A) At ~1.5 hours post-infection (hpi), MavQ-enriched, PI3P-positive vesicular-tubular structures bud from the LCV. Note that the LCV subdomain positive for MavQ is not PI3P-positive. The neck connecting the LCV and vesicular-tubular structures is marked by asterisks. See also movie S12. (B) At ~5 hpi, MavQ weakly decorates the ER in addition to vesicular-tubular structures (instances marked by asterisks). Scale bars, 3 μm (A), 5 μm (B). (**C**) Time-lapse confocal images of a live COS-7 cell expressing anti-SunTag, mCherry-ER, and mTagBFP2-2xFYVE$_{\text{Hrs}}$ at ~3.5 hpi with opsonized, unlabeled *L. pneumophila* expressing SunTag$_{24x}$-MavQ. A wave of MavQ dissociation from the ER occurs (direction of propagation indicated by the arrow), and MavQ-enriched, PI3P-positive vesicles emerge during the process (instances marked by asterisks). Scale bar, 5 μm. See also movies S15 and S16. (**D**) Model depicting interlinked positive and negative feedback loops between MavQ and SidP that drive spatiotemporal oscillation of MavQ to facilitate ER membrane remodeling.

MavQ exhibits spatiotemporal oscillatory behavior and remodels the host ER membrane during infection

We asked whether any membrane remodeling occurs in association with PI3P turnover during *Legionella* infection and whether MavQ is involved in this process. We observed dynamic PI3P-positive vesicular-tubular structures originating near the LCV (fig. S7A and movie S11) as well as sparks of the PI3P signal around the LCV (fig. S7B and movie S11) at ~1 hour post-infection (hpi). These PI3P-positive vesicular-tubular structures underwent fusion and fission and subsequently developed into vesicular-tubular networks (fig. S7A and movie S11). The T4SS was required for the formation of these PI3P-positive structures, and deletion of *mavQ* diminished their occurrence (fig. S7C).

We adopted the SunTag system (*13*) and developed a live cell imaging approach to track the dynamics of translocated *Legionella* effectors in the host cell. As a proof of principle, we tracked the *Legionella* effector SidC, which binds to the LCV during infection (*1*). In uninfected cells, constitutively expressed anti-SunTag (single-chain variable fragment antibody to GCN4, fused to GFP, the solubility tag GB1, and a nuclear localization sequence) was sequestered in the nucleus (fig. S7D). When the cell was challenged with a *Legionella* strain expressing SunTag$_{24x}$-SidC (SidC tagged with 24 copies of the GCN4 peptide), anti-SunTag redistributed to the polar regions of the LCV within 0.5 hpi (fig. S7E), indicating the onset of SunTag24x-SidC translocation at the bacterial poles. Notably, this redistribution was T4SS-dependent (fig. S7, E and F). At ~0.5 to 1 hpi, the entire membrane of some LCVs was decorated with SidC (fig. S7G). SidC remained associated with the LCV and did not redistribute to PI3P-positive vesicular-tubular structures or networks at later time points (fig. S7H).

When the cell was challenged with a *Legionella* strain expressing SunTag$_{24x}$-MavQ, anti-SunTag decorated the polar regions of the LCV and sometimes the entire membrane starting at ~1 to 1.5 hpi (fig. S8A). At this time point, the LCV had been associated with the ER or acquired ER-derived membranes (fig. S8A), and the majority of LCVs had lost PI3P that had accumulated during the early phase of LCV formation (fig. S8B). Remarkably, we observed budding of MavQ-enriched, PI3P-positive vesicular-tubular structures from the ER-associated LCV (Fig. 5A, fig. S8C, and movies S12 and S13). These structures later developed into dynamic vesicular-tubular networks spreading throughout the cytoplasm (fig. S9 and movie S14). We observed fast, "catastrophic" releases of MavQ together with the PI3P probe mTagBFP2-2xFYVE$_{Hrs}$ from vesicular-tubular structures to the cytosol, as well as shrinkage of these structures (fig. S9 and movie S14). Starting at ~3.5 hpi, MavQ decorated the whole ER network or ER subdomains in some cells (Fig. 5, B and C). Furthermore, MavQ was able to form protein waves in concert with PI3P turnover on the ER and redistribute from the ER to vesicular-tubular structures (Fig. 5C, fig. S10A, and movies S15 to S17). We also performed co-infection using a *Legionella* strain expressing SunTag$_{24x}$-SidP and a *Legionella* strain expressing untagged MavQ. SidP localized on transient PI3P-positive vesicular-tubular structures in the cell co-infected by both strains (fig. S10B and movie S18); this finding supports the role of SidP in facilitating the PI3P turnover and PI3P-positive membrane dynamics. Thus, MavQ translocated from *L. pneumophila* exhibits spatiotemporal oscillatory behavior and remodels the host ER membrane in the context of infection.

To assess the functional significance of MavQ and SidP, we monitored replication of *Legionella* strains bearing a catalytically inactive mutation of *mavQ* (Lp02, $mavQ^{D160A}$) or the *sidP* deletion (Lp02, Δ*sidP*) in *Acanthamoeba castellanii*. The mutant strains displayed an intracellular growth defect that was complemented by expression of wild-type MavQ or wild-type SidP (fig. S11). Thus, both MavQ and SidP are required for full *Legionella* virulence in a eukaryotic host.

Discussion

ER manipulation is a hallmark of host cell subversion by *L. pneumophila* (*1*). Our results reveal that the atypical *Legionella* PI 3-kinase MavQ alters the phosphoinositide identity of the ER and is capable of ER membrane remodeling.

During *Legionella* infection, the nascent LCV is first enriched with PI3P and then matures into a PI 4-phosphate (PI4P)–enriched vacuole (*1*). *Legionella* effectors are not required for the initial PI3P acquisition of the LCV (*14*). Sequential action of the *Legionella* PI 4-kinase LepB and the PI 3-phosphatase SidF may then mediate the PI3P-to-PI4P conversion in the LCV membrane (*2*, *4*). A recent report suggests that MavQ acts in tandem with LepB and SidF to supply PI4P in the LCV membrane (*15*). Although it is possible that MavQ indirectly contributes to PI4P generation in the LCV membrane and thus to intracellular *Legionella* growth, the following observations cause us to favor the idea that the membrane-remodeling capabilities of MavQ play more important roles: (i) coevolution of MavQ and SidP during *Legionella* speciation (*7*), and (ii) the requirement of biochemical synergy between MavQ and SidP for effective, dynamic membrane remodeling.

The positive and negative feedback loops between MavQ and SidP constitute a simple molecular circuit that displays oscillatory behavior and enables MavQ and SidP to form traveling waves on the membrane (Fig. 5D). We propose the following model based on Turing's discovery of reaction-diffusion pattern-forming systems (*16*): Initially, MavQ and SidP are freely diffusing in the cytosol, which would lead to a spatially homogeneous state of the MavQ-SidP system if there were no adenosine triphosphate (ATP) for MavQ to consume. However, with energy from cytosolic ATP, any MavQ molecule transiently associated with the ER membrane has a chance to convert PI into PI3P, which promotes the recruitment of other MavQ molecules through MavQ's PI3P-binding CTD. Local accumulation of membrane-bound MavQ then triggers the recruitment of SidP. Intuitively, membrane-bound SidP would cause MavQ and then SidP itself to dissociate from the membrane as a result of SidP-catalyzed PI3P hydrolysis. Nonetheless, if SidP diffuses more quickly on the membrane than MavQ does, it is possible that MavQ-SidP co-clusters will form on the membrane according to Turing's model (*16*, *17*). Such MavQ-SidP co-clusters may correspond to the MavQ- and SidP-enriched ER clumps or subdomains we observed. Once local concentrations of membrane-bound MavQ and SidP reach certain thresholds, membrane remodeling can occur as a consequence of protein crowding, possible membrane insertion of the MavQ^{CTD}, or even PI3P accumulation (*18*), ultimately giving rise to MavQ- and SidP-positive vesicles and tubules. These nascent ER-derived vesicles and tubules can fuse back with the ER or dissipate, thereby releasing MavQ and SidP back to the ER or the cytosol. This completes one cycle of spatiotemporal oscillation of MavQ and SidP. In addition, traveling waves can also occur in reaction-diffusion systems according to Turing's theory (*16*, *17*). Indeed, waves of MavQ and/or SidP form and propagate on the ER membrane.

The MavQ-SidP system resembles the exemplary minimal biochemical oscillator MinD-MinE, which forms traveling waves (*19*, *20*) that can be explained by a Turing-like reaction-diffusion mechanism (*20*). This makes the MavQ-SidP system a strong candidate for a simple two-component Turing-like reaction-diffusion system. Striking spatiotemporal patterns of phosphoinositides have been recently reconstituted in vitro using a kinase-phosphatase competitive reaction (*21*), and a synthetic biology investigation also demonstrated that artificial molecular circuits containing competing kinases and phosphatases are capable of exhibiting a particular self-organizing behavior (i.e., cell polarization) (*22*). Traveling waves of phosphoinositides occur at the plasma membrane of *Dictyostelium discoideum* and mast cells (*23*, *24*) and can be induced at the trans-Golgi network upon inhibition of the PI 4-kinase PI4KIIIβ (*25*). Our results exemplify the importance of self-organizing behaviors that result from chemically interacting kinases and phosphatases in complex cellular behaviors, and reveal a mechanism that intracellular bacterial pathogens use to remodel host cellular membranes for survival.

REFERENCES AND NOTES

1. B. Steiner, S. Weber, H. Hilbi, *Int. J. Med. Microbiol.* **308**, 49–57 (2018).
2. F. Hsu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13567–13572 (2012).
3. L. Toulabi, X. Wu, Y. Cheng, Y. Mao, *J. Biol. Chem.* **288**, 24518–24527 (2013).
4. N. Dong *et al.*, *Nat. Microbiol.* **2**, 16236 (2016).
5. H. E. Ledvina *et al.*, *Cell Host Microbe* **24**, 285–295.e8 (2018).
6. L. Holm, L. M. Laakso, *Nucleic Acids Res.* **44**, W351–W355 (2016).
7. M. L. Urbanus *et al.*, *Mol. Syst. Biol.* **12**, 893 (2016).
8. G. R. Hammond, S. Takasuga, T. Sasaki, T. Balla, *PLOS ONE* **10**, e0139957 (2015).
9. Y. J. Kim, M. L. Guzman-Hernandez, T. Balla, *Dev. Cell* **21**, 813–824 (2011).
10. E. L. Axe *et al.*, *J. Cell Biol.* **182**, 685–701 (2008).
11. T. Y. Tsai *et al.*, *Science* **321**, 126–129 (2008).
12. V. M. Rivera *et al.*, *Nat. Med.* **2**, 1028–1032 (1996).
13. M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, *Cell* **159**, 635–646 (2014).
14. S. Weber, M. Wagner, H. Hilbi, *mBio* **5**, e00839-13 (2014).
15. G. Li, H. Liu, Z. Q. Luo, J. Qiu, *EMBO Rep.* **22**, e51163 (2021).
16. A. M. Turing, *Philos. Trans. R. Soc. London Ser. B* **237**, 37–72 (1952).
17. J. Halatek, F. Brauns, E. Frey, *Philos. Trans. R. Soc. London Ser. B* **373**, 20170107 (2018).
18. H. F. Renard, L. Johannes, P. Morsomme, *Trends Cell Biol.* **28**, 274–286 (2018).
19. D. M. Raskin, P. A. de Boer, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4971–4976 (1999).
20. M. Loose, E. Fischer-Friedrich, J. Ries, K. Kruse, P. Schwille, *Science* **320**, 789–792 (2008).
21. S. D. Hansen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15013–15022 (2019).
22. A. H. Chau, J. M. Walter, J. Gerardin, C. Tang, W. A. Lim, *Cell* **151**, 320–332 (2012).

23. Y. Arai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12399–12404 (2010).
24. D. Xiong *et al.*, *Nat. Chem. Biol.* **12**, 159–166 (2016).
25. B. Mesmin *et al.*, *EMBO J.* **36**, 3156–3174 (2017).

ACKNOWLEDGMENTS

We thank members of the Tagliabracci and Liou laboratories for discussions, A. Darehshouri for help with TEM, and M. Gradowski for help with bioinformatics. Results shown in this report are derived from work performed at the Structural Biology Center, Advanced Photon Source, Argonne National Laboratory. **Funding:** This work was funded by NIH grants DP2GM137419 (V.S.T.), R01GM113079 (J.L.), T32GM008203-29 (V.A.L.), and F30HL143859-01 (M.H.B.); Welch Foundation grants I-1911 (V.S.T.) and I-1789 (J.L.); CPRIT grant RP170674 (V.S.T.); and Polish National Agency for Scientific Exchange scholarship PPN/BEK/2018/1/00431 (K.P.). Research reported in this publication was supported by the Office of the Director, NIH under awards S10OD021685 and S10OD025018. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. J.L. is a Sowell Family Scholar in Medical Research. V.S.T. is a Michael L. Rosenberg Scholar in Medical Research, a CPRIT Scholar (RR150033), and a Searle Scholar. **Author contributions:** T.-S.H., V.A.L., M.H.B., J.L., and V.S.T. designed the experiments; K.P. performed the bioinformatics; T.-S.H., V.A.L., M.H.B., and V.S.T. performed molecular cloning, strain construction, and protein purification; V.A.L. and D.R.T. performed crystallization and structure determination; A.O. performed biophysical characterization; V.A.L. and M.H.B. performed biochemical experiments; T.-S.H. performed imaging experiments and cellular assays, and developed the live cell imaging method for tracking *Legionella* effectors during infection; and T.-S.H., V.A.L., M.H.B., K.P., and V.S.T. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All materials developed in this study will be made available upon request. The atomic coordinates have been deposited in the Protein Data Bank with accession code 7M7A.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/935/suppl/DC1
Materials and Methods
Figs. S1 to S11
Table S1
Movies S1 to S18
References (*26–56*)
MDAR Reproducibility Checklist

6 September 2019; resubmitted 17 February 2021
Accepted 14 April 2021
Published online 29 April 2021
10.1126/science.aay8118

CORONAVIRUS

Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9

Chunlei Jiao[1], Sahil Sharma[2]†, Gaurav Dugar[2]†‡, Natalia L. Peeck[1], Thorsten Bischler[3], Franziska Wimmer[1], Yanying Yu[1], Lars Barquist[1,4], Christoph Schoen[5], Oliver Kurzai[5,6], Cynthia M. Sharma[2]*, Chase L. Beisel[1,4]*

CRISPR-Cas systems recognize foreign genetic material using CRISPR RNAs (crRNAs). In type II systems, a trans-activating crRNA (tracrRNA) hybridizes to crRNAs to drive their processing and utilization by Cas9. While analyzing Cas9-RNA complexes from *Campylobacter jejuni*, we discovered tracrRNA hybridizing to cellular RNAs, leading to formation of "noncanonical" crRNAs capable of guiding DNA targeting by Cas9. Our discovery inspired the engineering of reprogrammed tracrRNAs that link the presence of any RNA of interest to DNA targeting with different Cas9 orthologs. This capability became the basis for a multiplexable diagnostic platform termed LEOPARD (leveraging engineered tracrRNAs and on-target DNAs for parallel RNA detection). LEOPARD allowed simultaneous detection of RNAs from different viruses in one test and distinguished severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and its D614G (Asp^{614}→Gly) variant with single-base resolution in patient samples.

CRISPR-Cas immune systems degrade foreign genetic material through the guidance of CRISPR RNAs (crRNAs) (*1*, *2*). crRNAs are encoded as spacer-repeat subunits within a system's CRISPR array (*3*). Each crRNA typically undergoes processing from a precursor transcribed from the array and then partners with the system's Cas effector nuclease to direct cleavage of target nucleic acids. Within type II systems, the source of Cas9 nucleases and many CRISPR technologies (*4*, *5*), crRNA processing and subsequent DNA targeting by Cas9 requires a trans-activating crRNA (tracrRNA) (*6*–*8*). The tracrRNA hybridizes to the "repeat" portion of each crRNA within the transcribed array. Host-derived ribonuclease (RNase) III then cleaves the formed RNA stem to generate a processed crRNA:tracrRNA duplex utilized by Cas9 (*6*). What remains unclear is whether crRNAs are confined to CRISPR-Cas loci or can be derived from elsewhere in the genome. Here, we show that crRNAs can be derived from host RNAs outside the CRISPR-Cas locus, inspiring a Cas9-based diagnostic platform that allows scalable detection of multiple biomarkers in a single test.

Cellular RNAs bound to Cas9 from *C. jejuni* resemble crRNAs

Our prior work interrogating RNAs bound to Cas9 from *Campylobacter jejuni* NCTC11168 (CjeCas9) revealed crRNA-guided RNA targeting by CjeCas9 (*9*). To further explore RNA binding partners of CjeCas9, we repeated the immunoprecipitation and RNA sequencing (RIP-seq) approach of epitope-tagged Cas9 using *C. jejuni* strain CG8421 harboring only two spacers in its endogenous type II-C CRISPR-Cas system (Fig. 1, A and B, and fig. S1, A and B). RIP-seq identified the CRISPR-tracrRNA locus as well as 205 RNA fragments derived from cellular RNAs enriched with Cas9-3xFLAG (Fig. 1C, fig. S1C, and table S1). Analyses of the enriched fragments using MEME (*10*) revealed two significant sequence motifs across all three replicates (Fig. 1D). Motif #1 was complementary to 13 nucleotides (nts) within the guide portion of crRNA2, in line with RNA targeting by crRNAs in NCTC11168 (*9*). Motif #2 was complementary to 21 nts within the tracrRNA anti-repeat domain. As this domain normally hybridizes to the crRNA repeat as part of crRNA biogenesis (fig. S2, A and B), motif #2 raised the possibility that these cellular RNAs were hybridizing with the tracrRNA, potentially becoming RNAs that function like crRNAs.

We explored this possibility through two routes. First, for the enriched RNA fragments with motif #2, we aligned the motif within each RNA fragment with a spacer-repeat pair, and we measured the length of each fragment corresponding to the spacer or repeat (Fig. 1E). Most frequently, the spacer part was 15 nt longer than a canonical crRNA spacer whereas the repeat part was the same size as a canonical crRNA repeat, similar to slightly extended versions of crRNAs. Second, we predicted how each RNA fragment base pairs with the tracrRNA anti-repeat (*11*). Predicted binding affinities were significantly higher for RNA fragments with motif #2 than for fragments without the motif ($p = 3 \times 10^{-7}$) (Fig. 1F). However, multiple RNA fragments were predicted to strongly pair with the tracrRNA anti-repeat despite lacking motif #2 (Fig. 1F), likely due to bulges in the RNA duplex creating discontinuities in the motif. For these RNA fragments and those containing motif #2, the predicted interactions between each RNA and the tracrRNA anti-repeat consistently contained imperfect RNA duplexes, with the most extensive pairing near the 3′ end of the anti-repeat (Fig. 1G and fig. S2). We made similar observations for motif #1 (fig. S1D).

---

[1] Helmholtz Institute for RNA-based Infection Research (HIRI)/Helmholtz-Centre for Infection Research (HZI), 97080 Würzburg, Germany. [2] Molecular Infection Biology II, Institute of Molecular Infection Biology, University of Würzburg. 97080 Würzburg, Germany. [3] Core Unit Systems Medicine, University of Würzburg, 97080 Würzburg, Germany. [4] Medical Faculty, University of Würzburg, 97080 Würzburg, Germany. [5] Institute for Hygiene and Microbiology, University of Würzburg, 97080 Würzburg, Germany. [6] Leibniz Institute for Natural Product Research and Infection Biology, Hans-Knoell-Institute, Jena, 07745 Germany.
***Corresponding author. Email: cynthia.sharma@uni-wuerzburg.de (C.M.S.); chase.beisel@helmholtz-hiri.de (C.L.B.)**
†These authors contributed equally to this work.
‡Present address: Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, Netherlands.

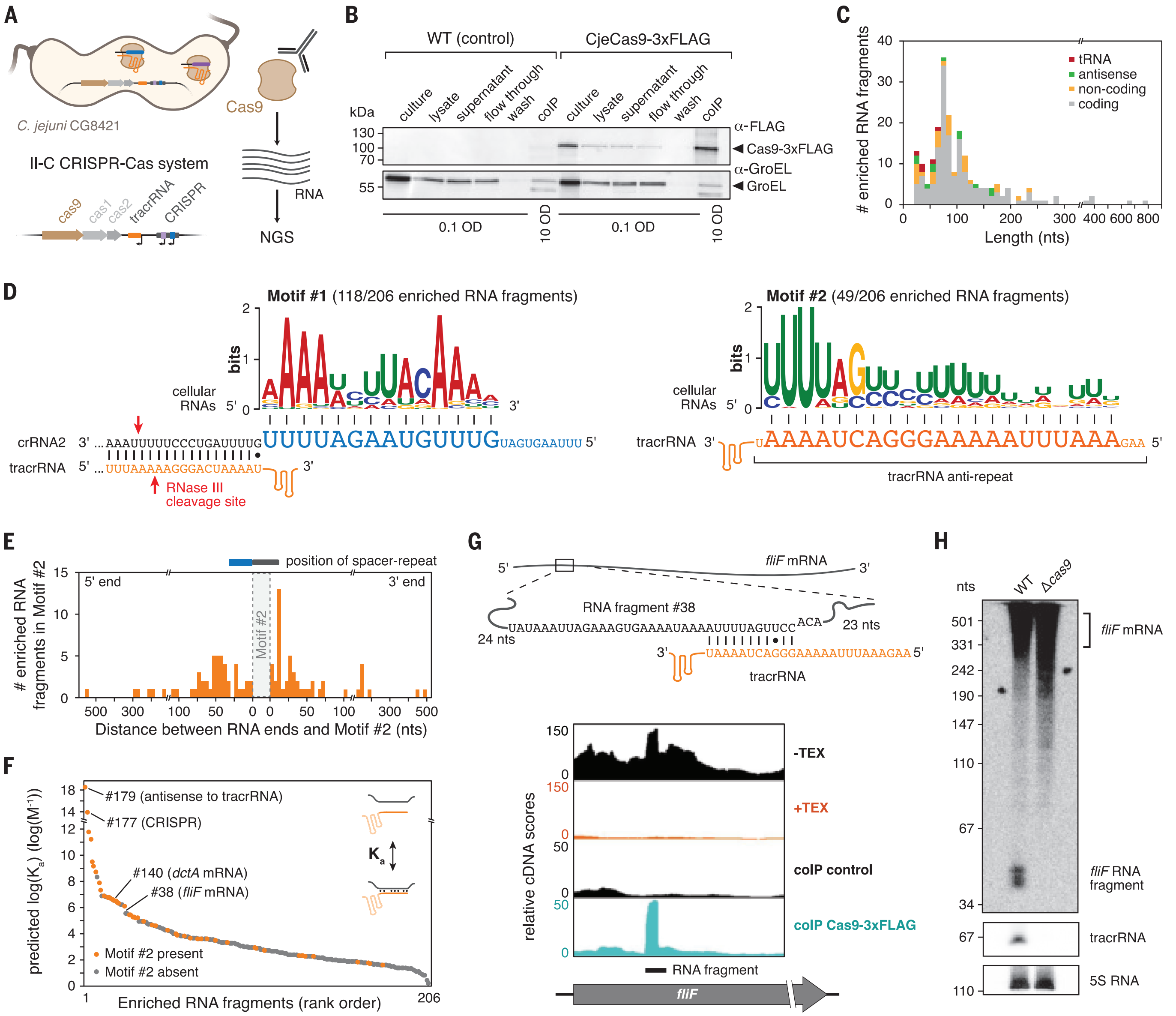


Fig. 1. Fragments of cellular RNAs bound by Cas9 in *Campylobacter jejuni* resemble crRNAs. (**A**) Coimmunoprecipitation and sequencing RNAs bound to Cas9-3xFLAG from *C. jejuni* CG8421 using RIP-seq. (**B**) Western blot analysis of samples from *C. jejuni* strains with Cas9-3xFLAG or untagged WT control before and after immunoprecipitation. (**C**) Size range of the cellular RNA fragments identified through RIP-seq. Colors indicate the class of RNA. (**D**) Two motifs extracted by MEME from the enriched RNA fragments and their predicted interaction with crRNA2 or the tracrRNA. *E*-value = 8.8×10^{-22} (motif 1), 8.9×10^{-14} (motif 2). (**E**) Distribution of RNA lengths centered around motif #2. (**F**) Predicted binding affinity (K_a) between the tracrRNA anti-repeat and each enriched RNA fragment. Orange indicates the presence of motif #2. (**G**) Predicted *fliF* mRNA:tracrRNA duplex and mapped reads from differential RNA-seq (top) or RIP-seq (bottom) performed in CG8421. (**H**) Northern blot analysis for *fliF* RNAs from CG8421 WT and *cas9*-deletion strains.

These crRNA-like RNAs raised the question of whether these same RNAs were present in our prior RIP-seq analysis with strain NCTC11168 (*9*). We found that 7 of the 96 enriched fragments were predicted to bind the tracrRNA anti-repeat more tightly than at least one crRNA (fig. S3, A and B). Two of these RNAs (derived from *fliF* and *dctA* mRNAs) matched those found in CG8421 (Fig. 1G and figs. S2D and S3C). The RNA fragment derived from the *fliF* mRNA could be detected by Northern blot in both total RNA and RIP-seq samples yet disappeared following deletion of *cas9* (Fig. 1H and fig. S4, A and B). The *dctA* RNA fragment was only weakly detected in one strain (fig. S4). Although *cas9* deletion did not significantly perturb FliF protein concentrations in vivo under standard growth conditions (fig. S5), deleting the CRISPR array in NCTC11168 increased levels of the *fliF* RNA fragment (fig. S4C). Finally, the *fliF* RNA fragment was a processing product, as confirmed with Terminator exonuclease treatment (Fig. 1G). These crRNA-like RNAs thus are also present in *C. jejuni* NCTC11168 and likely exist in other *C. jejuni* strains on account of the shared tracrRNA binding site in *fliF* (fig. S6).

Coimmunoprecipitated RNAs can function as noncanonical crRNAs that direct DNA targeting by Cas9

Cas9 binding, predicted tracrRNA pairing, and the length distribution of many of these enriched RNA fragments suggested that the tracrRNA pairs with endogenous RNAs, resulting in “noncanonical” crRNAs (ncrRNAs) (Fig. 2A and fig. S2A). The ncrRNAs therefore

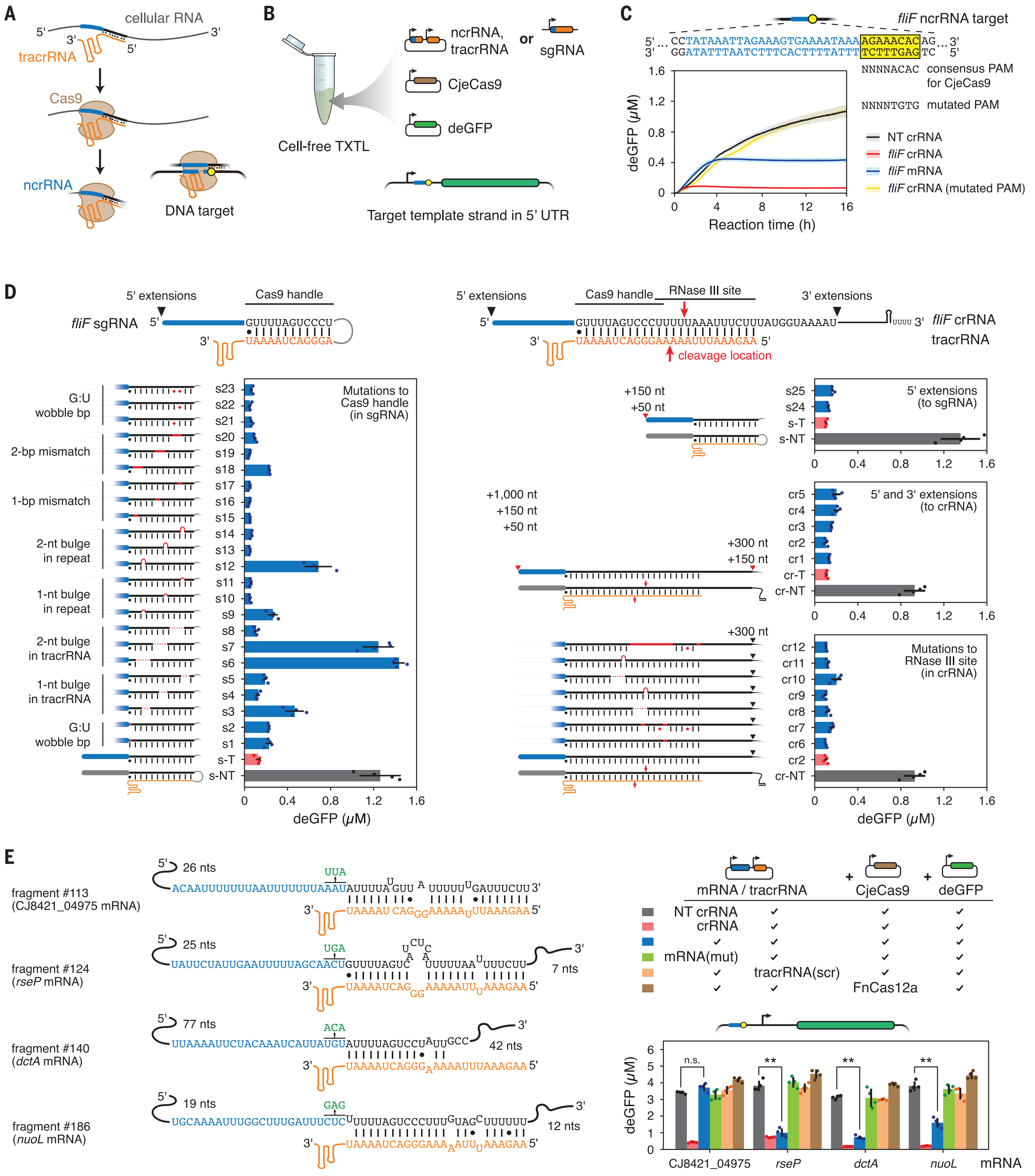


Fig. 2. Noncanonical crRNAs can direct DNA cleavage by CjeCas9. (**A**) General process for ncrRNA generation. (**B**) Applying the TXTL assay to characterize putative ncrRNAs. (**C**) DNA targeting through the *fliF* ncrRNA in TXTL. Lines and shaded regions indicate the mean and standard deviation from four separately mixed replicates. NT, nontargeting. (**D**) Systematic evaluation of mutating the repeat:anti-repeat duplex for CjeCas9 with TXTL. Endpoint GFP levels are shown. Mutations and extensions to the *fliF* sgRNA-crRNA repeat are indicated in red. See table S1 for sequences. (**E**) DNA targeting by selected ncrRNAs predicted in TXTL. Check marks indicate use of the construct above the line. mRNA(mut): mRNA encoding the ncrRNA with point mutations in the predicted "seed" region of the guide. tracrRNA(scr): tracrRNA with the anti-repeat sequence scrambled. Values in (D) and (E) represent the mean and standard deviation from four separately mixed replicates. **$p < 0.001$. n.s., not significant.

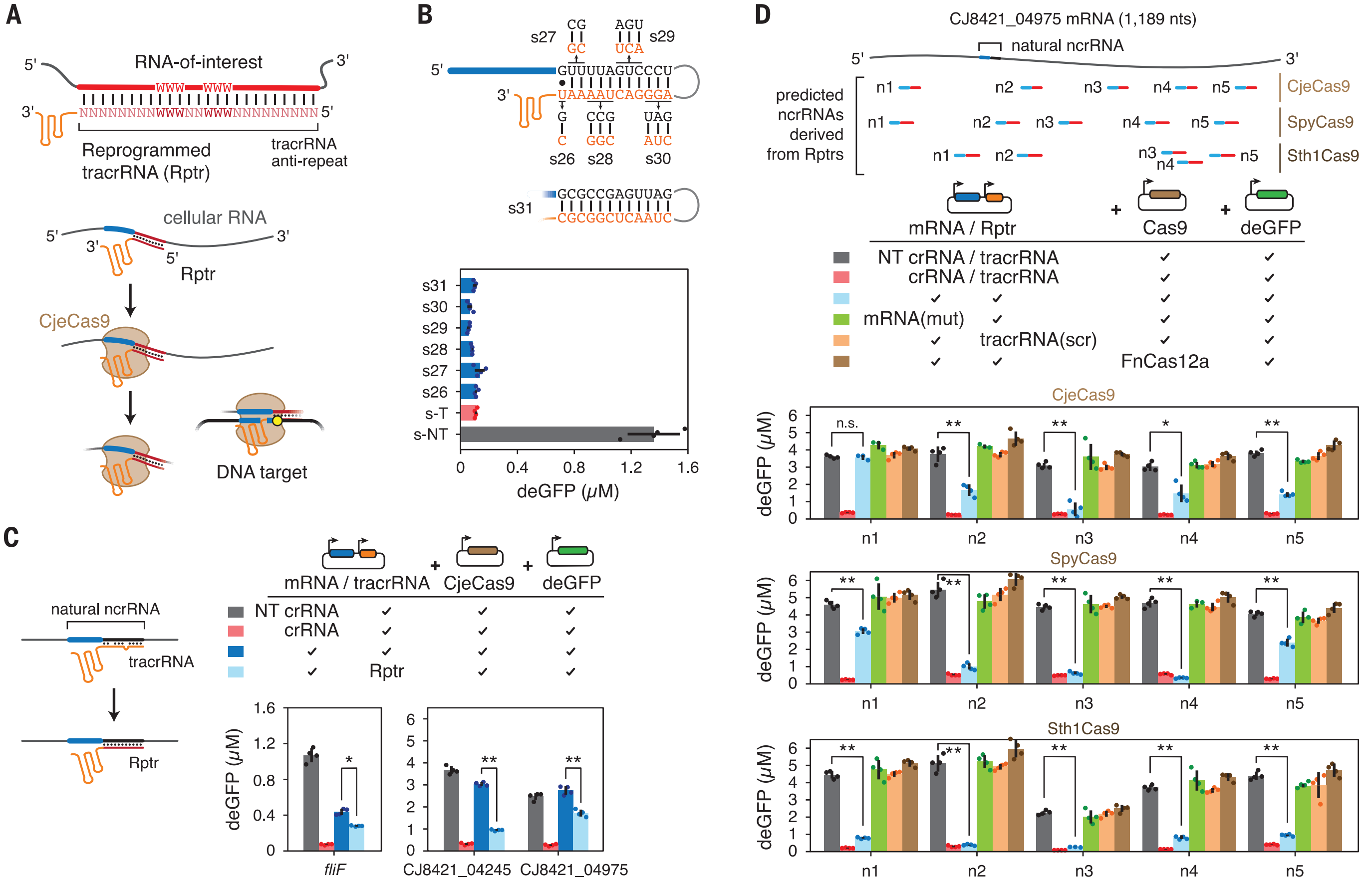


Fig. 3. Reprogrammed tracrRNAs co-opt RNA transcripts to guide different Cas9 orthologs. (**A**) Design of reprogrammed tracrRNAs (Rptrs) utilized by CjeCas9. W = A or T. PAM, yellow circle. (**B**) Toleration of mutations to the RNA duplex of an sgRNA that preserve secondary structure in TXTL. (**C**) Enhancing DNA targeting by less functional or non-functional ncrRNAs by converting the tracrRNA into a Rptr in TXTL. (**D**) Sequence-specific DNA targeting in TXTL using Rptrs compatible with three different Cas9 nucleases. Values in (B) to (D) represent the mean and standard deviation from four replicates in TXTL. *$p < 0.01$; **$p < 0.001$. n.s., not significant.

would be expected to direct Cas9 to complementary DNA targets flanked by a protospacer-adjacent motif (PAM), similar to a canonical crRNA (*12*). As none of the genes giving rise to the detected ncrRNAs has a correctly placed PAM, the ncrRNAs are not expected to direct Cas9 to cleave their originating genomic site (table S1).

To evaluate ncrRNA-dependent targeting, we exploited a cell-free transcription-translation (TXTL) assay previously used to characterize CRISPR-Cas systems (*13*–*15*). As part of the assay, DNA constructs encoding CjeCas9, an RNA guide, and a green fluorescent protein (GFP) reporter harboring a target sequence flanked by a recognized PAM are added to the TXTL reaction. GFP fluorescence is then measured over time as a readout of DNA binding and cleavage by CjeCas9 (Fig. 2B and fig. S7A). We focused on examining the *fliF* ncrRNA given its presence in both *C. jejuni* strains and detection by Northern blotting. Applying this assay to the tracrRNA and mRNA comprising the entire *fliF* coding region (1683 nts) (Fig. 2C), we found that expressing the mRNA reduced GFP levels 2.5-fold compared with a nontargeting crRNA ($p = 5.4 \times 10^{-5}$). Expressing the equivalent crRNA reduced GFP levels 15.1-fold compared with the nontargeting control. The reduced GFP silencing for the *fliF* mRNA versus the crRNA potentially reflects not only reduced targeting efficiency but also delayed complex formation. Overall, the TXTL results offer evidence that mRNA-derived ncrRNAs can direct DNA targeting by Cas9.

The reduced performance of the *fliF* mRNA in TXTL could be due to how an ncrRNA deviates from a standard crRNA. These deviations include the crRNA repeat sequence, the secondary structure of the duplex formed with the tracrRNA anti-repeat, and 5′ or 3′ extensions to the repeat that do not undergo efficient processing. To evaluate these deviations, we systematically mutated or extended the standard crRNA, either as a single guide RNA (sgRNA) to ensure duplex formation or as a crRNA:tracrRNA pair, and evaluated GFP silencing in TXTL (Fig. 2D and table S1). CjeCas9 could accommodate some mutations within the region of the repeat:anti-repeat duplex in the sgRNA implicated in nuclease binding (*16*). The more disruptive mutations spanned more nts, were closer to the 5′ end of the repeat, or resulted in a bulge in the tracrRNA (e.g., s3, s6, s7, s9, s12, s18 in Fig. 2D, left). Observed differences in GFP silencing do not appear to arise from variable sgRNA levels (fig. S7B). Extending the sgRNA-crRNA ends or mutating the region cleaved by RNase III within the crRNA had minimal impact on GFP silencing (Fig. 2D, right). Overall, the tracrRNA can tolerate deviations from a standard crRNA as long as pairing through the 3′ end of the tracrRNA anti-repeat is maintained.

We applied insights from our mutational analyses to prioritize putative ncrRNAs from *C. jejuni* CG8421 for functional tests in TXTL. In total, we identified eight RNA fragments predicted to base pair extensively with the 3′ end of the tracrRNA anti-repeat (Fig. 2E and fig. S8). We then assessed GFP silencing by expressing up to 350 nts upstream and downstream of each associated ncrRNA-encoding gene with tracrRNA, CjeCas9, and the GFP reporter harboring each cognate DNA target. Of the eight tested RNAs, three (from *rseP*, *nuoL*, and *dctA*) yielded a >twofold reduction in GFP reporter levels compared with a nontargeting crRNA control ($p < 0.001$). Furthermore, targeting was directed specifically through the predicted ncrRNA, as mutating the "seed"

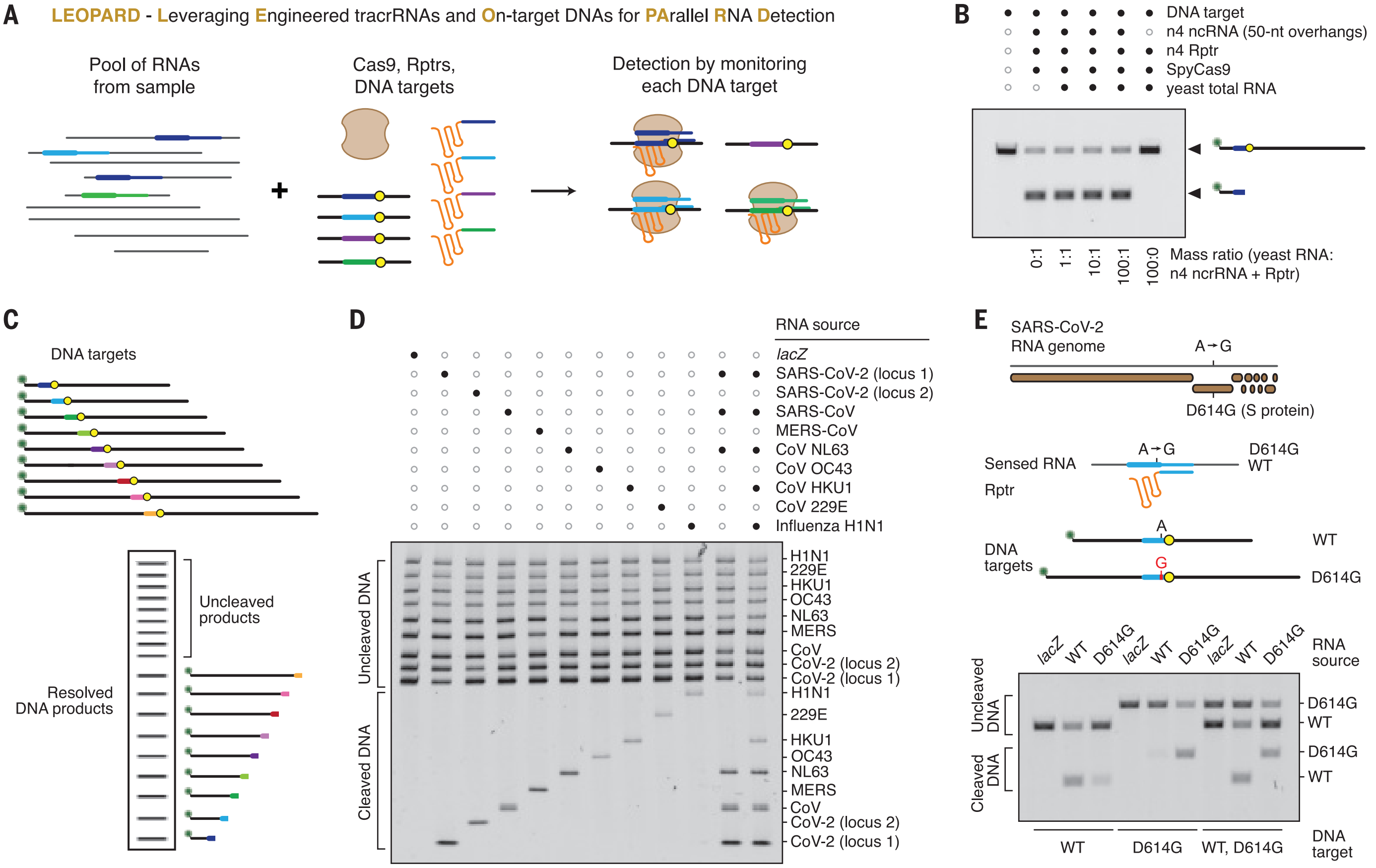


Fig. 4. Reprogrammed tracrRNAs enable multiplexable RNA detection with single-base resolution in vitro. (**A**) Overview of the multiplexed diagnostic platform LEOPARD. (**B**) Highly specific cleavage of target DNA by SpyCas9 in vitro. Targeting was directed by the ncrRNA associated with the n4 locus in CJ8421_04975. (**C**) Multiplexed monitoring of DNA targets. (**D**) Parallel detection of nine different RNA fragments associated with respiratory viruses by denaturing polyacrylamide gel electrophoresis. Filled dots in (D) and (E) indicate the presence of the component listed on the right. (**E**) Specific detection of the D614G point mutation within the spike protein of SARS-CoV-2. Each detected T7-transcribed RNA in (B), (D), and (E) comprised the intended ncrRNA sequence with 50-nt extensions on either end.

region of the putative ncrRNA (*17*), scrambling the tracrRNA anti-repeat, or replacing CjeCas9 with the orthogonal FnCas12a nuclease fully relieved GFP repression (Fig. 2E). Multiple factors, such as mRNA folding or accessibility during translation, may explain why the other five ncrRNAs did not exhibit targeting activity in TXTL, as a linear-regression model built around the sgRNA mutants had limited ability to predict the targeting activity of these ncrRNAs (supplementary text S1). The current lack of predictability parallels guide design for RNA-targeting Cas13a nucleases, which only became predictable with extensive datasets and machine learning (*18*).

Beyond TXTL, we assessed ncrRNA function as part of DNA targeting in *C. jejuni* CG8421 and in *Escherichia coli*. For CG8421, transformation interference assays did not yield any significant DNA targeting directed by ncrRNAs derived from the *rseP*, *dctA*, and *nuoL* mRNAs (fig. S9A), likely due to low ncrRNA abundance compared with the strain's crRNAs under the examined growth conditions. For *E. coli*, overexpressing the *dctA* mRNA, CjeCas9, and the tracrRNA led to moderate (15.5-fold) clearance of a transformed plasmid with the putative *dctA* ncrRNA target ($p = 0.0036$), but not when the tracrRNA anti-repeat was scrambled (1.6-fold) ($p = 0.068$) (fig. S9B). We therefore conclude that ncrRNAs derived from mRNAs can elicit DNA targeting in both in vivo and cell-free systems.

The tracrRNA can be reprogrammed to direct Cas9 activity by an RNA of interest

The conversion of a cellular RNA into an ncrRNA was based on sequences bearing complementarity to the tracrRNA anti-repeat, analogous to natural crRNA biogenesis (fig. S2A). What if the tracrRNA anti-repeat sequence could be changed to hybridize to other RNAs while maintaining the appropriate structure for Cas9 recognition? If so, then the resulting reprogrammed tracrRNA (Rptr, pronounced "raptor") could specifically derive an ncrRNA from a cellular RNA. The resulting ncrRNAs can then guide Cas9 to matching DNA targets (Fig. 3A). Although tracrRNA engineering has rarely been explored outside of sgRNAs or crRNA:tracrRNA duplexes (*19*), multiple studies have shown that the repeat:anti-repeat duplex of the sgRNA for the *Streptococcus pyogenes* Cas9 (SpyCas9) can be extensively modified as long as the secondary structure is maintained (*20*, *21*).

CjeCas9 recognizes a perfect RNA duplex formed between the crRNA repeat and tracrRNA anti-repeat (fig. S2B) (*16*). Based on our mutational analysis of the crRNA repeat, we already observed that Cas9 can accommodate several mutations within the crRNA repeat: tracrRNA anti-repeat duplex (Fig. 2D). Therefore, we evaluated GFP silencing in TXTL after mutating both sides of the duplex in the *fliF* sgRNA while preserving the secondary structure (s26-s31, Fig. 3B). GFP silencing was maintained even when exchanging the sequence of the entire duplex (s31, Fig. 3B). Next, we reprogrammed the tracrRNA anti-repeat to form perfect 25–base pair duplexes with three putative ncrRNAs (derived from *fliF*, CJ8421_04245, CJ8421_04975) exhibiting at most modest GFP silencing in TXTL (Fig. 2, C and E, and fig. S8). In all three cases, GFP repression was significantly enhanced with Rptrs compared with wild-type (WT) tracrRNAs ($p = 1 \times 10^{-6}$ to 0.0011), even if repression was not as strong as with the canonical crRNA:tracrRNA pair (Fig. 3C). Finally, we reprogrammed the tracrRNA anti-repeat to base pair with entirely new regions of an mRNA (Fig. 3D). Starting with the CJ8421_04975 mRNA, we designed five different CjeCas9 Rptrs hybridizing to different locations (n1 to n5) in the mRNA (Fig. 3D and fig. S10A). Of these Rptrs, four yielded significantly reduced GFP levels compared with nontargeting crRNA controls ($p = 6 \times 10^{-7}$ to 0.002). Notably, mutating the predicted seed region, scrambling the tracrRNA anti-repeat, or replacing CjeCas9 with FnCas12a restored GFP expression (Fig. 3D). Northern blot analysis from TXTL-extracted RNAs further revealed no detectable processed RNAs with a size resembling that of mature ncrRNA. Complete ncrRNA processing to a size similar to that of canonical crRNAs therefore may not be necessary for DNA targeting by CjeCas9 (fig. S11), in line with the dispensability of RNase III for crRNA-mediated DNA targeting through the II-C CRISPR-Cas system in *Neisseria meningitidis* (*22*).

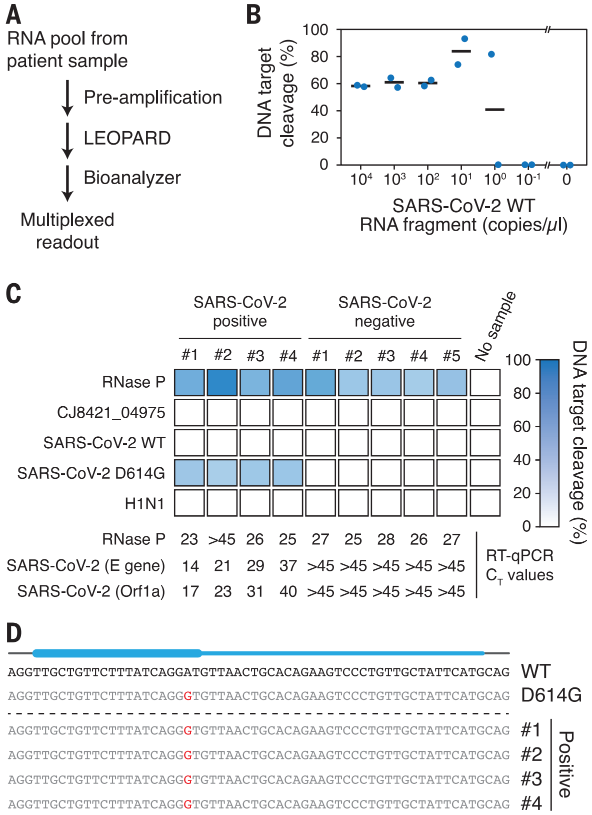


Fig. 5. LEOPARD with RNA preamplification and Bioanalyzer readout allows for multiplexed detection of SARS-CoV-2 in patient samples. (**A**) General workflow for LEOPARD with target-specific preamplification and DNA target resolution on a Bioanalyzer. (**B**) Sensitivity of the workflow for detecting dilutions of an in vitro–transcribed WT SARS-CoV-2 RNA fragment. One microliter was added for each test. Bars represent the average of independent duplicates. (**C**) Multiplexed detection of five RNAs in patient samples confirmed positive or negative for SARS-CoV-2 by RT-qPCR. (**D**) Sanger sequencing results of the detected region in SARS-CoV-2 cDNA from the positive patient samples. Blue bar: Position of the ncrRNA, with the thick part indicating the resulting ncrRNA portion.

Given the functionality of CjeCas9 Rptrs, we asked whether tracrRNAs for other Cas9 homologs can be similarly reprogrammed. We selected the well-characterized *Streptococcus pyogenes* Cas9 (SpyCas9) and the *Streptococcus thermophilus* CRISPR1 Cas9 (Sth1Cas9) as examples. In both cases, we devised design rules for Rptrs based on the known secondary structure of the crRNA:tracrRNA duplex and the preference of RNase III to cleave double-stranded RNA with AT-rich sequences (fig. S10A) (*20*, *23*). All 10 designed Rptrs significantly reduced GFP levels compared with the nontargeting crRNA control ($p = 1 \times 10^{-7}$ to 1×10^{-4}) (Fig. 3D). As before, GFP expression was restored by disrupting the seed sequence in the ncrRNA guide, scrambling the tracrRNA anti-repeat, or swapping either of the Cas9's for FnCas12a. In many cases, the extent of GFP silencing approached that of the targeting crRNA control. We also evaluated plasmid clearance with Rptrs in *E. coli* for all three Cas9 orthologs, finding that each could elicit efficient plasmid clearance for at least one tested Rptr (fig. S10B). The targeted plasmid was efficiently cleared even when expressing the sensed mRNA at low levels (fig. S12A) or when deleting RNase III (fig. S12B). Overall, the tracrRNA for different Cas9 orthologs can be converted into Rptrs to elicit DNA targeting based on the presence of a selected cellular RNA.

Reprogrammed tracrRNAs enable sequence-specific detection by Cas9

By linking DNA targeting to an RNA of interest, Rptrs offer a valuable opportunity for RNA detection and a different paradigm for CRISPR diagnostics. Current CRISPR diagnostics principally rely on Cas12a or Cas13 searching for double-stranded DNA or RNA targets in a sample, where target recognition elicits nonspecific single-stranded DNA or RNA cleavage of a fluorescent reporter (*24*–*26*). The nonspecific readout practically limits one test to one target sequence. By contrast, Rptrs convert sensed RNAs into ncrRNAs, which would direct Cas9 to matching DNA. Cas9 binding or cleavage of a DNA sequence would then indicate the presence of the sensed RNA in the sample. Because the sequence of each DNA target is distinctive, large numbers of target sequences could be monitored in parallel in one test. We call the resulting diagnostic platform LEOPARD, for leveraging engineered tracrRNAs and on-target DNAs for parallel RNA detection (Fig. 4A).

To begin assessing LEOPARD, we performed a simplified in vitro reaction using T7-transcribed RNAs, commercially available SpyCas9 protein, and linear DNA targets (fig. S13A). We began with RNA corresponding to one of the synthetic ncrRNA loci within CJ8421_04975 (n4 under SpyCas9, Fig. 3D). Introducing an annealing step to hybridize the Rptr to the T7-transcribed ncrRNA yielded DNA target cleavage without adding RNase III or RNase A for ncrRNA:Rptr processing (fig. S13, B and C). The cleavage efficiency was also similar to that of the equivalent crRNA:tracrRNA pair, even when the ncrRNA sequence was extended on either end (fig. S13D). The time scale of the annealing step could also be minimized by rapid cooling of the samples (fig. S13E). With the annealing step, efficient cleavage occurred with a 100-fold excess of yeast total RNA but only when the ncrRNA was present (Fig. 4B). LEOPARD therefore can report the presence of a specific RNA of interest based on cleavage of a DNA target and can be streamlined through further optimization.

LEOPARD allows for multiplexed RNA detection by Cas9 with single-base resolution

Realizing the full multiplexing potential of LEOPARD requires monitoring many DNA targets at once. To initially demonstrate this multiplex capability, we devised a readout scheme based on resolving distinct cleavage products from pooled DNA targets by gel electrophoresis (Fig. 4C). Each target is labeled with a fluorophore on one end, producing only two visualizable products—cleaved and uncleaved. We then applied this scheme to specifically detect nine ~150-nt RNA fragments associated with respiratory viruses, including two from SARS-CoV-2 coronavirus (the causative agent of COVID-19), six from other coronaviruses, and one from influenza H1N1 (Fig. 4D and fig. S14). Each DNA target was cleaved by Cas9 only in the presence of the corresponding RNA, even when detecting three or five specific RNA fragments in the same reaction (Fig. 4D and fig. S14).

As the viral RNA sequences were selected to minimize homology, we asked if LEOPARD could detect even a single-nucleotide difference. The Asp^{614}→Gly (D614G) mutation in the spike protein of SARS-CoV-2 served as an example, as it comprises a single base change (A23403G) that increased infectivity and drove global spread (*27*). By placing this nt change within the seed region of the target, we could detect the WT or D614G RNA using one Rptr combined with either the WT or D614G target (Fig. 4E). The matching DNA target was preferentially cleaved when testing each target individually, although some cleavage of the nonmatching target was observed. However, combining the two targets in a single reaction yielded discernable cleavage only for the matching target, presumably through preferential binding and cleavage of the perfect target by Cas9 (*28*). LEOPARD therefore can confer multiplexed RNA detection in a single reaction with single-base resolution.

To extend LEOPARD beyond this proof-of-principle demonstration, we made two additions. First, we added target-specific reverse transcription–polymerase chain reaction (RT-PCR) and in vitro transcription similar to Cas13-based diagnostics (*29*) to improve assay sensitivity beyond the threshold set by detection of a cleaved DNA product. Second, we resolved DNA targets using a Bioanalyzer as a more practical readout (Fig. 5A). Applying this modified workflow to sense the in vitro–transcribed WT SARS-CoV-2 RNA fragment, we could detect as little as approximately one copy, or 1.7 aM in the original dilution, of this RNA (Fig. 5B and fig. S15A) compared with 3×10^8 copies, or 0.6 nM in the original dilution, without preamplification (fig. S15B). As this sensitivity would be sufficient for detecting SARS-CoV-2 RNA in patient samples, we applied LEOPARD to evaluate samples confirmed positive or negative for SARS-CoV-2 by RT-qPCR (Fig. 5C and table S1). The positive samples reflected a range of SARS-CoV-2 RNA concentrations down to ~2 aM. We then probed for both SARS-CoV-2 and the D614G variant as well as influenza H1N1, the *C. jejuni* CJ8421_04975 mRNA (n4 under SpyCas9, Fig. 3D) as a nonhuman negative control, and the mRNA encoding human RNase P to confirm correct administration of the nasal swab using four Rptrs [one for both WT and D614G SARS-CoV-2 (Fig. 4E)] and five DNA targets. Of the four SARS-CoV-2–positive and five SARS-CoV-2–negative samples tested, RNase P mRNA but not the CJ8421_04975 mRNA or H1N1 RNA was detected in all nine samples (Fig. 5C and fig. S16). Notably, we detected the D614G variant of SARS-CoV-2 in all four positive samples, which was confirmed by sequencing preamplified cDNA (Fig. 5D). Although the sample size is small, detection of this variant suggests that it was spreading in Germany when the samples were collected. WT or D614G SARS-CoV-2 RNA was not detected in any of the negative samples, paralleling the RT-qPCR results (Fig. 5C). Each reaction allowed for parallel testing for five different RNAs, including controls, that would require separate reactions for other diagnostic platforms. These findings demonstrate the practical utility of LEOPARD for multiplexed RNA detection.

Discussion

Starting from the characterization of a native CRISPR-Cas9 system in the bacterial pathogen *C. jejuni*, we discovered that cellular transcripts can be the source of noncanonical crRNAs through hybridization with the tracrRNA. This discovery adds ncrRNAs to the list of RNA guides found in nature, including crRNAs, scaRNAs that similarly pair with the tracrRNA anti-repeat, and “natural” sgRNAs formed through upstream transcription of the tracrRNA (*30*, *31*). These prior examples are all encoded within CRISPR-Cas loci. So far, it remains unclear whether ncrRNAs serve a physiological role in *C. jejuni*. Future studies therefore could help clarify whether ncrRNAs derived from outside CRISPR arrays are spurious off-target products that are tolerated by the host or if they confer yet-to-be-discovered functions extending beyond adaptive immunity.

We further demonstrated that Rptrs can link the presence of an RNA of interest to sequence-specific DNA targeting by Cas9. This capability could enable in vivo applications with Cas9 such as multiplexed transcriptional recording or transcription-dependent editing. The most immediate application involved multiplexed RNA detection in vitro through LEOPARD. LEOPARD adds to the existing CRISPR diagnostic platforms principally based on Cas12a or Cas13 (*24*–*26*, *29*) while offering scalable multiplexing in a single reaction. Our reliance on gel electrophoresis or a Bioanalyzer provided a proof-of-principle demonstration of multiplexed detection, although both are difficult to implement beyond a dozen targets. Instead, incorporating microarrays or next-generation sequencing (*32*) can potentially monitor up to millions of targets by linking the presence of a specific RNA to binding of labeled Cas9 or cleavage of labeled DNA target at a specific location on a chip. Either approach could also enhance assay sensitivity due to the limited number of DNA molecules in a given cluster, potentially circumventing the need for RNA preamplification. Simpler setups involving lateral flow assays could also be developed, paralleling other CRISPR diagnostics (*33*, *34*). With further development, LEOPARD could become a powerful diagnostic tool not only for the detection of viral variants distinguished by individual nts but also for applications such as screening for cancer mutations, identifying pathogens and antibiotic-resistance markers, or determining gene expression profiles for drug susceptibility. Moreover, extending Rptrs to tracrRNA-dependent nucleases within type V systems could help incorporate their distinctive attributes, such as signal amplification or programmable transposition (*24*, *35*).

While systematically perturbing the standard crRNA:tracrRNA duplex, we found that many deviations–particularly outside of the 5′ end of the repeat–were tolerated by Cas9 and still led to targeting of designed DNA targets. However, despite this promiscuity, targeting is still determined by the upstream guide sequence and the requirement for a flanking PAM. Thus, both anti-repeat hybridization and guide-dependent DNA targeting may limit off-targeting activity. Although we did not observe any detectable off-targeting in vitro or in vivo, future work could devise design rules for Rptrs that account for potential off-targeting as well as on-target activity, similar to existing sgRNA design algorithms (*36*). In turn, these rules would help advance the utilization of any RNA into a sequence-specific guide for CRISPR technologies.

REFERENCES AND NOTES

1. R. Barrangou *et al*., *Science* **315**, 1709–1712 (2007).
2. K. S. Makarova *et al*., *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
3. S. J. J. Brouns *et al*., *Science* **321**, 960–964 (2008).
4. G. J. Knott, J. A. Doudna, *Science* **361**, 866–869 (2018).
5. A. V. Anzalone, L. W. Koblan, D. R. Liu, *Nat. Biotechnol.* **38**, 824–844 (2020).
6. E. Deltcheva *et al*., *Nature* **471**, 602–607 (2011).
7. T. Karvelis *et al*., *RNA Biol.* **10**, 841–851 (2013).
8. M. Jinek *et al*., *Science* **337**, 816–821 (2012).
9. G. Dugar *et al*., *Mol. Cell* **69**, 893–905.e7 (2018).
10. T. L. Bailey *et al*., *Nucleic Acids Res.* **37**, W202–W208 (2009).
11. J. N. Zadeh *et al*., *J. Comput. Chem.* **32**, 170–173 (2011).
12. R. T. Leenay, C. L. Beisel, *J. Mol. Biol.* **429**, 177–191 (2017).
13. R. Marshall *et al*., *Mol. Cell* **69**, 146–157.e3 (2018).
14. C. Liao *et al*., *Nat. Commun.* **10**, 2948 (2019).
15. T. Jacobsen *et al*., *Nucleic Acids Res.* **48**, 5624–5638 (2020).
16. M. Yamada *et al*., *Mol. Cell* **65**, 1109–1121.e3 (2017).
17. X. Wu, A. J. Kriz, P. A. Sharp, *Quant. Biol.* **2**, 59–70 (2014).
18. H.-H. Wessels *et al*., *Nat. Biotechnol.* **38**, 722–727 (2020).
19. T. Scott, R. Urak, C. Soemardy, K. V. Morris, *Sci. Rep.* **9**, 16104 (2019).
20. A. E. Briner *et al*., *Mol. Cell* **56**, 333–339 (2014).
21. A. C. Reis *et al*., *Nat. Biotechnol.* **37**, 1294–1301 (2019).
22. Y. Zhang *et al*., *Mol. Cell* **50**, 488–503 (2013).
23. Y. Altuvia *et al*., *Nucleic Acids Res.* **46**, 10530–10531 (2018).
24. J. S. Chen *et al*., *Science* **360**, 436–439 (2018).
25. A. East-Seletsky *et al*., *Nature* **538**, 270–273 (2016).
26. J. S. Gootenberg *et al*., *Science* **356**, 438–442 (2017).
27. B. Korber *et al*., *Cell* **182**, 812–827.e19 (2020).
28. E. A. Boyle *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5461–5466 (2017).
29. C. M. Ackerman *et al*., *Nature* **582**, 277–282 (2020).
30. R. E. Workman *et al*., *Cell* **184**, 675–688.e19 (2021).
31. H. K. Ratner *et al*., *Mol. Cell* **75**, 498–510.e5 (2019).
32. C. Jung *et al*., *Cell* **170**, 35–47.e13 (2017).
33. J. P. Broughton *et al*., *Nat. Biotechnol.* **38**, 870–874 (2020).
34. J. S. Gootenberg *et al*., *Science* **360**, 439–444 (2018).
35. J. Strecker *et al*., *Science* **365**, 48–53 (2019).
36. R. E. Hanna, J. G. Doench, *Nat. Biotechnol.* **38**, 813–823 (2020).
37. T. Bischler, K. U. Förstner, D. Maticzka, P. R. Wright, PEAKachu: A peak calling tool for CLIP/RIP-seq data, Version 0.2.0, Zenodo (2021). https://zenodo.org/record/4669966.
38. L. Barquist, Perl script for evaluating tracr:target interactions with Nupack, Zenodo (2021). https://zenodo.org/record/4668694.

ACKNOWLEDGMENTS

We thank T. Achmedov for assistance with Northern blotting; J. Yu and E. Vialetto for help with cloning; D. Collias for helping devise the acronym LEOPARD; D. Collias, P. Fineran, I. Mougiakos, and

J. Vogel for critical feedback on the manuscript; P. Guerry for sharing the *C. jejuni* CG8421 strain; R. Louwen for serum against CjeCas9; C. Homberger for technical support with the Bioanalyzer; N. Caliskan for providing human DNA; V. Corman for providing SARS-CoV-2 RNA for test validation; and T. Höfert and L. Zubrod for conducting RT-qPCR with COVID-19 patient samples. **Funding:** Support was provided by Deutsche Forschungsgemeinschaft SH 580/9-1 (C.M.S.) and BE 6703/1-1 (C.L.B.) Interdisciplinary Center for Clinical Research Würzburg (IZKF) project Z-6 Free State of Bavaria for COVID-19 research at the IHM. **Author contributions:** Conceptualization: C.J., S.S., G.D., C.M.S., C.L.B.; Discovery of ncrRNAs: G.D., C.M.S.; Methodology: C.J., S.S., G.D., C.M.S., C.L.B.; NGS Software: T.B.; Formal Analysis: C.J., S.S., G.D., T.B., C.M.S., C.L.B.; Investigation: C.J., S.S., G.D., N.L.P., T.B., F.W.; Linear-regression model: Y.Y., L.B.; Obtaining and assessing patient samples: C.S., O.K.; Writing – Original Draft: C.J., S.S., C.L.B.; Writing – Review and Editing: C.J., S.S., G.D., N.L.P., T.B., F.W., C.S., C.M.S., C.L.B.; Visualization: C.J., S.S., C.L.B.; Supervision: C.M.S., C.L.B.; Funding Acquisition: C.M.S., C.L.B.. **Competing interests:** A provisional patent application has been filed on the related concept by C.J., G.D., C.M.S., and C.L.B.. The other authors declare that they have no competing interests. **Data and materials availability:** The NGS data from RIP-seq and differential RNA-seq were deposited into NCBI GEO under accession number GSE156266. Code for automating NUPACK folding predictions and calling enriched RNA peaks are available through Zenodo (*37*, *38*). All other data in the main text or the supplementary materials are available upon reasonable request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/941/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Table S1
References (*39*–*52*)
MDAR Reproducibility Checklist

9 September 2020; accepted 9 April 2021

10.1126/science.abe7106

REPORTS

QUANTUM COMPUTING

Quantum walks on a programmable two-dimensional 62-qubit superconducting processor

Ming Gong[1,2,3]†, Shiyu Wang[1,2,3]†, Chen Zha[1,2,3]†, Ming-Cheng Chen[1,2,3], He-Liang Huang[1,2,3], Yulin Wu[1,2,3], Qingling Zhu[1,2,3], Youwei Zhao[1,2,3], Shaowei Li[1,2,3], Shaojun Guo[1,2,3], Haoran Qian[1,2,3], Yangsen Ye[1,2,3], Fusheng Chen[1,2,3], Chong Ying[1,2,3], Jiale Yu[1,2,3], Daojin Fan[1,2,3], Dachao Wu[1,2,3], Hong Su[1,2,3], Hui Deng[1,2,3], Hao Rong[1,2,3], Kaili Zhang[1,2,3], Sirui Cao[1,2,3], Jin Lin[1,2,3], Yu Xu[1,2,3], Lihua Sun[1,2,3], Cheng Guo[1,2,3], Na Li[1,2,3], Futian Liang[1,2,3], V. M. Bastidas[4], Kae Nemoto[5], W. J. Munro[4,5], Yong-Heng Huo[1,2,3], Chao-Yang Lu[1,2,3], Cheng-Zhi Peng[1,2,3], Xiaobo Zhu[1,2,3]*, Jian-Wei Pan[1,2,3]*

Quantum walks are the quantum mechanical analog of classical random walks and an extremely powerful tool in quantum simulations, quantum search algorithms, and even for universal quantum computing. In our work, we have designed and fabricated an 8-by-8 two-dimensional square superconducting qubit array composed of 62 functional qubits. We used this device to demonstrate high-fidelity single- and two-particle quantum walks. Furthermore, with the high programmability of the quantum processor, we implemented a Mach-Zehnder interferometer where the quantum walker coherently traverses in two paths before interfering and exiting. By tuning the disorders on the evolution paths, we observed interference fringes with single and double walkers. Our work is a milestone in the field, bringing future larger-scale quantum applications closer to realization for noisy intermediate-scale quantum processors.

The quantum principles of superposition and entanglement allow a more powerful form of random walk, termed quantum walk (QW) (*1*). These walks have attracted considerable attention with many applications known in quantum transport (*2*), quantum simulations (*3*, *4*), quantum search algorithms (*5*, *6*), and even universal quantum computing (*7*, *8*). The universality of QWs has been shown with encoded quantum computation, meaning that they are a universal quantum computation primitive and may give an exponential algorithmic speed-up (*9*). In particular, QWs with multiple walkers show a quantum advantage (*10*, *11*) that is superior to ones that use only a single walker. Motivated by QWs' rich potential applications, numerous proof-of-principle experimental demonstrations have been performed in a wide variety of hardware platforms, ranging from photonics (*12*, *13*), trapped ions (*14*, *15*), and neutral atoms (*16*) to nuclear magnetic resonance (*17*) and superconducting qubits (*18*, *19*).

It is well known that QW-based quantum search algorithms require at least a two-dimensional (2D) configuration (*20*). Furthermore, easy circuit programmability with an arbitrary number of walkers is an essential requirement to encode applications, in which the configuration can be changed on a walk-by-walk basis, including the adjustability of tunneling amplitude and graph structure (*21*). Achieving both of these simultaneously has proved experimentally challenging. Superconducting circuits provide the nonlinear interaction Hamiltonian necessary for the universal quantum computation, making them one of the leading quantum computer approaches (*22*). Along with the real-time programmability, they are now an excellent candidate system for the realization of fully configurable 2D QWs.

In our work, we started with the design of a moderate-scale 2D superconducting qubit array. Immediately, there was the problem associated with planar wiring and how it can be realized to control all the qubits as the size of the 2D array increases. One solution has been 3D wiring by using techniques such as "flip chip" (*23*) or "through-silicon vias" (*24*). In this work, we provide an alternative technical-friendly solution based on "pass-through holes" (*25*). This is applied to an 8-by-8 qubit array (Fig. 1A) that is composed of 16 units whose circuit diagram is shown in Fig. 1B. Two of the qubits, U03Q2 and U22Q1 (Fig. 1C), and one coupling resonator (between U10Q0 and U10Q3) are nonfunctional.

We represent the effective Hamiltonian of the qubit system using the Bose-Hubbard model as

$$\hat{H} = \sum\nolimits_{j\in\{Q_i\}} \hbar\omega_j \hat{a}_j^\dagger \hat{a}_j + \frac{\hbar U_j}{2}\hat{n}_j(\hat{n}_j - 1) + \sum\nolimits_{j\in\{Q_i\},i\in\left\{C_{Q_i}\right\}} \hbar J_{\mathrm{eff}}^{i,j}\left(\hat{a}_i^\dagger \hat{a}_j + \hat{a}_i \hat{a}_j^\dagger\right) \quad (1)$$

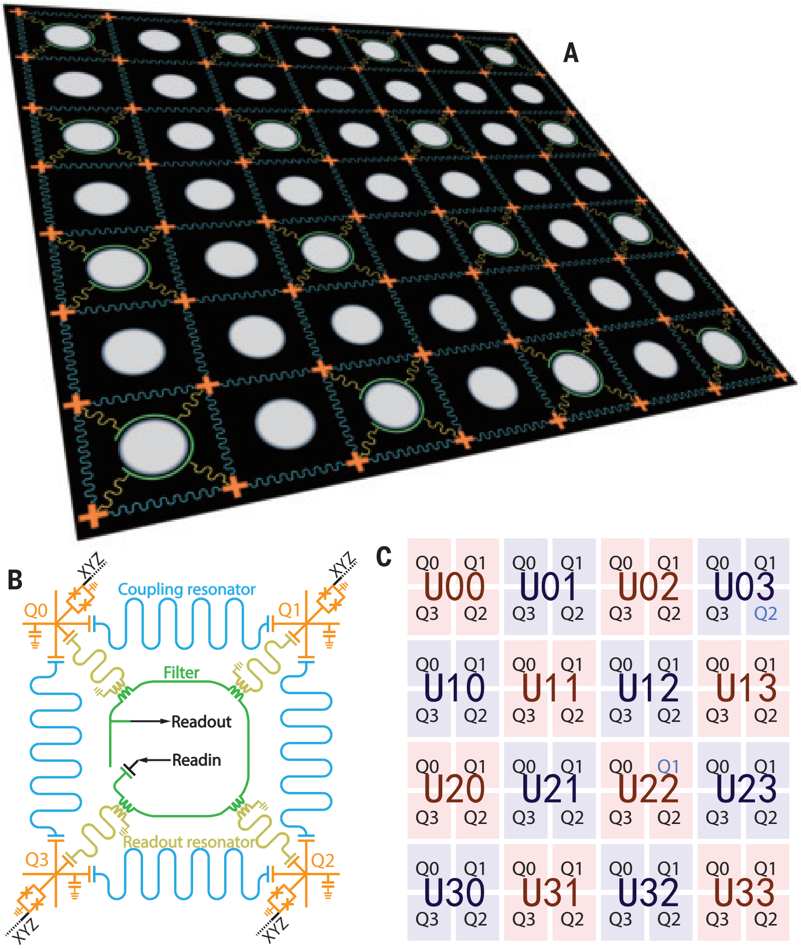
where $\hat{a}_j^\dagger$ is the qubits bosonic creation operator, $\hat{a}_j$ is the qubits bosonic annihilation operator, ω_j is the j^{th} qubit frequency, U_j is the anharmonicity, and $J_{\mathrm{eff}}^{i,j}$ is the effective coupling strength between Q_i and Q_j by means of a large detuned coupling resonator, where i

---

[1] Hefei National Laboratory for Physical Sciences at the Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China. [2] Shanghai Branch, CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China. [3] Shanghai Research Center for Quantum Sciences, Shanghai 201315, China. [4] NTT Basic Research Laboratories and Research Center for Theoretical Quantum Physics, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan. [5] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.

***Corresponding author. E-mail: xbzhu16@ustc.edu.cn (X.Z.); pan@ustc.edu.cn (J.-W.P.)** †These authors contributed equally to this work.

Fig. 1. The layout and architecture of the superconducting quantum processor. (**A**) The schematic diagram of the 2D superconducting quantum processor. The orange crosses represent the qubits arranged in an 8-by-8 array. The gray circles are the pass-through holes (*25*) for 3D wiring. The electrodes for wiring are not shown for simplification. (**B**) The circuit diagram of a unit of the qubit array. Each qubit (orange) has an *XYZ* control line (black) for microwave and pulse control. The qubit couples to an individual $\lambda/4$ readout resonator (yellow), which in turn is commonly coupled to a filter (green). Two neighboring qubits are dispersively coupled to each other through a $\lambda/2$ coupling resonator (blue). (**C**) The labels of qubits. Two broken qubits, U03Q2 and U22Q1, are marked in blue.



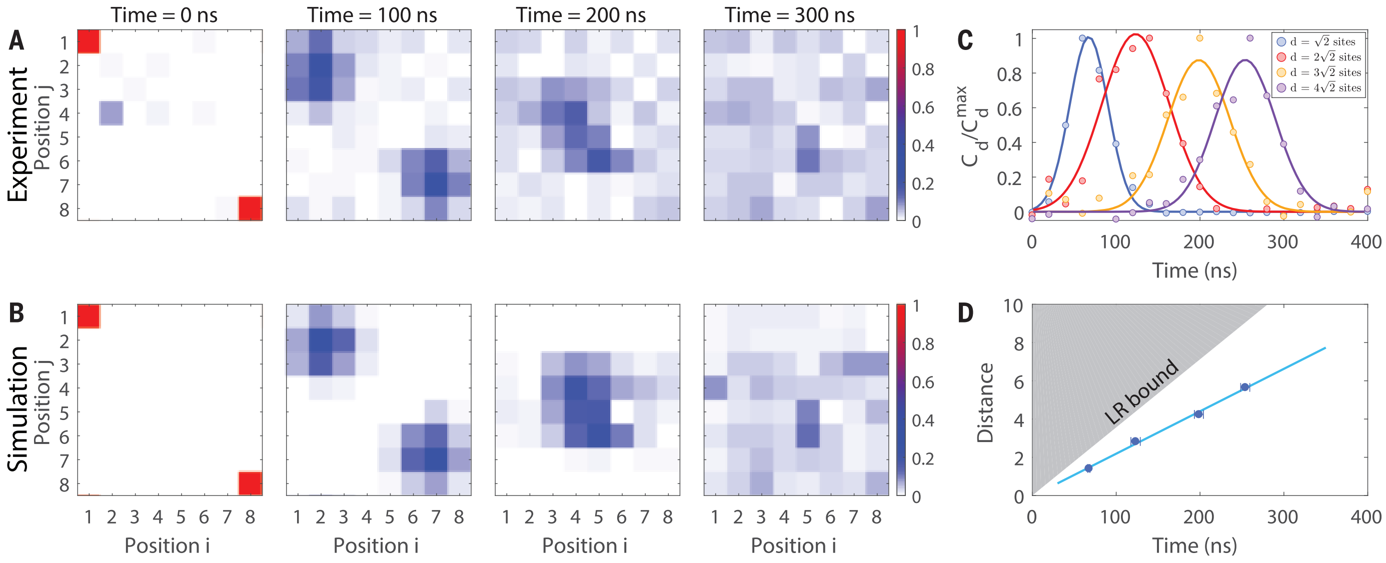


Fig. 2. QWs on a 2D superconducting qubits array. (**A**) The evolution of the measured populations $\langle n_j \rangle$ of all qubits at times t = 0, 100, 200, and 300 ns, respectively, with the two walkers initialized on qubits U00Q0 and U33Q2. (**B**) Numerical simulation of the qubits population evolution under the same conditions as in (A). (**C**) The correlation function as a function of time in the case of single-particle QW. The blue, red, orange, and purple circles represent the measured data points for the correlation function between the initial excitation site and the sites on the diagonal with distance $d = \sqrt{2}, 2\sqrt{2}, 3\sqrt{2}$, and $4\sqrt{2}$sites, respectively. The corresponding solid curves are Gaussian fittings to the data, and the propagation fronts are the center of the Gaussian fittings. (**D**) The propagation velocity and the LR bounds. Using a linear fit (blue line) of the propagation fronts (blue circles) with distance, we determined the propagation velocity as 22.2 ± 2.0 site/μs. The gray shadow indicates LR bound with $v_{\max}$ = 35.7 site/μs.

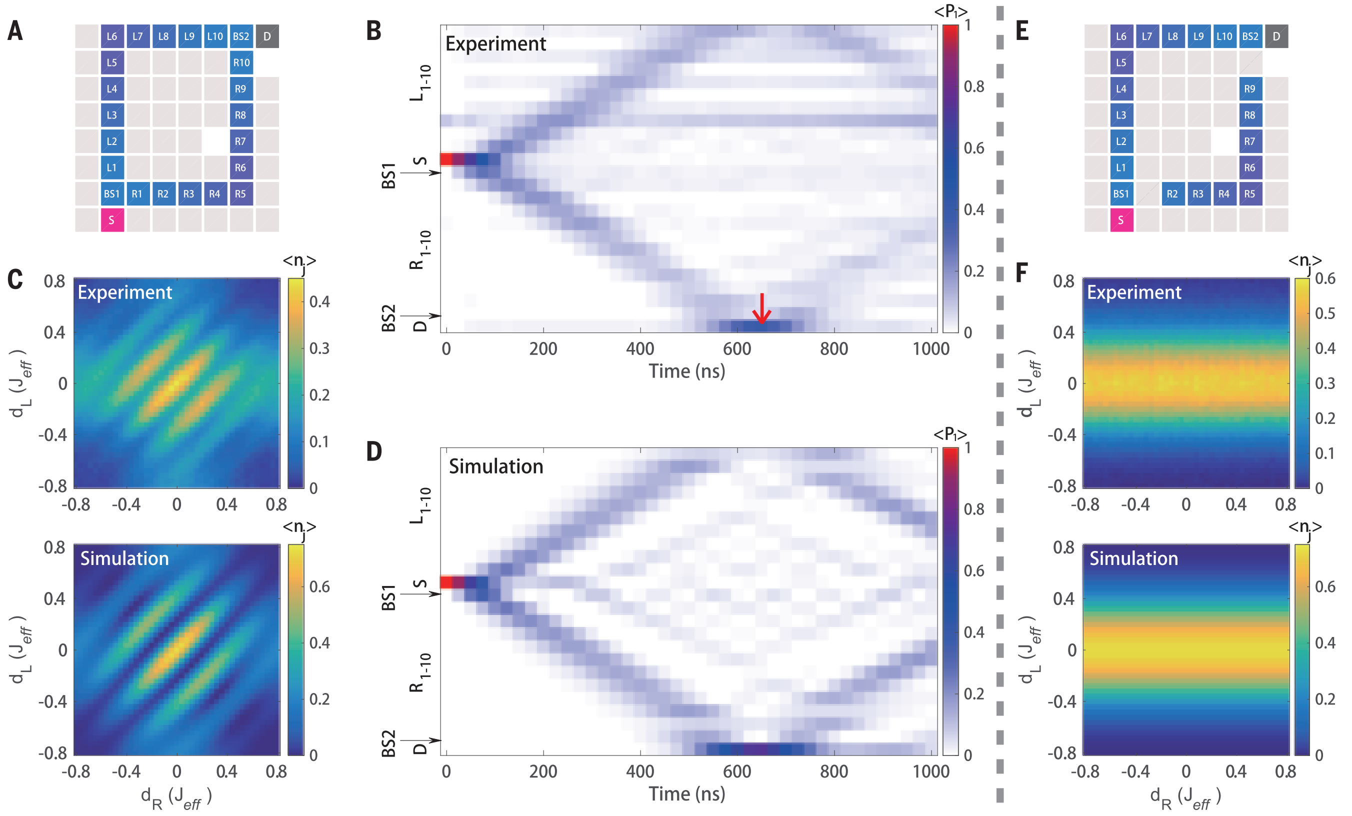


Fig. 3. Single-particle MZ interferometer. (**A**) The circuit diagram of the programmable paths for the realization of the single-particle MZ interferometer in a qubit array. (**B** and **D**) Illustration of the dynamics evolution of population $\langle n_j \rangle$ of all relevant sites in experiment and simulation, respectively. The red arrow indicates the time t = 650 ns when the population of D is maximized. (**C**) The experimental and simulated population of D at t = 650 ns under different disorder steps in two paths. (**E**) The circuit diagram of the interferometer with the $\{R\}$ path blocked at R_1 and R_{10}. (**F**) The experimental and simulated population of D under different disorder steps in two paths with $\{R\}$ path blocked.

labels the qubit in group $\{C_{Q_i}\}$, which couples to Q_i. For the realization of continuous-time QWs (CTQWs), we tuned all qubits to the same interaction frequency for time-independent evolution, and the effective evolution Hamiltonian is given by

$$\hat{H}_{\mathrm{evo}} = \sum\nolimits_{i\in\{Q_i\},j\in\{C_{Q_i}\}} \hbar J_{\mathrm{eff}}^{i,j}\left(\hat{a}_i^\dagger \hat{a}_j + \hat{a}_i \hat{a}_j^\dagger\right) \qquad (2)$$

which forms an interference network. By setting the qubits at the interaction frequency of 5.02 GHz, we determined the effective coupling strengths $J_{\mathrm{eff}}^{i,j}$ by measuring two-qubit swapping oscillations and established $J_{\mathrm{eff}}/2\pi$ = 2.01 ± 0.07 MHz.

We began by exploring CTQWs using one and two walkers by exciting one or two qubits on U00Q0 and U33Q2. Once the initial states were prepared, we tuned all qubits to the interaction frequency and allowed the system to naturally evolve under Eq. 2 for a certain time. We then measured the population $\langle \hat{n}_j \rangle$ of all 62 qubits in their σ_z basis for evolution times ranging from 0 to 600 ns. For each time point, we performed 50,000 single-shot measurements. In Fig. 2A, we present the experimental results for the two-walker QW, with a comparison from numerical simulations (Fig. 2B). In the supplementary materials (*25*), we show the results for the single-walker QWs and the fidelity of the 62-qubit evolution as a function of time. The high-fidelity evolutions indicate the high-accuracy characterization and high-precision control of our system.

For QWs, we also sought to determine the propagation velocity of the walker(s) through the network compared with the Lieb-Robinson (LR) bound (*26*). To achieve this, focusing on the single-walker situation for simplicity, we used the two-site correlation function defined by $C_{ij}(t) = \langle \hat{\sigma}_z^i \hat{\sigma}_z^j \rangle - \langle \hat{\sigma}_z^i \rangle \langle \hat{\sigma}_z^j \rangle$ (Fig. 2C) (*19*). As shown in Fig. 2D, we determined the propagation velocity as 22.2 ± 2.0 site/μs. The maximal group velocity for 2D systems (*27*, *28*) is given by $v_{\max} = 2\sqrt{2}J_{\mathrm{eff}}\left(1 - 16J_{\mathrm{eff}}^2/9U^2\right)$, which equates to $v_{\max}$ = 35.7 site/μs in our system. Now, $v < v_{\max}$ clearly shows that our propagation velocity is limited by the LR bound. The difference is attributed to the short distance and disorders (*25*, *27*, *29*).

The CTQW demonstration established a solid basis for the realization of programmable QWs. Furthermore, our ability to accurately vary the frequency of each qubit enabled us to define propagation paths for the quantum walkers. This is critical for QW-based quantum computing, for which we need to deal with graph problems with different structures. As shown in Fig. 3A, we defined two intersecting paths in our 62-qubit superconducting processor to demonstrate a Mach-Zehnder (MZ) interferometer, where the qubits in the path are tuned to the interaction frequency of 5.02 GHz, whereas those not involved are biased to 4.97 GHz. After exciting the site S, the walker will propagate to $BS1$, where it is split and transmitted along two spatially separated paths (L_1 to L_{10}) and (R_1 to R_{10}). These paths are reconnected at $BS2$, from which the walker arrives at site D. The time evolution of all sites' population is measured from time (t) = 0 to 1000 ns. Figure 3B clearly shows the single walker traversing both the $\{L\}$ and $\{R\}$ paths. At t = 650 ns, we observed a refocusing of the QW, with the population as high as 0.43. We found an excellent agreement compared with the numerical simulations (Fig. 3D).

Our flexibility in adjusting the qubit frequencies provides another freedom that we can exploit associated with the phase on paths,

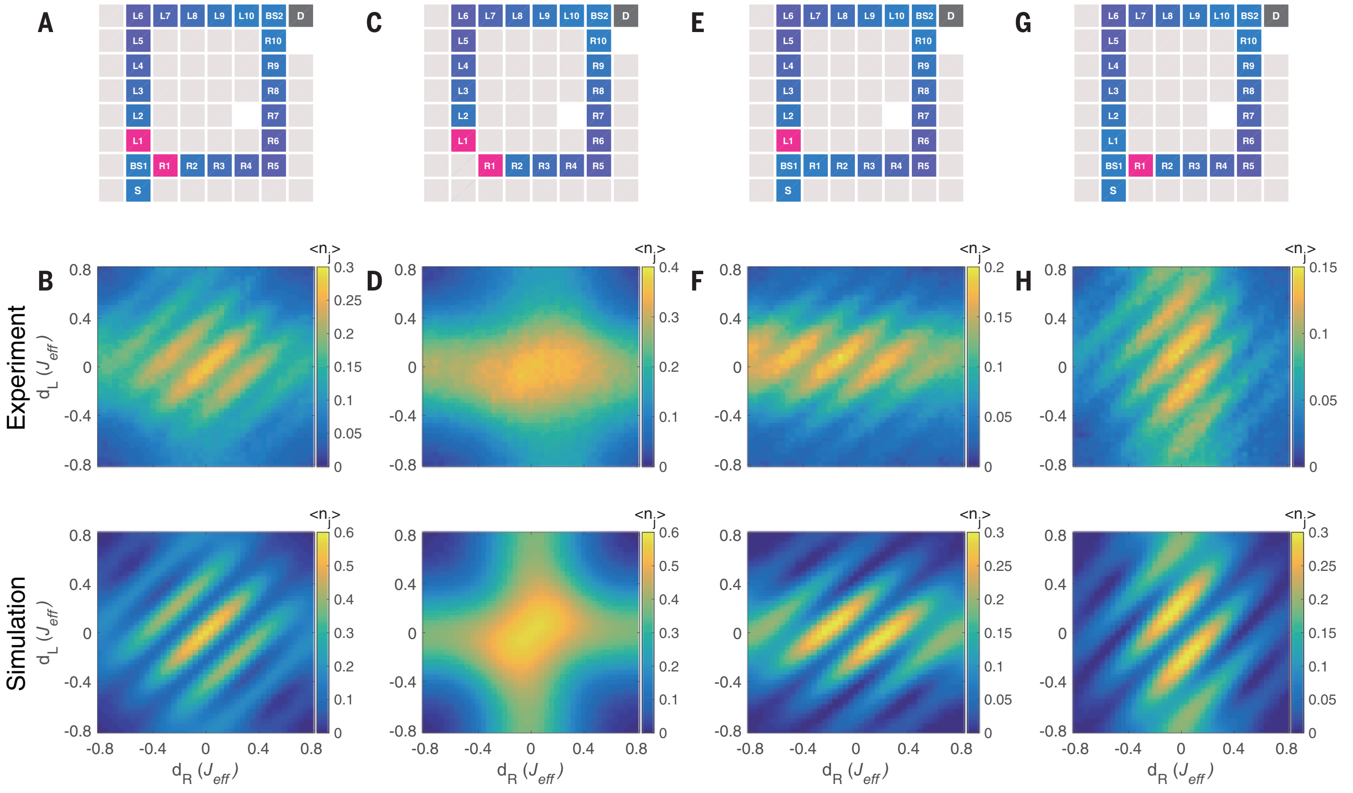


Fig. 4. Two walkers in the MZ interferometer. (**A**, **C**, **E**, and **G**) Circuit diagrams of the programmable paths for the Mach-Zehnder interferometer. The initial state composed of either a single or double walker is prepared by the excitation of the sites marked with pink. (**B**, **D**, **F**, and **H**) The experimental and numerically simulated population of D at t = 550 ns are shown for the various configurations below the respective circuit diagrams.

which is realized with the change of disorders. For the $\{R\}$ path elements, we adjusted the disorder of the sites R_1 to R_5 from d_R to $5d_R$, respectively, whereas for sites R_6 to R_{10}, we did the opposite, changing from $5d_R$ to d_R. Similar disorder changes (scaling as d_L) are made in the $\{L\}$ path. By controlling the disorder sizes, we measured the population on site D at t = 650 ns and observed interference fringes (Fig. 3C). To confirm the origin of these fringes, we blocked the path of $\{R\}$ on R_1 and R_{10} (Fig. 3E) and found no interference fringes (Fig. 3F). Such results show that the disorder not only changes the tunneling amplitude between neighboring sites but also provides the quantum walker a different phase accumulated in propagation that gives rise to the interference fringes. Moreover, for that interference to have occurred, our walker must have maintained coherence as it traversed a superposition of distinct spatially separated paths. The generation of those nonlocal correlations is essential for the development of QW-based universal quantum computation.

The question that next arose was what occurs when we have multiple walkers in our MZ interferometer. We then created two walkers on sites L_1 and R_1 (Fig. 4A) by exciting these respective qubits and then let the system evolve. We measured the population on site D after t = 550 ns and observed the interference fringes in Fig. 4B. This is a similar pattern to what we observed in the single-walker case. To determine the origin of this interference fringes in the two-walker case, we performed a number of control experiments, beginning with the removal of sites $BS1$ and S (Fig. 4C), which stopped both of the walkers from back-propagating to take their alternate path. As shown in Fig. 4D, no interference fringes were observed anymore. This indicates that the pattern comes from the interference between the single-particle forward- and back-propagation. Next, we created a single walker at either site L_1 or R_1 (Fig. 4, E and G), respectively, and let it walk through the interferometer. The results both clearly show interference fringes (Fig. 4, F and H). However, neither one of them nor their sum (*25*) are the same as what we observed in Fig. 4B. This reinforces our observation that the two walkers present in the MZ interferometer must have interacted with each other. Such results agree well with our understanding of transmon qubit physics in the hard-core boson limit (*30*), where $|U/J_{\mathrm{eff}}| \sim 120$.

Our successful demonstration of QWs in two dimensions and the corresponding realization of MZ interferometers clearly shows the potential of these superconducting qubit processors. With the remarkable control of not only the qubits frequencies but also the tunneling amplitude and phase between neighboring sites, these superconducting circuit–based QWs are an elegant approach for the exploration of hard-core boson interference beyond that achievable in photonic systems. Furthermore, multiple-walker realizations will push us into the quantum advantage realm as the excitation number and/or the processor size increases. Last, the demonstration of programmable QWs in superconducting quantum processors is a technological milestone that provides a solid basis for more complex quantum many-body simulations and in the future can be further applied to quantum search algorithms and even universal quantum computing.

REFERENCES AND NOTES

1. Y. Aharonov, L. Davidovich, N. Zagury, *Phys. Rev. A* **48**, 1687–1690 (1993).
2. O. Mülken, A. Blumen, *Phys. Rep.* **502**, 37–87 (2011).
3. A. Aspuru-Guzik, P. Walther, *Nat. Phys.* **8**, 285–291 (2012).
4. N. Lambert *et al.*, *Nat. Phys.* **9**, 10–18 (2013).
5. N. Shenvi, J. Kempe, K. B. Whaley, *Phys. Rev. A* **67**, 052307 (2003).
6. A. M. Childs, J. Goldstone, *Phys. Rev. A* **70**, 022314 (2004).
7. A. M. Childs, *Phys. Rev. Lett.* **102**, 180501 (2009).

8. A. M. Childs, D. Gosset, Z. Webb, *Science* **339**, 791–794 (2013).
9. A. M. Childs *et al.*, in *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing* (ACM, 2003), pp. 59–68.
10. P. P. Rohde, T. C. Ralph, *Phys. Rev. A* **85**, 022332 (2012).
11. H.-S. Zhong *et al.*, *Science* **370**, 1460–1463 (2020).
12. A. Peruzzo *et al.*, *Science* **329**, 1500–1503 (2010).
13. C. Chen *et al.*, *Phys. Rev. Lett.* **121**, 100502 (2018).
14. H. Schmitz *et al.*, *Phys. Rev. Lett.* **103**, 090504 (2009).
15. F. Zähringer *et al.*, *Phys. Rev. Lett.* **104**, 100503 (2010).
16. M. Karski *et al.*, *Science* **325**, 174–177 (2009).
17. C. A. Ryan, M. Laforest, J.-C. Boileau, R. Laflamme, *Phys. Rev. A* **72**, 062317 (2005).
18. V. V. Ramasesh, E. Flurin, M. Rudner, I. Siddiqi, N. Y. Yao, *Phys. Rev. Lett.* **118**, 130501 (2017).
19. Z. Yan *et al.*, *Science* **364**, 753–756 (2019).
20. A. Tulsi, *Phys. Rev. A* **78**, 012310 (2008).
21. M. S. Underwood, D. L. Feder, *Phys. Rev. A* **85**, 052314 (2012).
22. F. Arute *et al.*, *Nature* **574**, 505–510 (2019).
23. B. Foxen *et al.*, *Sci. Tech.* **3**, 014005 (2017).
24. M. Vahidpour *et al.*, arXiv:1708.02226 [physics.app-ph] (2017).
25. Materials and methods are available as supplementary materials.
26. E. H. Lieb, D. W. Robinson, *Commun. Math. Phys.* **28**, 251–257 (1972).
27. M. Cheneau *et al.*, *Nature* **481**, 484–487 (2012).
28. Y. Takasu *et al.*, *Sci. Adv.* **6**, eaba9255 (2020).
29. C. K. Burrell, J. Eisert, T. J. Osborne, *Phys. Rev. A* **80**, 052319 (2009).
30. Y. Lahini *et al.*, *Phys. Rev. A* **86**, 011603 (2012).

ACKNOWLEDGMENTS

The authors thank the University of Science and Technology of China Center for Micro- and Nanoscale Research and Fabrication for supporting the sample fabrication. The authors also thank QuantumCTek Co. for supporting the fabrication and the maintenance of room-temperature electronics. **Funding:** This research was supported by the National Key R&D Program of China (grant 2017YFA0304300), the Chinese Academy of Sciences, Anhui Initiative in Quantum Information Technologies, Technology Committee of Shanghai Municipality, National Science Foundation of China (grants 11905217 and 11774326), Shanghai Municipal Science and Technology Major Project (grant 2019SHZDZX01), Natural Science Foundation of Shanghai (grant 19ZR1462700), and Key-Area Research and Development Program of Guangdong Provice (grant 2020B0303030001). This work was also supported in part by the Japanese MEXT Quantum Leap Flagship Program (MEXT Q-LEAP), grant JPMXS0118069605. **Author contributions:** X.Z. and J.-W.P. conceived the research. M.G., S.W, C.Z., M.-C.C., and X.Z. designed the experiment. S.W., M.G., Q.Z., Y.Z., Y.Y., F.C., C.Y., and X.Z. designed the sample. S.W., C.Z., H.R., H.D., K.Z., S.C., and Y.-H.H. prepared the sample. S.G., H.Q., and H.D. prepared the Josephson parametric amplifiers. Y.W. developed the programming platform for the experiments. M.G., S.W., C.Z., Y.Z., S.L., C.Y., J.Y., D.F., D.W., and H.S. contributed to the building of the ultra low-temperature and low-noise measurement system. J.L., Y.X., F.L., C.G., L.S., N.L., and C.-Z.P. developed the room-temperature electronics. M.G., S.W., C.Z., M.-C.C., H.-L.H., V.M.B., K.N., W.J.M., C.-Y.L., and X.Z. performed the data analysis. All authors contributed to the discussions of the results and the preparation of the manuscripts. X.Z. and J.-W.P. supervised the whole project. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/948/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Table S1
References
Movies S1 to S4

28 January 2021; accepted 20 April 2021
10.1126/science.abg7812

ZEOLITE CATALYSIS

Role of the ionic environment in enhancing the activity of reacting molecules in zeolite pores

Niklas Pfriem[1], Peter H. Hintermeier[1], Sebastian Eckstein[1], Sungmin Kim[2], Qiang Liu[1,3], Hui Shi[1,4], Lara Milakovic[1], Yuanshuai Liu[1,5], Gary L. Haller[1], Eszter Baráth[1], Yue Liu[1]*, Johannes A. Lercher[1,2]*

Tailoring the molecular environment around catalytically active sites allows for the enhancement of catalytic reactivity through a hitherto unexplored pathway. In zeolites, the presence of water creates an ionic environment via the formation of hydrated hydronium ions and the negatively charged framework aluminum tetrahedra. The high density of cation-anion pairs determined by the aluminum concentration of a zeolite induces a high local ionic strength that increases the excess chemical potential of sorbed and uncharged organic reactants. Charged transition states (carbocations for example) are stabilized, which reduces the energy barrier and leads to higher reaction rates. Using the intramolecular dehydration of cyclohexanol on H-MFI zeolites in water, we quantitatively show an enhancement of the reaction rate by the presence of high ionic strength as well as show potential limitations of this strategy.

Zeolites, which are Brønsted- or Lewis-acidic microporous tectosilicates, are widely applied in chemical industry for sorption, separation, and catalysis (*1*–*4*). In the most classic case, the acid character is introduced by substitution of metal cations with a 3+ formal charge (*5*). It has been speculated that the proximity of charge and dipoles in zeolite pores should give rise to strong field effects, but the effects were not systematically quantifiable (*6*, *7*). A long series of investigations has shown that Brønsted acid sites (BASs) have constant acid strength for sorption and catalysis, as long as the concentrations of substituting tetrahedral atoms do not exceed a certain threshold (*8*, *9*). The high intrinsic catalytic activity of zeolites has, therefore, been attributed to the notable stabilization of transition states in the constraints of pores (*10*–*14*).

Recent experiments have shown that this beneficial aspect of transition-state stabilization also holds true when the catalyzed reaction is performed in the presence of water, allowing for the generation of hydrated hydronium ions ($\mathrm{H_3O^+}_{\mathrm{hydr.}}$). The catalytic activity of these hydronium ions is up to two orders of magnitude higher than the respective specific activity of $\mathrm{H_3O^+}_{\mathrm{hydr.}}$ in an aqueous acid solution, as probed by alcohol dehydration (*10*, *11*, *15*). In the environment of the zeolites, the $\mathrm{H_3O^+}_{\mathrm{hydr.}}$ occupy a well-defined location at aluminum tetrahedra (which causes the BAS in the solid) and are located in much closer proximity in zeolite pores than in the liquid phase. This proximity has been shown to lead to an increase in the activity coefficient of organic molecules in the zeolite pores, resembling the situation in the aqueous phase in the presence of high concentrations of cations and anions of dissolved acids, bases, and salts (*15*). The direct consequence of the higher activity coefficient is a decrease in the interaction strength of the sorbed organic molecules, much like the decreasing interaction with the solvent in an aqueous solution of increasing density of cation-anion pairs. This allows for a rigorous translation of the physical chemistry of nonideality in an aqueous phase into the constraints of nanopores.

The higher activity coefficient is caused by an increase in the excess chemical potential compared with an ideal environment, for example, of a zeolite having no BAS. Conceptually, the increase in the excess chemical potential of the reacting molecule decreases the energy difference to the transition state and, in consequence, should lead to a higher rate, even if the transition state is not stabilized. In addition, a polar transition state will be stabilized by the polar environment compared with the uncharged reactant. The confinement and the well-defined close spacing of $\mathrm{H_3O^+}_{\mathrm{hydr.}}$ additionally stabilize the transition state, acting positively to reduce the free-energy barrier in the confines of zeolites (*10*, *11*).

Here, we show that such a scenario can be realized and that the combination of the increase in the thermodynamic activity of reacting molecules and the steric constraints of the zeolite pores leads to strong deviations from the expected sympathetic variation of catalytic activities with active sites of constant acid-base properties. We use the dehydration of cyclohexanol on a series of H-MFI zeolites with a wide range of BAS [$\mathrm{H_3O^+}_{\mathrm{hydr.}}$] [0.05 to

---

[1]Department of Chemistry and Catalysis Research Center, Technical University of Munich, Lichtenbergstrasse 4, 85747 Garching, Germany. [2]Institute for Integrated Catalysis, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA. [3]Paul Scherrer Institute, Forschungsstrasse 111, 5232 Villigen PSI, Switzerland. [4]School of Chemistry and Chemical Engineering, Yangzhou University, Siwangting Road 180, 225009 Yangzhou, Jiangsu, China. [5]Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Songling Road 189, Laoshan District, Qingdao, China.
***Corresponding author. Email: yue.liu@tum.de (Y.L.); johannes.lercher@tum.de (J.A.L.)**

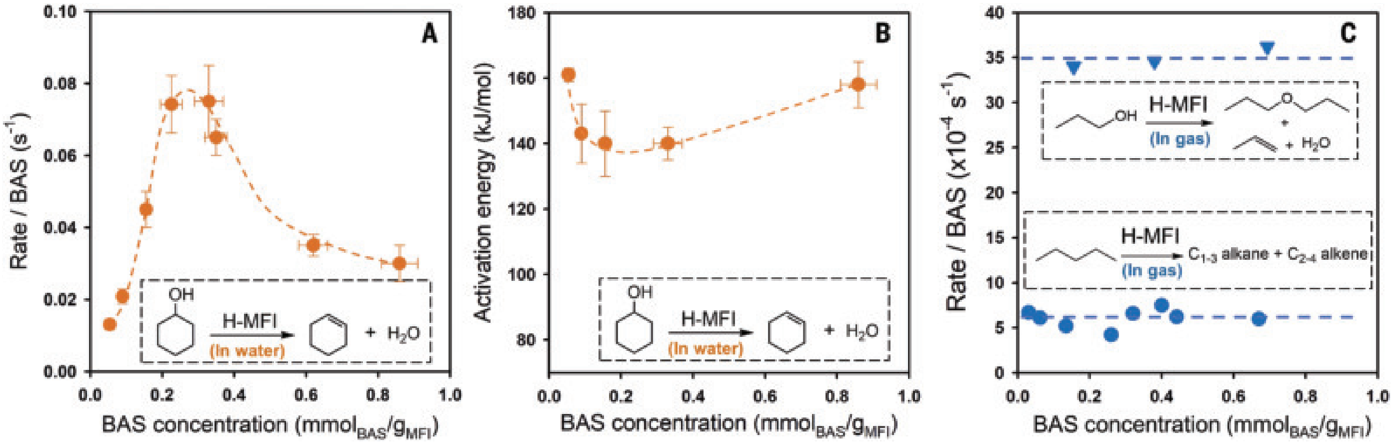


Fig. 1. BAS-normalized reaction rate in H-MFI–catalyzed reactions in water and gas phases. (**A** and **B**) Rate at 423 K (A) and activation energy (B) of cyclohexanol dehydration on H-MFI with different BAS concentrations in water. (**C**) Cracking of *n*-pentane at 763 K (circles) and dehydration of 1-propanol at 433 K (triangles) on H-MFI with different BAS concentrations in the gas phase. Data for *n*-pentane cracking are from (*25*) and (*16*); data for 1-propanol dehydration on H-MFI with 0.69 mmol/g_{MFI} are from (*26*).

0.86 millimoles per gram of MFI (mmol/g_{MFI})] to show that the positive effect of enhancing the excess chemical potential of reacting molecules leads to an optimum density of active sites.

When catalyzed by $H_3O^+_{hydr.}$, cyclohexanol dehydrates to cyclohexene in the aqueous phase on zeolite H-MFI. The reaction rate is independent of the aqueous phase concentration of cyclohexanol above 0.1 mol/liter (zero-order-reaction region; fig. S1). The rate normalized to the concentration of BAS, the turnover frequency (TOF), increased sixfold from a BAS concentration of 0.054 to 0.36 mmol/g_{MFI} and then decreased by 60% to a BAS concentration of 0.86 mmol/g_{MFI} (Fig. 1A). Concurrently, the activation energy decreased from 161 to 140 kJ/mol and increased to 158 kJ/mol afterward (Fig. 1B). By contrast, for gas phase reactions such as *n*-pentane cracking and 1-propanol dehydration, the variation in BAS concentration did not change the TOF, that is, the acid strength of the sites involved can be considered as being constant (Fig. 1C). The identical catalytic activity of BASs in such a series of zeolites has been shown before for cracking (*16*, *17*). The characterization with a base molecule also showed an identical strength of the BAS (*18*). Although it has been reported that different Al locations influence the catalytic activity of zeolites (*19*, *20*), the constant TOFs of *n*-pentane cracking and 1-propanol dehydration on all the tested H-MFIs allow us to exclude the probability that selective Al locations on any specific T sites or in pairs affect reactivity.

Then, the question arises as to why BASs exhibited substantial differences in the presence of water. To address this question, we analyzed the differences of BASs induced by their environments. In the gas phase, BASs are predominately covalent hydroxy groups, located on oxygen bridging between silicon-oxygen and aluminum-oxygen tetrahedra in the zeolite. The hydroxy groups are moderately polar and have negligible volume. In water, the BASs are converted to $H_3O^+_{hydr.}$ bonded ionically to the zeolite framework, with a positive charge and a specific volume.

In contrast to a homogeneous solution in which $H_3O^+_{hydr.}$ are highly dispersed throughout the liquid volume, the $H_3O^+_{hydr.}$ in H-MFI are confined in the limited space of zeolite pores. Consequently, this leads to a very high local concentration that cannot be changed by adding more water. Figure 2A (black line) shows the concentrations of $H_3O^+_{hydr.}$ per H-MFI unit cell varying between 0.3 and 5 in the samples tested, corresponding to a local density of 0.1 to 1.6 mol/liter, using the H-MFI unit cell volume of about 5.2 nm^3 (*21*). If we consider further that the micropore volume of H-MFI is only ~1.3 nm^3 per unit cell (~0.14 cm^3/g_{MFI}; table S1), then the $H_3O^+_{hydr.}$ has a molarity in the range of 0.4 to 6.4 mol/liter in the micropore space. Such high local concentrations of $H_3O^+_{hydr.}$ have two consequences: a very high local ionic strength and a competition between the $H_3O^+_{hydr.}$ and the substrate for the micropore space.

As an ion pair, $H_3O^+_{hydr.}$ with the corresponding anions induces an ionic environment. In an aqueous homogeneous electrolyte solution, this leads to nonideality by which the ions (including $H_3O^+_{hydr.}$) have concentration-dependent activity coefficients (γ). A solute is stabilized ($\gamma < 1$) or destabilized ($\gamma > 1$) by the presence of a specific concentration of cations and anions, reflecting an increased or decreased excess chemical potential of the solute. The ionic strength (*I*) is the most critical variable determining the extent of deviation from an ideal solution. It is defined as the sum of the product of the charge (z_i) squared and the concentration (c_i) of all the ions (Eq. 1). Thus, γ is expressed as a function of *I*, γ(*I*):

$$I = \frac{1}{2} \cdot \sum c_i z_i^2 \qquad (1)$$

Although classical ionic strength is defined for a homogeneous solution, we have shown previously that the concept of nonideality and ionic strength is transferrable to zeolites in water and applicable to quantitatively explain adsorption properties (*15*). By considering the zeolite to be a "quasi solid electrolyte," its local ionic strength in a micropore is also defined by Eq. 1, with z_i being 1 and c_i being volumetric densities of $H_3O^+_{hydr.}$ and the negatively charged framework site (Z^-) in the micropores, i.e., concentration normalized to micropore volume. Figure 2A (blue line) shows the ionic strength of the investigated H-MFI. The high local concentration of $H_3O^+_{hydr.}$ in H-MFI micropores induces high local ionic strength. For example, H-MFI (Si/Al 15) has a BAS concentration of 0.86 ± 0.05 mol/g_{MFI}, corresponding to an ionic strength of 4.9 ± 0.3 mol/liter. The ionic strength increases monotonically with the BAS concentration (Fig. 2A), and the curve bends at high BAS concentration, because of the expansion of the unit cell at high Al concentration that dilutes the volumetric concentration of $H_3O^+_{hydr.}$.

The nonideality also affects the catalytic reaction rate. In general, the TOF under a certain ionic strength TOF(*I*) differs from that under ideal conditions $TOF_{(ideal)}$—that is, TOF at zero ionic strength—by factors of the activity coefficient of ground state $\gamma_{GS}(I)$ and transition state $\gamma_{TS}(I)$ (Eq. 2A; detailed derivations

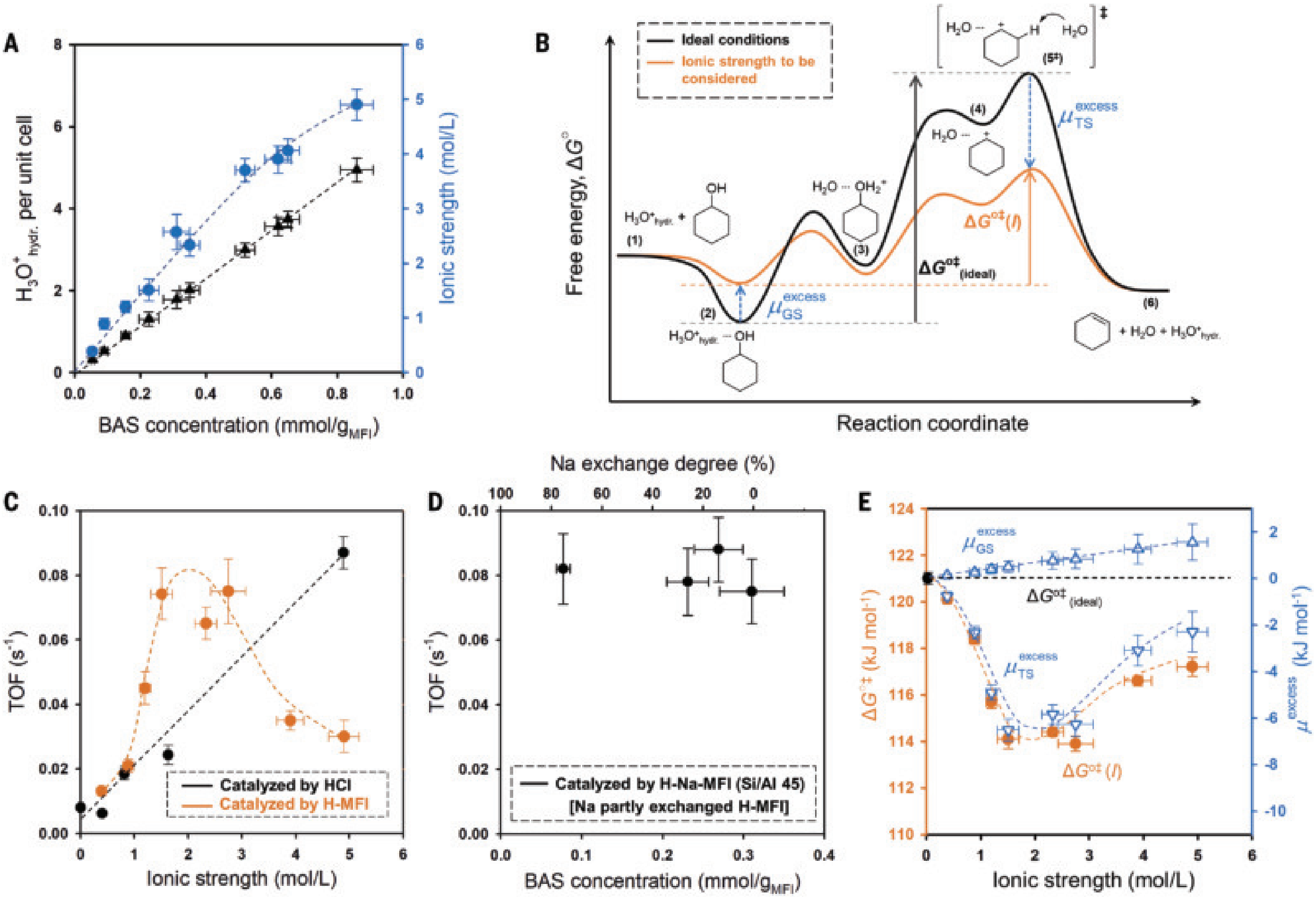


Fig. 2. Impact of local $H_3O^+_{hydr.}$ concentration and ionic strength on the dehydration of cyclohexanol catalyzed by $H_3O^+_{hydr.}$. (**A**) Unit cell–normalized concentrations of $H_3O^+_{hydr.}$ (triangles) and ionic strength (circles) as a function of BAS ($H_3O^+_{hydr.}$) concentration. (**B**) Elementary steps and their energies in dehydration of cyclohexanol on $H_3O^+_{hydr.}$ in H-MFI zeolite in water under ideal and nonideal conditions: (1) reactant cyclohexanol and $H_3O^+_{hydr.}$, (2) cyclohexanol associated with $H_3O^+_{hydr.}$, (3) protonated cyclohexanol, (4) cyclohexyl carbenium ion, (5‡) transition state of deprotonation of cyclohexyl carbenium ion by water, and (6) product cyclohexene, water, and $H_3O^+_{hydr.}$. The "H_2O···" represents the interactions with solvent water. (**C**) TOF as a function of ionic strength under the catalysis of HCl (black circles) at 453 K and H-MFI (orange circles) at 423 K. The ionic strength in HCl solution is varied by changing the concentration of LiCl electrolyte. (**D**) TOF of Na partly exchanged H-MFI (Si/Al 45). (**E**) Reaction free-energy barriers and excess chemical potential of the ground state (GS) and transition state (TS) under the ideal condition and under an ionic strength. The calculations of free-energy barriers and excess chemical potential are provided in supplementary text S1.

are in supplementary text S1 and S2). The $TOF_{(ideal)}$ is determined by $\Delta G^{o\ddagger}_{(ideal)}$, which is the free-energy barrier under the ideal condition calculated by applying the transition-state formula (Eq. 2B):

$$TOF(I) = TOF_{(ideal)} \cdot \frac{\gamma_{GS}(I)}{\gamma_{TS}(I)} \qquad (2A)$$

$$TOF_{(ideal)} = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{o\ddagger}_{(ideal)}}{RT}\right) \qquad (2B)$$

where k_B, T, h, and R denote the Boltzmann constant, temperature, Plank constant, and ideal gas constant, respectively. The activity coefficient determines the excess chemical potential (μ^{excess}) according to $\mu^{excess} = RT\ln\gamma$. For the dehydration of cyclohexanol by $H_3O^+_{hydr.}$, it proceeds stepwise via association with $H_3O^+_{hydr.}$, protonation of the OH group, C–O cleavage to form the cyclohexyl carbenium ion, and deprotonation of the cyclohexyl carbenium ion (Fig. 2B). This stepwise mechanism is shown to dominate from low $H_3O^+_{hydr.}$ concentration (<0.01 mol/liter) in homogeneous acids to high local $H_3O^+_{hydr.}$ concentration (~2.7 mol/liter) in zeolites (*10*, *11*). The free-energy barrier is the energy difference between the transition state, the deprotonation of cyclohexyl carbenium ion by water ($C_6H_{11}^+\cdots H_2O$), and the ground state of adsorbed cyclohexanol associated with $H_3O^+_{hydr.}$ ($H_3O^+_{hydr.}\cdots C_6H_{11}OH$). Cyclohexanol is a neutral molecule in the ground state, whereas it is a positively charged cyclohexyl carbenium ion ($C_6H_{11}^+$) in the transition state. A neutral molecule is normally destabilized by the specific ionic strength ($\mu_{GS}^{excess} > 0$), given by Eq. 3A, which shows a proportional increase with I. The term K_s denotes the Setschenow constant. By contrast, a cation or an anion is normally stabilized by the presence of an ionic environment ($\mu_{TS}^{excess} < 0$), given by the extended Debye-Hückel equation (e.g., Truesdell-Jones equation; Eq. 3B), where a is the ion diameter and A, B, and b are constants:

$$\mu_{GS}^{excess} = 2.303 \cdot RTK_s I \qquad (3A)$$

$$\mu_{TS}^{excess} = 2.303RT \cdot \left(-\frac{A\sqrt{I}}{1+aB\sqrt{I}} + bI\right) \qquad (3B)$$

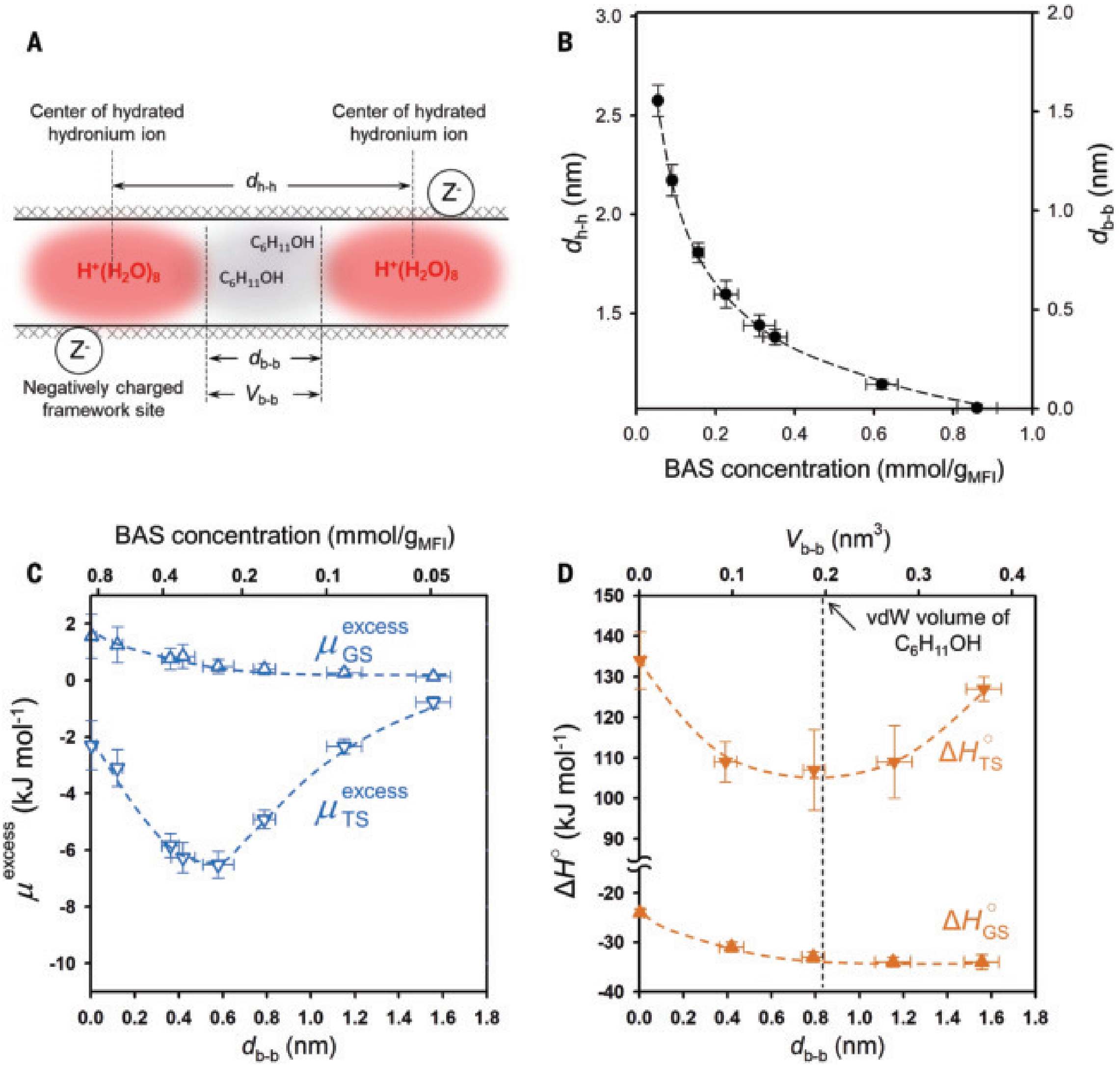


Fig. 3. Impact of the distance between $H_3O^+_{hydr.}$ in the H-MFI micropore on the energy of the ground state and the transition state of cyclohexanol dehydration catalyzed by $H_3O^+_{hydr.}$. (**A**) Schematic illustration of $H_3O^+_{hydr.}$ and cyclohexanol in H-MFI micropore channels and the mean distance d_{h-h} between two neighboring $H_3O^+_{hydr.}$ and the mean distance d_{b-b} and volume V_{b-b} between the boundaries of neighboring $H_3O^+_{hydr.}$. (**B**) The d_{h-h} and d_{b-b} as a function of BAS concentration. (**C** and **D**) Excess chemical potential (C) and enthalpy of the ground and transition states (D) as a function of d_{b-b} and V_{b-b}. The d_{h-h} is estimated by the cubic root of the average zeolite volume normalized to the number of $H_3O^+_{hydr.}$; the d_{b-b} is calculated by subtracting the length of $H_3O^+_{hydr.}$ from d_{h-h}; the V_{b-b} is calculated by a cylinder model with length of d_{b-b} and the diameter of the H-MFI micropore channel.

As illustrated in Fig. 2B, the positive μ^{excess}_{GS} and negative μ^{excess}_{TS} in an ionic environment lead to a lower energy barrier compared with that under ideal conditions [$\Delta G^{o\ddagger}(I) < \Delta G^{o\ddagger}_{(ideal)}$].

The rate of cyclohexanol dehydration by HCl in homogeneous aqueous solution showed a positive correlation with the ionic strength in concentrated LiCl solution (black line in Fig. 2C). Its TOF at 453 K is less than 0.01 s^{-1} at very low ionic strength, whereas it increased to more than 0.08 s^{-1} at a high ionic strength of ~5 mol/liter. In analogy to HCl, the TOF of cyclohexanol dehydration on H-MFI is replotted as a function of ionic strength (orange line in Fig. 2C). It increases with increasing ionic strength up to ~2 mol/liter; however, in contrast to HCl, it drops at higher ionic strength.

To further explore the impact of ionic strength, a series of Na$^+$ partly exchanged H-MFI (Si/Al 45) were prepared. Partly exchanging $H_3O^+_{hydr.}$ by Na$^+$ decreases the BAS concentration of H-MFI while maintaining the same ionic strength. As shown in Fig. 2D, with Na exchange degrees up to 75%, the TOFs of H-Na-MFI remained unchanged, indicating that the variation of the concentration of $H_3O^+_{hydr.}$ while maintaining a constant ionic strength does not change the reaction rate. The comparison of H-Na-MFI (Si/Al 45) with a Na$^+$ exchange degree of 75% with H-MFI (Si/Al 200) exemplifies the impact of the ionic strength. Both samples have close BAS concentrations (77 versus 90 μmol/g_{MFI}); however, the former has a higher ionic strength than the latter (2.6 versus 0.9 mol/liter), and, indeed, the TOF is fourfold higher on H-Na-MFI (Si/Al 45) than on H-MFI (Si/Al 200), that is, 0.082 versus 0.021 s^{-1}. Thus, these results demonstrate that it is not the location of hydronium ions or the hydronium ion concentrations but the ionic strength that is critical for the specific catalyzed rate.

The free-energy barrier and excess chemical potential in the reactions are shown in Fig. 2E. The decrease of μ^{excess}_{TS} is more important than the increase of μ^{excess}_{GS}; thus, the lower free-energy barrier is mostly caused by the stabilization of the transition state. The inverse-volcano trend of μ_{TS}^{excess} with ionic strength seems at first sight consistent with Eq. 3B whereby the negative first term dominates at low ionic strength and the positive second term dominates at high ionic strength. However, the monotonic increase of TOFs with ionic strength at all concentrations under HCl catalysis excludes this possibility.

Next, we explore the reason for the rate drop with $H_3O^+_{hydr.}$ at higher concentrations. In contrast to catalysis in aqueous HCl, in H-MFI, both cyclohexanol and $H_3O^+_{hydr.}$ reside in the zeolitic micropore channels, and their relative locations are depicted in Fig. 3A. The average distance between two $H_3O^+_{hydr.}$ neighbors ($d_{h\text{-}h}$) represents the distance between $H_3O^+_{hydr.}$-Z^- pairs. The distance between their boundaries is $d_{b\text{-}b}$, with a volume of $V_{b\text{-}b}$. In the space between them resides the sorbed cyclohexanol. At the molecular level, the change of electrolyte concentration is equivalent to the change of $d_{h\text{-}h}$ and $d_{b\text{-}b}$. Thus, the increase of BAS concentration in H-MFI leads to a shorter distance between $H_3O^+_{hydr.}$, inducing a decrease in $d_{h\text{-}h}$, $d_{b\text{-}b}$, and $V_{b\text{-}b}$. Figure 3B shows that increasing the BAS concentration from 0.05 to 0.86 mmol/g_{MFI} reduced the $d_{h\text{-}h}$ from 2.6 to 1.0 nm and the $d_{b\text{-}b}$ from 1.6 to almost 0.0 nm. It should be noted here that the composition of $H_3O^+_{hydr.}$ [$H^+(H_2O)_8$] was determined to be invariant with temperature (supplementary text S3).

The individual energy levels of reaction ground and transition states in terms of enthalpy (ΔH°_{GS}, ΔH°_{TS}) and excess chemical potential (μ_{GS}^{excess}, μ_{TS}^{excess}) on all studied H-MFI are depicted as a function of $d_{b\text{-}b}$ (Fig. 3, C and D). The enthalpy of ground state ΔH°_{GS} of sorbed cyclohexanol is stable at a distance $d_{b\text{-}b}$ larger than 0.8 nm, whereas it increases sharply with the decrease of $d_{b\text{-}b}$, suggesting a repulsion between $H_3O^+_{hydr.}$ and sorbed cyclohexanol. The same trend is observed for μ_{GS}^{excess}. By contrast, the transition state has both ΔH°_{TS} and μ_{TS}^{excess} in reverse-volcano curves with the minimum at ~0.8 and ~0.6 nm of $d_{b\text{-}b}$. The $d_{b\text{-}b}$ of 0.8 nm is a critical distance at which $V_{b\text{-}b}$ is 0.2 nm^3, the same volume of one cyclohexanol molecule in liquid phase. At this distance, the volume between two neighboring $H_3O^+_{hydr.}$ is equal to the van der Waal (vdW) volume of cyclohexanol. At lower $d_{b\text{-}b}$, the $V_{b\text{-}b}$ becomes smaller than the vdW volume of cyclohexanol, causing a strong vdW repulsion that increases the energy of the transition state and decreases the TOF. In an open aqueous solution, the work to separate charges (hydronium ions and the anions of the zeolite) will be replaced only by a volume increase without constraining the sorbed alcohol.

Thus, we show how the bulkiness and the charges of $H_3O^+_{hydr.}$ constrained in zeolite micropores combine to influence the catalytic activity. For H-MFI, the $H_3O^+_{hydr.}$ is a cluster with a composition of $H^+(H_2O)_8$ and a specific volume, for example, ~0.24 nm^3 at room temperature. It is anchored to the exchange sites of the zeolite framework by Coulombic forces and competes with substrate (e.g., cyclohexanol) to occupy the micropore volume. Because the stabilization of water in the $H^+(H_2O)_8$ ion is larger than the stabilization of the sorbed organic molecule, that is, cyclohexanol, the latter adsorbs only in the volume between neighboring $H^+(H_2O)_8$ in the micropore channel. This adds spatial constraints to the substrate additional to that of the micropore framework. At high $H_3O^+_{hydr.}$ concentrations, the volume between neighboring $H^+(H_2O)_8$ becomes smaller than the vdW volume of substrate. This leads to strong vdW repulsion between substrate and $H^+(H_2O)_8$ and consequently results in a decrease in the reaction rate. Such vdW repulsion might not apply for the case of smaller alcohols, for example, ethanol, because they were shown to be capable of replacing water molecules in the $H_3O^+_{hydr.}$ cluster, forming $(C_2H_5OH)(H_3O^+)(H_2O)_n$ cluster in zeolite channels (*22*).

Thus, the charge of $H_3O^+_{hydr.}$ creates an ionic environment that can be expressed as the ionic strength in H-MFI micropores. Such an environment increases the standard free energy and excess chemical potential of the substrate ground state, that is, it destabilizes the uncharged reacting ground state of cyclohexanol. The ionic environment stabilizes the positively charged transition state. Both effects together lead to a decrease of the activation free energy and enhance the reaction rate. Such ionic environments in zeolites will also exist with other solvents. For example, methanol forms protonated clusters $H^+(CH_3OH)_n$ in small-pore zeolites (*23*), which would additionally allow for the stabilization of the charged transition state via the modulating solvent permittivity (*24*).

The combination of the enhancement by the charged environment and the limitations by the additional spatial constraints from $H_3O^+_{hydr.}$ leads to a maximum in the catalytic activity of $H_3O^+_{hydr.}$ with varying $H_3O^+_{hydr.}$ concentrations. For H-MFI, the highest rates are observed when the volume between two neighboring $H_3O^+_{hydr.}$ equals to the vdW volume of substrate, that is, the concentrations of substrate and $H_3O^+_{hydr.}$ in the micropores are almost equal.

The final question is, however, how general these conclusions about reactivity in constrained environments are. To probe this, we explored a wide series of H-BEA zeolites (table S1) and report here the first results of this study. The dependence of the rate of alcohol dehydration on the concentration of hydronium ions exhibits an analogous volcano-like dependence (Fig. 4A). In particular, the variation of activation enthalpy and entropy with ionic strength shows similar profile shapes to that of H-MFI (Fig. 4, B and C), indicating that the ionic strength influences catalytic

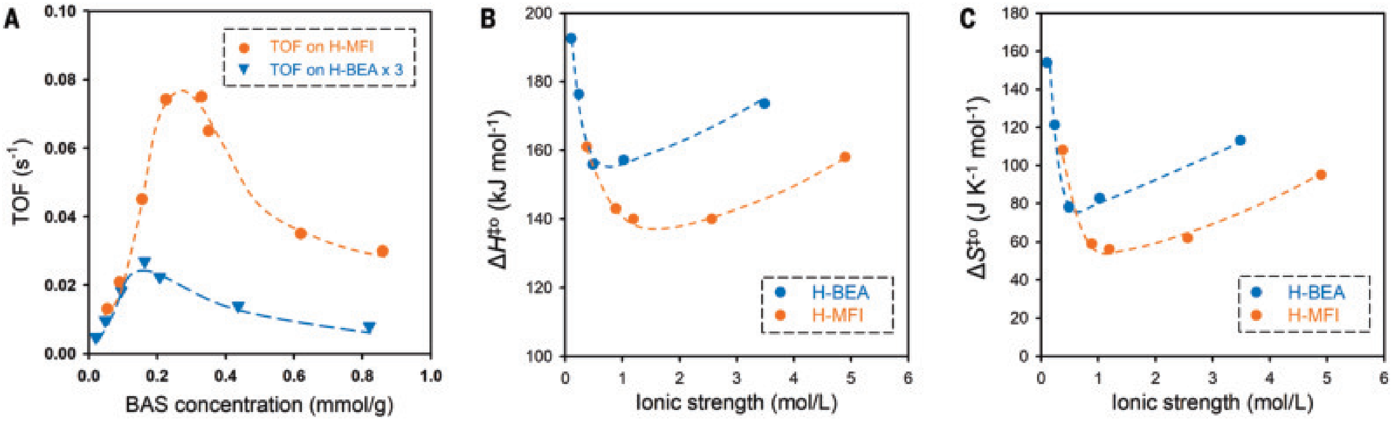


Fig. 4. Comparison of the dehydration reactions of cyclohexanol catalyzed by $H_3O^+_{hydr.}$ in H-MFI and H-BEA. (**A**) Impact of BAS concentration on the TOF at 423 K. (**B** and **C**) Activation enthalpy ($\Delta H^{\ddagger o}$) (B) and activation entropy ($\Delta S^{\ddagger o}$) (C) as a function of ionic strength.

activity through the same mechanism. Note the curves shifting up and toward lower ionic strength on H-BEA compared with H-MFI, which are attributed to the larger pore size of H-BEA that induces less vdW stabilization of transition state. It is to be expected that the specific enhancement and the position of the maximum will subtly depend on the size of the reacting molecule and the difference in polarity between the reacting substrate and the transition state. Overall, the results demonstrate unequivocally that the quantitative interpretation of the catalytic activity will allow for the prediction of the most suitable microporous catalyst for reactions in the presence of active sites that are associated with ions.

REFERENCES AND NOTES

1. R. Gounder, E. Iglesia, *J. Am. Chem. Soc.* **131**, 1958–1971 (2009).
2. P. Tian, Y. Wei, M. Ye, Z. Liu, *ACS Catal.* **5**, 1922–1938 (2015).
3. W. Vermeiren, J. P. Gilson, *Top. Catal.* **52**, 1131–1161 (2009).
4. E. T. C. Vogt, B. M. Weckhuysen, *Chem. Soc. Rev.* **44**, 7342–7370 (2015).
5. G. L. Woolery, G. H. Kuehl, H. C. Timken, A. W. Chester, J. C. Vartuli, *Zeolites* **19**, 288–296 (1997).
6. D. Barthomeuf, *J. Phys. Chem.* **88**, 42–45 (1984).
7. C. Mirodatos, D. Barthomeuf, *J. Catal.* **114**, 121–135 (1988).
8. H. Stach *et al.*, *J. Phys. Chem.* **96**, 8480–8485 (1992).
9. A. Auroux, *Top. Catal.* **4**, 71–89 (1997).
10. H. Shi, S. Eckstein, A. Vjunov, D. M. Camaioni, J. A. Lercher, *Nat. Commun.* **8**, 15442 (2017).
11. Y. Liu *et al.*, *Nat. Commun.* **8**, 14113 (2017).
12. S. Wang, E. Iglesia, *J. Catal.* **352**, 415–435 (2017).
13. A. J. Jones, S. I. Zones, E. Iglesia, *J. Phys. Chem. C* **118**, 17787–17800 (2014).
14. V. J. Margarit *et al.*, *ACS Catal.* **9**, 5935–5946 (2019).
15. S. Eckstein *et al.*, *Angew. Chem. Int. Ed.* **58**, 3450–3455 (2019).
16. S. Schallmoser *et al.*, *J. Catal.* **316**, 93–102 (2014).
17. W. O. Haag, R. M. Lago, P. B. Weisz, *Nature* **309**, 589–591 (1984).
18. D. J. Parrillo, C. Lee, R. J. Gorte, *Appl. Catal. A* **110**, 67–74 (1994).
19. A. Janda, A. T. Bell, *J. Am. Chem. Soc.* **135**, 19193–19207 (2013).
20. B. C. Knott *et al.*, *ACS Catal.* **8**, 770–784 (2018).
21. D. H. Olson, G. T. Kokotailo, S. L. Lawton, W. M. Meier, *J. Phys. Chem.* **85**, 2238–2243 (1981).
22. J. S. Bates, B. C. Bukowski, J. Greeley, R. Gounder, *Chem. Sci.* **11**, 7102–7122 (2020).
23. J. R. Di Iorio *et al.*, *J. Catal.* **380**, 161–177 (2019).
24. D. Y. Murzin, *Catal. Sci. Technol.* **6**, 5700–5713 (2016).
25. W. Luo *et al.*, *Green Chem.* **21**, 3744–3768 (2019).
26. Y. Zhi *et al.*, *J. Am. Chem. Soc.* **137**, 15781–15794 (2015).

ACKNOWLEDGMENTS

We thank O. Y. Gutiérrez and D. M. Camaioni for fruitful discussions and S. Prodinger and M. A. Derewinski for providing the H-BEA zeolite. **Funding:** This work (J.A.L. and S.K.) was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (BES), Division of Chemical Sciences, Geosciences and Biosciences (Transdisciplinary Approaches to Realize Novel Catalytic Pathways to Energy Carriers, FWP 47319). G.L.H. was supported by the DOE Office of Science and BES, Division of Chemical Sciences, Geosciences and Biosciences, under grant no. DE-FG02-01ER15183, to Yale University, Department of Chemical and Environmental Engineering. **Author contributions:** Yue L. and J.A.L. conceived the research; N.P. performed the catalytic reaction by HCl and determined the activation energy on H-MFI; P.H.H., H.S., L.M., Yua.L., and E.B. measured the physicochemical property and catalytic activity of H-MFI, wherein P.H.H. first reported the volcano-like activity of H-MFI; S.E. and Q.L. measured the adsorption property of H-MFI; S.K. measured the reactions on H-BEA; and G.L.H. helped with data analysis. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available within the main text and the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/952/suppl/DC1
Materials and Methods
Supplementary Text S1 to S3
Figs. S1 to S5
Table S1
References (*27–29*)

2 March 2021; accepted 23 April 2021
Published online 6 May 2021
10.1126/science.abh3418

EXOSKELETONS

Removing energy with an exoskeleton reduces the metabolic cost of walking

Michael Shepertycky[1]*, Sarah Burton[2], Andrew Dickson[2], Yan-Fei Liu[2], Qingguo Li[1]*

Evolutionary pressures have led humans to walk in a highly efficient manner that conserves energy, making it difficult for exoskeletons to reduce the metabolic cost of walking. Despite the challenge, some exoskeletons have managed to lessen the metabolic expenditure of walking, either by adding or storing and returning energy. We show that the use of an exoskeleton that strategically removes kinetic energy during the swing period of the gait cycle reduces the metabolic cost of walking by 2.5 ± 0.8% for healthy male users while converting the removed energy into 0.25 ± 0.02 watts of electrical power. By comparing two loading profiles, we demonstrate that the timing and magnitude of energy removal are vital for successful metabolic cost reduction.

Humans are exceptional walkers, with our gait continuously adapting over evolutionary (*1*) and developmental time scales (*2*). In the absence of frictional losses, walking on level ground at a constant speed theoretically requires no energy input (*3*), as passive mechanisms permit the exchange of kinetic and potential energy (*4*, *5*). In reality, however, walking is metabolically expensive. Walking requires more energy than any other activity of daily living (*6*, *7*). During walking, humans expend metabolic energy, in part, to restore dissipated kinetic energy (*8*).

Exoskeletons have been developed to assist the lower-limb muscles in restoring dissipated kinetic energy to improve walking efficiency. Active (powered) exoskeletons use actuators, such as electric motors, to perform a portion of the positive joint work that is naturally done by concentric muscle contractions (*9*, *10*). These devices essentially add energy to the human-device system to supplement the metabolic energy required for walking (fig. S1). Passive (unpowered) exoskeletons use elastic mechanisms to aid users' muscle-tendon units in storing and transferring energy from one gait phase to another. By assisting energy recycling, these devices decrease the force that muscle-tendon units must hold during isometric contractions (*11*).

Although existing devices assist users by adding or recycling energy, is it possible to improve walking efficiency by purely removing energy? This would require a different design approach: Instead of assisting muscles in concentric or isometric contractions, devices would need to assist muscles in producing force during eccentric contractions. The strategy of removing energy is similar to those employed by some biomechanical energy harvesters optimized to produce electrical power (*12*–*14*). Unfortunately, these devices have not been able to provide their users with a metabolic benefit over carrying their own weight, let alone an advantage over normal walking [(*12*–*14*); also see supplementary materials]. Because the metabolic cost of walking can be explained primarily by the cost of actively generating muscle force (*15*), we hypothesized that we could optimize these devices for metabolic cost reduction by designing them with the aim of decreasing muscle activation instead of producing electrical power.

Assisting muscles in producing force during eccentric contractions is challenging. Substantial negative joint work during walking results from the passive deformation of soft tissues (*8*), such as ligaments. These tissues do not require metabolic energy and so cannot be metabolically assisted. Furthermore, some soft tissues store and return energy that would otherwise be dissipated (*5*, *8*); thus, disrupting these energy return mechanisms could lead to a substantial energetic penalty because the body must compensate through concentric contractions. Because eccentric muscle contractions are already metabolically efficient compared

---

[1] Department of Mechanical and Materials Engineering, Queen's University, Kingston, ON K7L 3N6, Canada. [2] Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada.
***Corresponding author. Email: 6mys@queensu.ca (M.S.); ql3@queensu.ca (Q.L.)**

with concentric contractions (*16*), assisting muscles in eccentric contractions should yield only a small energetic advantage. Therefore, assisting muscles in eccentric contractions may require the exoskeleton to apply a load more accurately and precisely, regarding both magnitude and timing, than is required for concentric contraction and powered exoskeletons (*9*, *10*, *17*).

We built a lightweight, backpack-mounted exoskeleton (1.059 kg; Fig. 1) to assist the hamstrings in producing force by applying a knee flexion moment during the terminal swing phase of the gait cycle. This device consists of cables that are attached by harnesses to the user's lower legs and extend upward to connect to input pulleys. As the user's knee extends forward during the swing period, the swinging leg's input cable is unspooled and drives the input pulley. These pulleys work in combination with a gear train and roller clutches to convert the cables' linear movements into rotational motion while simultaneously amplifying and integrating the cables' motions before engaging a single generator. During stance, the clutches decouple the input pulleys from the gear train, allowing the cables to be gathered back onto the input pulleys by returning springs. The clutches also enable the device to apply mechanical load only during the swing period, leaving the lower limbs unencumbered by the exoskeleton during the stance period. The back electromotive force of the generator applies the mechanical load to the input cables and, in turn, to the user (further described in the supplementary materials, movies S1 and S2, and data S3).

Biomechanical energy harvesters traditionally apply a mechanical loading profile onto their users that is proportional to the motion input into the device (*12*–*14*). This technique is simple to implement because it requires no active control, but it may sacrifice user assistance. In this study, we implemented an active control system that enabled our device to apply a controllable mechanical load by regulating the generator's back electromotive force. Exoskeletons have successfully reduced the metabolic energy requirement of walking by using assistive loading profiles that mimic either natural joint moment or power profiles (*18*–*21*). However, these profiles do not account for the contribution of passive elements that cannot be assisted. To specifically target the hamstrings and avoid interfering with natural energy return mechanisms, we implemented a profile that resembles the muscles' contribution to the net knee moment during the terminal swing phase. We derived this muscle-centric profile by subtracting the estimated contribution of passive elements, based on results of Whittington *et al.* (*22*), from the net knee moment profile. We then scaled the magnitude of the resultant profile to the user's body mass. The control system and loading profiles are described in the supplementary materials, figs. S2 and S3, and data S4.

We conducted treadmill-walking experiments with 10 healthy male participants to determine the metabolic and biomechanical effects of the muscle-centric profile compared with both a traditional harvesting profile and normal walking. Minimal changes in both knee joint angles and net joint moments indicate that neither loading profile hindered normal knee function (Fig. 2, A to C), which can cause deviations from normal walking such as a stiff knee gait (*14*). The lack of effect on normal knee function was further confirmed by a lack of changes in spatial-temporal parameters such as stride length and ground contact time. The average angles, moments, and powers for the ankle, knee, and hip joints are presented in fig. S4 and data S5 and S6.

One difference between the two loading profiles (Fig. 2E) was that the moment applied about the user's knee by the traditional profile peaked during mid-swing (80.1 ± 0.5% gait cycle, mean ± SEM), whereas that applied by the muscle-centric profile peaked during terminal swing (90.7 ± 0.2% gait cycle). Both loading profiles significantly reduced the estimated average muscle contribution to the negative net knee moment during the swing period [n = 10 participants; repeated-measures analysis of variance (ANOVA), $F_{2,18} = 17.806$, $P < 0.001$, $\eta_p^2 = 0.664$], with the traditional profile reducing it by 11.0 ± 1.8% [least significant difference (LSD) post hoc analysis: $P < 0.001$] and the muscle-centric profile reducing it by 12.0 ± 2.0% ($P = 0.001$; Fig. 2D).

Although both profiles removed a similar amount of mechanical power (traditional: 2.01 ± 0.12 W, muscle-centric: 1.98 ± 0.09 W; Fig. 2F) and applied a similar percentage of the muscles' angular impulses (traditional: 17.4 ± 1.3%, muscle-centric: 17.4 ± 1.5%; Fig. 2G), the muscle-centric profile reduced the energy requirement of walking by 2.5 ± 0.8% [n = 10; one-way repeated-measures ANOVA; $F_{2,18} = 12.714$, $P < 0.001$, $\eta_p^2 = 0.586$; LSD post hoc analysis: mean diff. (MD) = -0.082 ± 0.024 W kg^{-1}, $P = 0.008$, 95% CI (-0.137, -0.027)], whereas the traditional profile increased the energy requirement by 3.1 ± 1.1% [MD = 0.106 ± 0.039 W kg^{-1}, $P = 0.024$, 95% CI (0.018, 0.195)].

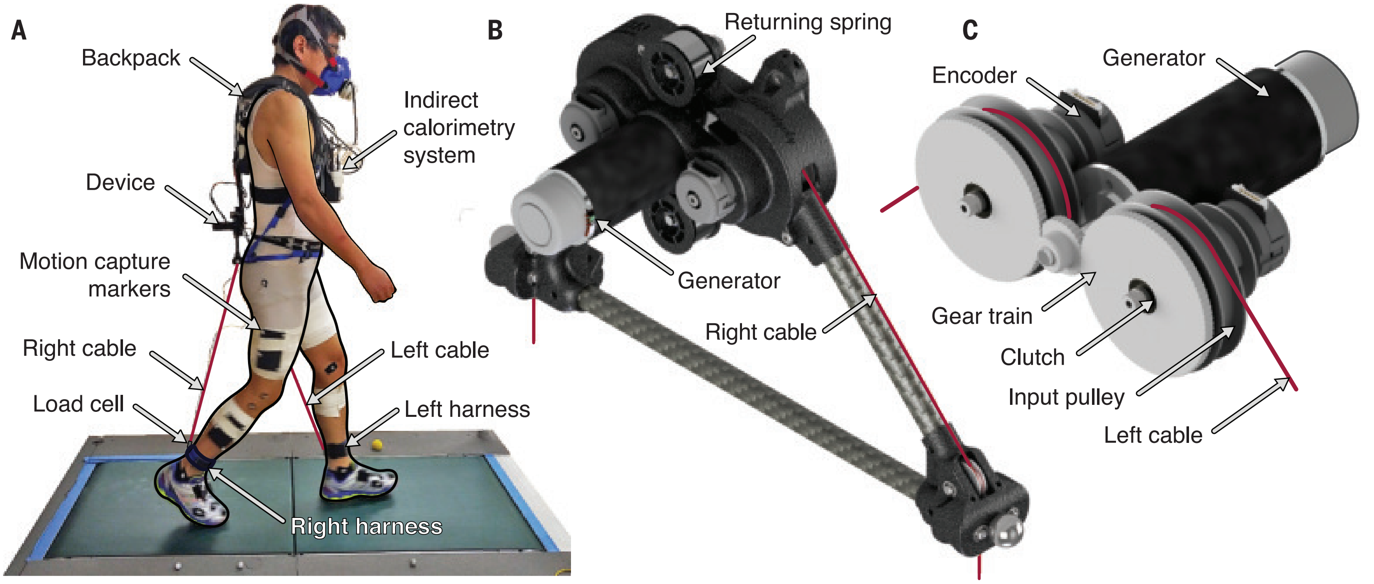


Fig. 1. Energy-removing exoskeleton. (**A**) A participant walking on the split-belt force-sensing treadmill while wearing the exoskeleton and measurement equipment. (**B**) Outer view of the device. (**C**) View of the device's internal components.

The difference between profiles was 0.19 ± 0.05 W kg^{-1} [MD = 0.188 ± 0.046 W kg^{-1}, *P* = 0.003, 95% CI (0.085, 0.291); Fig. 3 and table S1].

Because active skeletal muscles are the primary consumers of metabolic energy during walking (*23*), the decrease in metabolic cost observed with the muscle-centric profile was accompanied by an 11.2 ± 3.8% reduction in average hamstring activation from normal walking ($n = 10$; repeated-measures ANOVA, $F_{2,18} = 4.506$, $P = 0.026$, $\eta_p^2 = 0.334$; LSD post

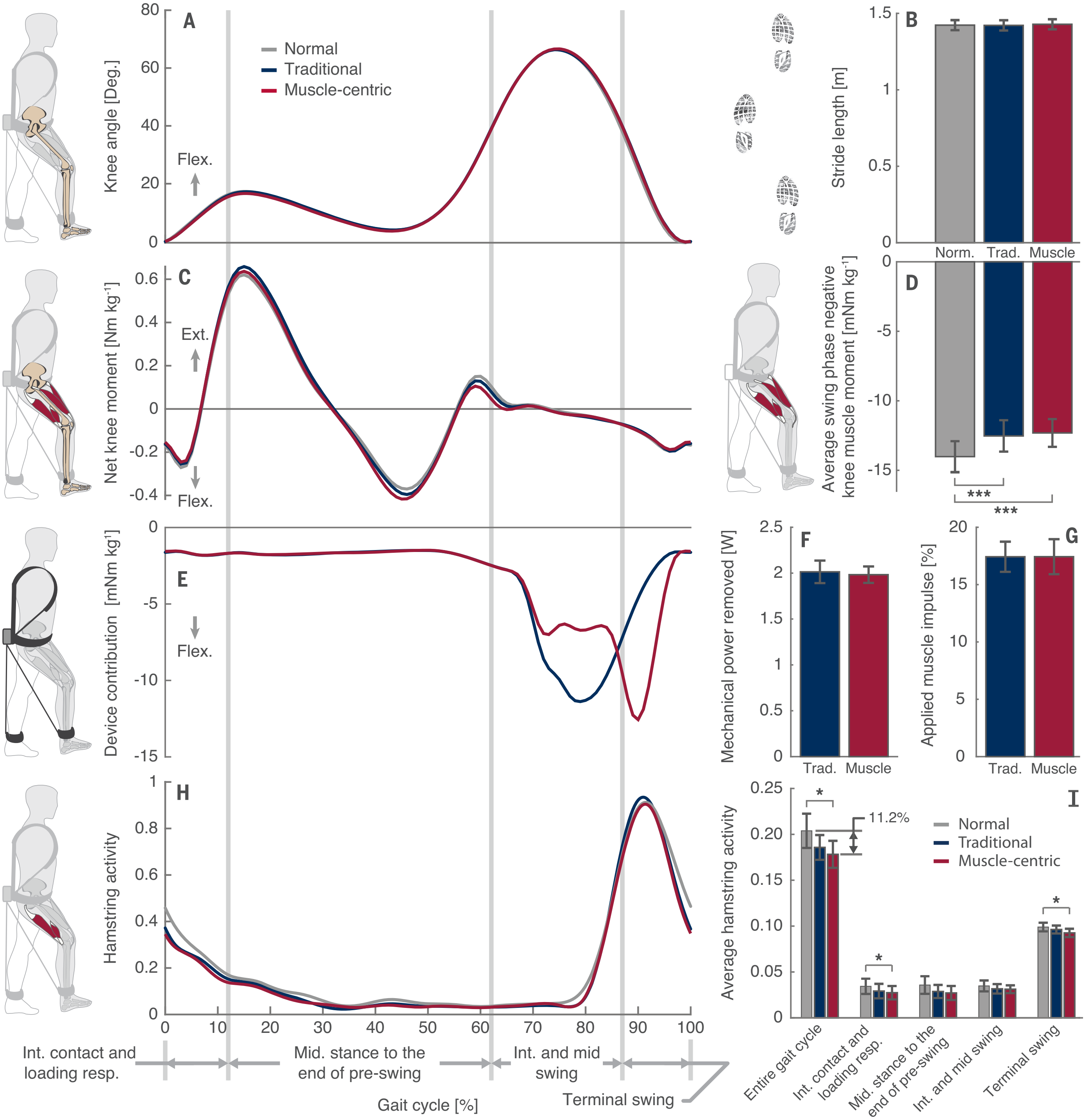


Fig. 2. Right leg sagittal plane knee kinematics, kinetics, and muscle activation averaged across participants. (**A**) Knee joint angles over a gait cycle for the three walking activities: normal walking (gray), traditional profile (blue), and muscle-centric profile (red). (**B**) Average stride lengths. (**C**) Net knee moment over a gait cycle normalized to body mass. (**D**) Average estimated negative muscle moment contribution during the swing period. (**E**) Moment applied by the device about the user's knee over a gait cycle. (**F**) Average mechanical power removed by the device. (**G**) Percentage of the negative muscle impulse applied by the device during the swing period. (**H**) Hamstring muscle activity over a gait cycle. (**I**) Average hamstring muscle activity for an entire gait cycle and specific gait phases (right groupings). $*P < 0.05$, $***P < 0.001$. Data are presented as means ± SEMs. Flex., flexion; Ext., extension; Int., initial; resp., response.

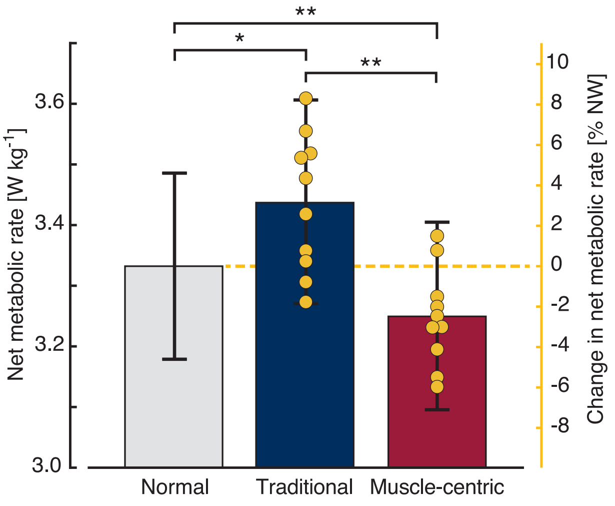


Fig. 3. Net metabolic rate. The average net metabolic rate for the three walking activities (gray, normal walking; blue, traditional profile; red, muscle-centric profile) is normalized to the participant's body mass (mean ± SEM). The left vertical axis represents the net metabolic rate (means and error bars; watts per kilogram). The right vertical axis represents the metabolic rate as a percent change from normal walking (NW; means and scatter). As such, no scatter is included for the normal walking condition. Net metabolic rate significantly differed between walking conditions (*$P < 0.05$, **$P < 0.01$). Each yellow circle represents one participant.

hoc analysis: $P = 0.035$; Fig. 2, H and I). This decrease primarily occurred during terminal swing (87 to 100% gait cycle; $n = 10$; one-way repeated-measures ANOVA; $F_{2,18} = 4.022$, $P = 0.036$, $\eta_p^2 = 0.309$; LSD post hoc analysis: $P = 0.021$) but continued into the initial contact and loading response phases (0 to 12% gait cycle; $n = 10$; one-way repeated-measures ANOVA; $F_{1.255,11.293} = 5.178$, $P = 0.037$, $\eta_p^2 = 0.365$; LSD post hoc analysis: $P = 0.026$). By contrast, the traditional profile did not significantly alter hamstring activity ($P = 0.102$).

Improper energy extraction could increase antagonistic muscle activity, as the body would need to perform positive muscle work to counteract the negative work performed by the device and recoup lost energy or provide joint stability. However, there was no significant change in overall average quadriceps activity in either loading condition, indicating that neither profile interfered with natural energy return mechanisms. Data for the quadriceps muscle group and individual muscles are presented in figs. S5 and S6 and data S5 and S6.

In addition to decreasing the metabolic cost of walking by using the muscle-centric profile, the device also converted the removed kinetic energy into 0.25 ± 0.02 W of electrical power. When applying the traditional profile, the exoskeleton produced twice the amount of electrical power (0.53 ± 0.03 W), albeit at a metabolic disadvantage.

The increase in metabolic cost associated with the traditional profile was not unexpected, but the cause is unclear. Although both profiles yielded similar reductions in average muscle joint moments, the traditional profile may not have reduced the muscle force and, consequently, the muscles' energy consumption to the same extent as the muscle-centric profile. This is because the muscle-tendon moment arms of the hamstrings are not directly proportional to muscle joint moments throughout gait (*24*). Additionally, the force generation and energy consumption of muscles depend on multiple parameters, including operating length and contractile velocity, which vary throughout the gait cycle (*25*, *26*). Furthermore, hamstrings are maximally extended during terminal swing, owing to simultaneous hip flexion and knee extension (*27*). As such, force production may be most metabolically costly during this stage. Therefore, force production may be more readily assisted during terminal swing relative to other swing phases (see supplementary materials). Additionally, the two profiles' different effects on the hip may have contributed to the observed differences in metabolic cost (see supplementary materials). Our results provide further evidence of the disconnect between lower-limb joint dynamics and metabolic expenditure (*11*, *28*). On the basis of these findings, we advocate for exoskeletons that use muscle-specific parameters and profiles to provide metabolic assistance (*29*).

We have demonstrated that exclusively removing energy is a viable method for reducing the metabolic cost of walking. Although greater metabolic cost reductions have been achieved through assisting concentric and isometric muscle contractions, our results demonstrate that assisting eccentric contractions is possible and can yield the concurrent benefit of electrical power production. These results indicate that lower-limb muscles can be assisted during the swing period of gait, despite previous assumptions that this period is primarily passive, requiring little to no muscle activity (*30*). A notable corollary to the present findings is that metabolic cost is likely not the only variable involved in gait adaptation, given that we observed little impact of our exoskeleton on gait parameters while reducing metabolic cost. Our findings highlight the importance of characterizing how muscles consume metabolic energy to perform mechanical tasks so that researchers can better understand activities such as walking and produce more efficient assistive devices.

REFERENCES AND NOTES

1. M. D. Sockol, D. A. Raichlen, H. Pontzer, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12265–12269 (2007).
2. H. Forssberg, *Exp. Brain Res.* **57**, 480–493 (1985).
3. M. Gomes, A. Ruina, *Phys. Rev. E* **83**, 032901 (2011).
4. T. McGeer, *Int. J. Robot. Res.* **9**, 62–82 (1990).
5. R. M. Alexander, *Am. Zool.* **24**, 85–94 (1984).
6. K. R. Westerterp, G. Plasqui, *Curr. Opin. Clin. Nutr. Metab. Care* **7**, 607–613 (2004).
7. K. R. Westerterp, *Front. Physiol.* **4**, 90 (2013).
8. K. E. Zelik, A. D. Kuo, *J. Exp. Biol.* **213**, 4257–4264 (2010).
9. L. M. Mooney, E. J. Rouse, H. M. Herr, *J. Neuroeng. Rehabil.* **11**, 151 (2014).
10. J. Kim *et al.*, *Science* **365**, 668–672 (2019).
11. S. H. Collins, M. B. Wiggin, G. S. Sawicki, *Nature* **522**, 212–215 (2015).
12. L. C. Rome, L. Flynn, E. M. Goldman, T. D. Yoo, *Science* **309**, 1725–1728 (2005).
13. J. M. Donelan *et al.*, *Science* **319**, 807–810 (2008).
14. M. Shepertycky, Q. Li, *PLOS ONE* **10**, e0127635 (2015).
15. R. Kram, C. R. Taylor, *Nature* **346**, 265–267 (1990).
16. T. W. Ryschon, M. D. Fowler, R. E. Wysong, A. Anthony, R. S. Balaban, *J. Appl. Physiol.* **83**, 867–874 (1997).
17. S. Galle, P. Malcolm, S. H. Collins, D. De Clercq, *J. Neuroeng. Rehabil.* **14**, 35 (2017).
18. P. Malcolm, W. Derave, S. Galle, D. De Clercq, *PLOS ONE* **8**, e56137 (2013).
19. L. M. Mooney, H. M. Herr, *J. Neuroeng. Rehabil.* **13**, 4 (2016).
20. M. Grimmer *et al.*, *J. Biomech.* **83**, 76–84 (2019).
21. B. T. Quinlivan *et al.*, *Sci. Robot.* **2**, eaah4416 (2017).
22. B. Whittington, A. Silder, B. Heiderscheit, D. G. Thelen, *Gait Posture* **27**, 628–634 (2008).
23. M. J. Joyner, D. P. Casey, *Physiol. Rev.* **95**, 549–601 (2015).
24. M. G. Hoy, F. E. Zajac, M. E. Gordon, *J. Biomech.* **23**, 157–169 (1990).
25. B. R. Umberger, K. G. M. Gerritsen, P. E. Martin, *Comput. Methods Biomech. Biomed. Engin.* **6**, 99–111 (2003).
26. E. M. Arnold, S. R. Hamner, A. Seth, M. Millard, S. L. Delp, *J. Exp. Biol.* **216**, 2150–2160 (2013).
27. S. L. Delp, A. S. Arnold, R. A. Speers, C. A. Moore, *J. Orthop. Res.* **14**, 144–151 (1996).
28. D. J. Farris, G. S. Sawicki, *J. Appl. Physiol.* **113**, 1862–1872 (2012).
29. O. N. Beck, L. K. Punith, R. W. Nuckols, G. S. Sawicki, *Exerc. Sport Sci. Rev.* **47**, 237–245 (2019).
30. T. M. Griffin, T. J. Roberts, R. Kram, *J. Appl. Physiol.* **95**, 172–183 (2003).

ACKNOWLEDGMENTS

We thank K. Deluzio for supplying experimental equipment in the Queen's University Human Mobility Research Laboratory and T. Bryant, C. Soutar, and P. Grenier for providing editorial feedback. **Funding:** This work was supported by grants from NSERC to Q.L. (RGPIN-2020-04771) and Y.-F.L. (RGPIN-2019-06635) and an NSERC Postgraduate Scholarship to M.S. **Author contributions:** M.S. and Q.L. designed the mechanical system. M.S. fabricated the device. M.S., S.B., A.D., and Y.-F.L. designed the electrical and control systems. M.S. and Q.L. designed and conducted experiments and processed and analyzed the data. All authors prepared, reviewed, and approved the manuscript. **Competing interests:** M.S., Y.-F.L., and Q.L. are the assignees on the patent of an energy harvesting technology related to the work in this manuscript (US Patent 9,407,125 B2). The authors declare no other competing interests. **Data and materials availability:** All data are included in the manuscript and supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/957/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S4
References (*31*–*91*)
Movies S1 and S2
Data S1 to S6

21 January 2020; resubmitted 27 October 2020
Accepted 8 March 2021
10.1126/science.aba9947

FERROELECTRICS

Control of polarization in bulk ferroelectrics by mechanical dislocation imprint

Marion Höfling[1], Xiandong Zhou[1], Lukas M. Riemer[2], Enrico Bruder[1], Binzhi Liu[3], Lin Zhou[3,4], Pedro B. Groszewicz†, Fangping Zhuo[1], Bai-Xiang Xu[1], Karsten Durst[1], Xiaoli Tan[3], Dragan Damjanovic[2], Jurij Koruza[1]*, Jürgen Rödel[1]*

Defects are essential to engineering the properties of functional materials ranging from semiconductors and superconductors to ferroics. Whereas point defects have been widely exploited, dislocations are commonly viewed as problematic for functional materials and not as a microstructural tool. We developed a method for mechanically imprinting dislocation networks that favorably skew the domain structure in bulk ferroelectrics and thereby tame the large switching polarization and make it available for functional harvesting. The resulting microstructure yields a strong mechanical restoring force to revert electric field–induced domain wall displacement on the macroscopic level and high pinning force on the local level. This induces a giant increase of the dielectric and electromechanical response at intermediate electric fields in barium titanate [electric field–dependent permittivity (ε_{33}) ≈ 5800 and large-signal piezoelectric coefficient (d_{33}*) ≈ 1890 picometers/volt]. Dislocation-based anisotropy delivers a different suite of tools with which to tailor functional materials.

Ferroelectricity, like ferromagnetism and superconductivity, belongs to the class of cooperative phenomena. Strain engineering allows us to substantially modify related properties of thin films with a judicious choice of substrate materials. Strain engineering has been demonstrated for ferromagnetic materials (*1*) and superconductors (*2*). For ferroelectric films, phase transition temperatures have been enhanced by 400°C (*3*), and even non-ferroelectric materials have been made ferroelectric (*4*). Introducing dislocations allows localized strain engineering and affects the polarization reversal process (*5*). Although dislocations are generally the cause of degraded performance in functional oxides and are usually avoided at all costs, there are some excellent examples demonstrating their utility. Dislocations enable strong vortex pinning in high-temperature superconductors (*6*) required for high critical currents, whereas in thermoelectrics, they increase the heat-to-electricity conversion efficiency by affecting phonon scattering (*7*). In oxides, dislocations are not only associated with the surrounding strain fields, but often also exhibit charged cores and screening space-charge layers (*8*). These peculiar mechanical and electrical characteristics offer a large, yet mostly unexplored, potential to alter not only thermal (*7*), but also electrical conductivity (*8*, *9*). Numerical simulations have also confirmed the potential for interaction of dislocations with fundamental order parameters such as spontaneous polarization (P_s) and strain (*10*, *11*).

Obtaining high intrinsic electromechanical response and manipulating the mobility of ferroelectric domain walls [labeled as extrinsic contribution (*12*)] are two key challenges that exist in bulk ferroelectrics. Some common approaches for increasing the ferroelectric properties are strain engineering (*4*), polarization rotation (*13*), construction of phase boundaries (*14*), and exploitation of critical points (*15*). State-of-the-art concepts to control domain wall motion include point defect doping (*16*, *17*), domain engineering (*18*, *19*), and texturing (*20*).

We developed a method that uses dislocations to interact with the fundamental order and field parameters of bulk ferroelectric oxides, i.e., polarization and strain, at different length scales. This directly affects the force on, and thereby the movement of, domain walls. It introduces a temperature-stable macroscopic restoring force acting against the applied electric field and allows us to tune the dielectric and electromechanical properties. We

---

[1] Department of Materials and Earth Sciences, Technical University of Darmstadt, 64287 Darmstadt, Germany. [2] Group for Ferroelectrics and Functional Oxides, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [3] Department of Materials Science and Engineering, Iowa State University, Ames, IA 50011, USA. [4] Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, USA.
†Present address: Department of Radiation Science and Technology, Delft University of Technology, Delft 2629JB, Netherlands.
***Corresponding author. Email: koruza@ceramics.tu-darmstadt.de (J.K.); roedel@ceramics.tu-darmstadt.de (J.R.)**

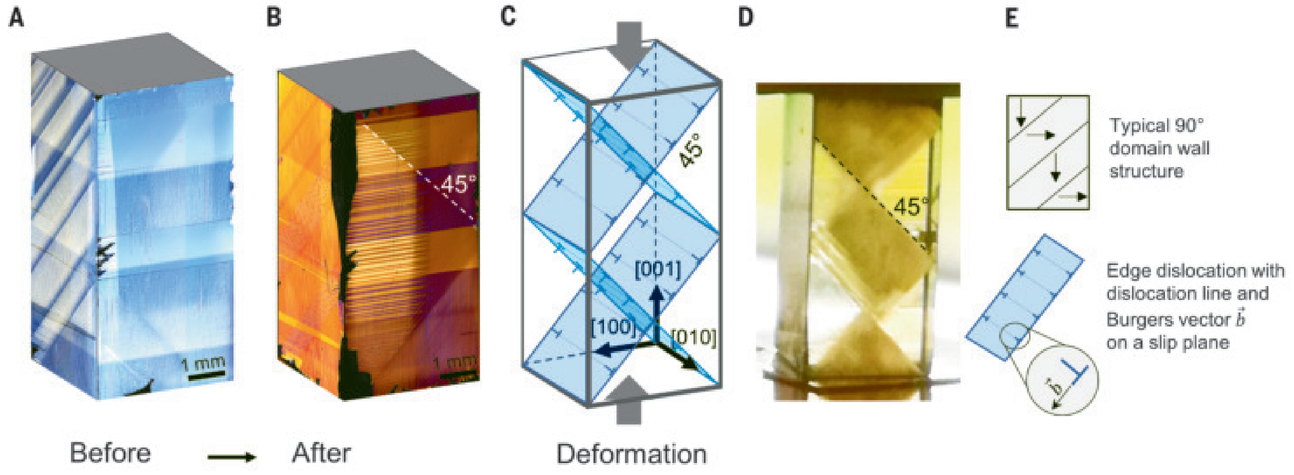

A
B
C
D
E
1 mm
1 mm
45°
45°
45°
[001]
[100]
[010]
Typical 90° domain wall structure
Edge dislocation with dislocation line and Burgers vector $\vec{b}$ on a slip plane
$\vec{b}$
Before
After
Deformation

Fig. 1. Changes of the ferroelectric domain configuration by mechanical dislocation imprint. (**A**) Typical domain arrangement for a tetragonal <001>-oriented undeformed $BaTiO_3$ single crystal (*23*) imaged with differential interference contrast, which causes the change in color. (**B**) The same imaging method used for the deformed sample. (**C** and **D**) Schematic presentation of the set of slip systems {101}<101> activated during uniaxial deformation at high temperature. This introduces a directional dislocation network, resulting in a reorientation of the domains according to the slip planes highlighted in the diagram in (C), yielding a highly unusual rhombic domain structure in the photograph shown in (D). Both the cloudy and clear regions in (B) and (D) exhibit a multidomain state with different domain density (fig. S2). (**E**) Schematic of the typical 90° domain wall alignment and the dislocations with dislocation line and Burgers vector.

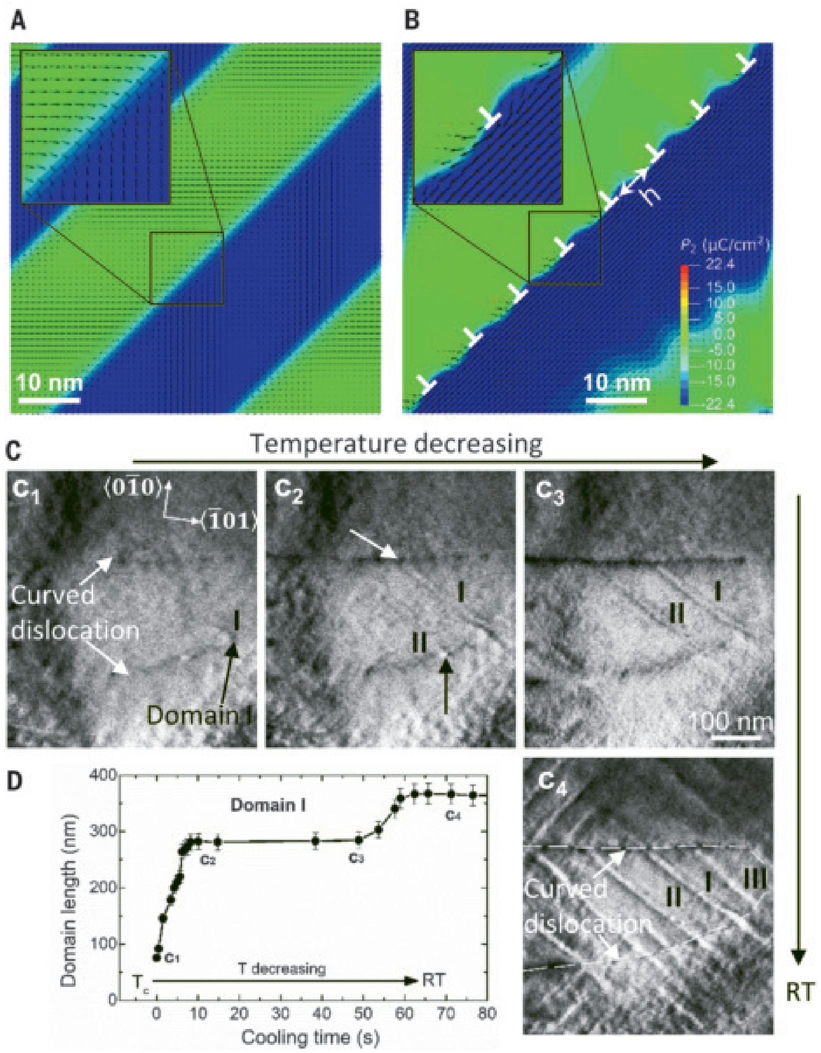


Fig. 2. Influence of dislocations on domain nucleation and mobility. (**A** and **B**) The steady state of local polarization without (A) and with (B) eight edge dislocations along the {101} plane as shown by phase-field simulations. Dislocations lead to both wavy domain walls and fluctuation of polarization amplitude, as demonstrated by the inset of (B). (**C**) In situ TEM study on a curved dislocation line during cooling through T_c (see movie S1). (**D**) Domain I nucleates at the lower segment of the dislocation (dark arrow in c_1), grows until it reaches the upper dislocation segment (white arrow in c_2), and is pinned there for a period of 40 s. Domains II (black arrow in c_2) and III are formed at the lower dislocation segment and grow until room temperature is reached (c_4). The temperature decrease is indicated with black arrows along the TEM images. Also shown is the link between simulated domain nucleation at the dislocation (B) and the formation of domains at the dislocation in temperature-dependent in situ TEM.

call our process mechanical dislocation imprinting because it is reminiscent of the mechanical imprinting used for engraving names or security codes. We also distinguish our process from one in which an electric charge is imprinted (*21*, *22*).

We used mechanical creep to introduce a directional dislocation network into <001>-oriented $BaTiO_3$ single crystals (fig. S1), and demonstrated that the dislocations act as nucleation sites for ferroelectric domain walls. The mechanical dislocation imprint causes a 19-fold increase in the large-signal piezoelectric coefficient (d_{33}*). Our methodology is readily accessible for furnishing bulk functional materials with permanent, tailored anisotropy.

The mechanical imprint is macroscopically reflected in the altered domain structure. The original lamellar domain configuration (*23*) (Fig. 1A) was changed into an unconventional rhombic domain structure with both clear and cloudy regions after high-temperature deformation (Fig. 1, B to D). Both regions exhibited a multidomain state with different domain density (fig. S2). The plastic deformation enforced an anisotropy with two-thirds of the domains in the deformed sample aligned along the deformation axis [001] (fig. S3). We used electron channeling contrast imaging (ECCI; fig. S4) to provide a microscopic view, featuring the orientation of imprinted glide planes on the surface. The dislocation spacing was in the range of 80 to 450 nm (fig. S4C). We used controlled indentation (fig. S4D) as a benchmark for the observed dislocation structures. We used transmission electron microscopy (TEM; fig. S5) to reveal a Burgers vector of $\boldsymbol{b} = a[101]$, where a is the lattice parameter. ECCI and TEM images confirmed the successful dislocation imprint by activation of the {101}<101> slip system.

Dislocations are associated with high local tensile and compressive stress fields (fig. S6, A and B) and, in some oxides, with a charged dislocation core. In ferroelectrics, this prompts a nucleation of domain walls (*24*). In our samples, this resulted in a rhombic domain structure (Fig. 1, B and D). We investigated the influence of the dislocations on the domain structure and the switching behavior using mechanically coupled phase-field (PF) simulations based on Ginzburg-Landau equations. The P_2 polarization component for a stable equilibrium configuration without dislocations is depicted in Fig. 2A, whereby straight 90° domain walls were formed on the {101} plane. We simulated the active slip system (Fig. 2B) with eight edge dislocations (Burgers vector $|\boldsymbol{b}| = 0.56$ nm), forming a dislocation array with $\boldsymbol{b} = a[101]$. From these results, we determined that the domain wall becomes kinked and broadens in the vicinity of a dislocation. Fluctuations of the local polarization and stress were strongly increased (Fig. 2B and fig. S6, A and B). This increase is in agreement with previous simulations investigating the impact of dislocation spacing, h, Burgers vector, and the strength of the dislocation-domain domain wall interaction (*10*, *11*) on the fluctuation of the local polarization.

We found a domain wall pinning effect of the dislocations using temperature-dependent in situ TEM, resulting in a controlled and localized nucleation and motion of the domain walls (Fig. 2C and movie S1). A similar effect had been predicted by simulations (*25*) and was reported in $Pb(Zr,Ti)O_3$ thin films with field-dependent in situ TEM (*26*). Upon cooling through the Curie temperature (T_c, 120°C), the first domain nucleated at the lower dislocation segment (I, dark arrow in Fig. 2C, c_1) and grew until it reached the upper dislocation segment. The latter pinned the domain temporarily (Fig. 2D) before it could overcome the dislocation barrier upon further cooling (Fig. 2C, c_4). At the same time, two new domains (II, III) nucleated at the lower dislocation segment. At room temperature (Fig. 2C, c_4), domain I was blocked by a {101} domain wall, whereas domains II and III were still blocked by the upper dislocation segment. The dislocation-domain wall interactions that we observed in the TEM were corroborated by PF simulations as being a function of electric field and temperature (fig. S7, A and B).

Macroscopically, the increase in population density of *c*-domains (out-of-plane P_s) versus *a*-domains (in-plane P_s) in the unpoled sample (fig. S3) caused reduction of the permittivity after deformation (fig. S8A) due to anisotropy of the dielectric tensor of BT with $\varepsilon_a > \varepsilon_c$. By contrast, the poled deformed samples retained a high fraction of *a*-domains (fig. S3), leading to an enhanced permittivity (by a factor of five) compared with the reference samples. Clearly, domain wall density will also come into play (*18*), but this appears to be a secondary issue.

Application of large-signal super-coercive fields (Fig. 3A) highlights the substantial impact of mechanical dislocation imprint. Despite loop saturation, polarization did not reach the expected spontaneous polarization value of 26 μC/cm^2, indicating domain wall pinning. Concurrently, we observed a fourfold increase in domain back-switching ($P_{back} = P_{max}^* - P_{rem}^*$) compared with the undeformed reference samples (fig. S9), which indicates a strong macroscopic restoring force. The apparent coercive field (E_c^*), was strongly enhanced, supporting domain wall pinning (for statistical confirmation, see fig. S9).

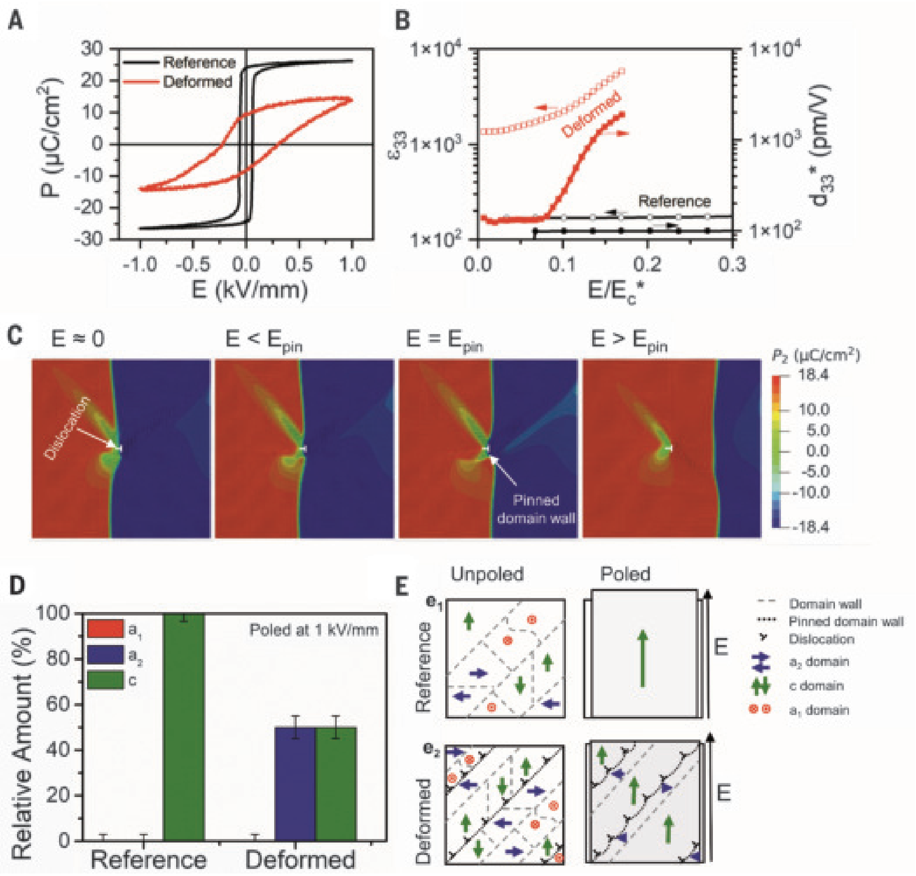


Fig. 3. Influence of mechanical dislocation imprint on electrical properties. (**A**) The presence of dislocations yields a decrease in P_{max}^* and pronounced back-switching. (**B**) Substantial increase in d_{33}^* and electric field–dependent permittivity (ε_{33}) for the poled deformed sample with increasing AC electric field (normalized by E_c^* at 1 kV/mm). (**C**) Simulation of the dislocation-domain wall interaction. When the electric field is relatively low, the domain wall is pinned at the dislocation and gets kinked and bent. When a pinning electric field is reached, the wall can break through the barrier of a single dislocation. (**D**) Comparison of the domain distribution of a reference and a deformed sample poled at 1 kV/mm, as obtained from nuclear magnetic resonance (NMR) measurements (fig. S3). (**E**) Schematic depiction of the domain structures. Due to pinning of the *a-c* domain walls at the dislocation and the resulting macroscopic restoring force, the total net expansion in the *c* direction is reduced compared with the undeformed reference sample, but the switchable strain is strongly enhanced.

The full potential of the dislocation-domain wall interaction was obtained in the intermediate E-field regime (below E_c^*). Increasing the amplitude of the applied AC electric field beyond the small-signal excitation resulted in stronger domain wall vibration, whereas the restoring force imposed by the presence of *a*-domains tuned this displacement and ensured a back-switching of domains in regions with high dislocation density. As a result, both the large-signal dielectric permittivity and the piezoelectric coefficient increased exponentially at ~$E/E_c^* = 0.1$ and reached $\varepsilon_{33}' \approx 5810$ and $d_{33}^* \approx 1890$ pm/V at $E/E_c^* = 0.17$, respectively (Fig. 3B). The d_{33}^* value was ~19 times higher compared with the undeformed reference sample ($d_{33}^* \approx 98$ pm/V, $\varepsilon_{33}' \approx 170$). Because of the long-range restoring force, this domain wall motion was almost anhysteretic below E_c^* (fig. S10). Note that the high permittivity and piezoelectric coefficient values remained stable up to 130,000 AC cycles (fig. S11) and a temperature of 75°C (fig. S12, A and B). In addition, the overall behavior featured a weak frequency dependence in d_{33}^* (fig. S13).

Our results indicate that the uniaxial stress activated four out of six available glide planes. The ensuing dislocation networks stabilized two domain wall variants (a_1-c and a_2-c) but disfavored the a_1-a_2 variant (fig. S3). Therefore, the imprinted mechanical dislocation structure caused an anisotropy in the domain structure, both in the unpoled and the poled state. The enforced presence of the *a*-domains provided a strain incompatibility to electric field–enforced *c*-domains, resulting in an elastic macroscopic restoring force akin to the case of thin films, in which the strain incompatibility is provided by the substrate. Locally, the dislocation network sitting on exactly the same plane as the domain wall provided a pinning force. The latter was confirmed by our numerical model (Fig. 3C) demonstrating pinning of a domain wall at a single dislocation. Domain wall bending and reversible domain wall movement increased up to $E = E_{pin}$ [pinning electric field]. Above E_{pin}, the domain wall overcame the local pinning potential. At the pinning electric field, domain walls at glide planes with low dislocation density became locally unpinned in an irreversible manner, but the overall remaining strain incompatibility ensured by glide planes with high dislocation density stabilized the *a-c* domain wall variant (Fig. 3D and fig. S14) and ensured complete back-switching (fig. S15). We visualized the increased density of *a-c* domain walls after deformation in the unpoled and poled state (Fig. 3E). We observed that mechanical dislocation imprint maintained a thermally stable piezoelectric coefficient up to 75°C. In a $SrTiO_3$ perovskite, dislocation structures have been reported to feature stability up to 600°C (*8*), offering pinning sites in a large temperature range. Similar reversal to the domain structure at zero electric field can only be enforced by external mechanical compressive stress (*27*). Current material design options include the use of point defects (*17*) or complex poling conditions stabilizing independently *a*-domains and *c*-domains in different sample volumes (*19*). These design options until now have not been used for practical purposes because they are stable only for a limited number of electrical cycles and at low temperatures (*19*) or limited electric field levels (*20*). Our dislocation-based mechanism is fundamentally different from acceptor doping (*17*), in which pinning centers are homogeneously distributed, allowing only relatively short motion of domain walls. In mechanically deformed samples, dislocations are concentrated in some parts of the sample and depinned domain walls can therefore move for longer distances. In bulk ceramics, templated grain growth (*28*, *29*) affords outstanding properties, as demonstrated for energy storage materials with supreme reliability (*30*).

Mechanical dislocation imprinting provides a powerful mechanism to extend the local pinning potential of dislocations to the macroscopic level of bulk ceramics. Whereas in thin films, the substrate provides the tool for strain engineering and misfit dislocations are indispensable (*31*), in bulk materials, dislocations have to be introduced through creep, plastic deformation, or new methods such as flash sintering (*32*). Uniaxial stress selects a specific dislocation structure that installs the in-plane strain permanently into the volume, thus providing a macroscopic restoring force. In addition, this method provides a dislocation structure for local effects such as domain wall pinning. This mechanical dislocation imprint both tames strong domain wall switching under applied electric field and allows the harvesting of this strain change caused by mechanically enforced back-switching. Subcoercive electric fields therefore reach extraordinary piezoelectric coefficients. This mechanism extends the available spectrum of tools for designing the functional properties of bulk functional materials such as ferroics and superconductors.

REFERENCES AND NOTES

1. R. S. Beach *et al*., *Phys. Rev. Lett.* **70**, 3502–3505 (1993).
2. I. Bozovic, G. Logvenov, I. Belca, B. Narimbetov, I. Sveklo, *Phys. Rev. Lett.* **89**, 107001 (2002).
3. K. J. Choi *et al*., *Science* **306**, 1005–1009 (2004).
4. J. H. Haeni *et al*., *Nature* **430**, 758–761 (2004).
5. S. V. Kalinin *et al*., *Adv. Mater.* **22**, 314–322 (2010).
6. B. Dam *et al*., *Nature* **399**, 439–442 (1999).
7. S. I. Kim *et al*., *Science* **348**, 109–114 (2015).
8. K. K. Adepalli, J. Yang, J. Maier, H. L. Tuller, B. Yildiz, *Adv. Funct. Mater.* **27**, 1700243–1700249 (2017).
9. L. Porz *et al*., *ACS Nano* acsnano.0c04491 (2020).
10. A. Kontsos, C. M. Landis, *Int. J. Solids Struct.* **46**, 1491–1498 (2009).
11. H. H. Wu, J. Wang, S. G. Cao, T. Y. Zhang, *Appl. Phys. Lett.* **102**, 232904 (2013).
12. D. Damjanovic, *Rep. Prog. Phys.* **61**, 1267–1324 (1998).
13. H. Fu, R. E. Cohen, *Nature* **403**, 281–283 (2000).
14. R. Guo *et al*., *Phys. Rev. Lett.* **84**, 5423–5426 (2000).
15. Z. Kutnjak, J. Petzelt, R. Blinc, *Nature* **441**, 956–959 (2006).
16. K. Carl, K. H. Härdtl, *Ferroelectrics* **17**, 473–486 (1978).
17. X. Ren, *Nat. Mater.* **3**, 91–94 (2004).
18. S. Wada *et al*., *Ferroelectrics* **355**, 37–49 (2007).
19. Q. Wang, F. Li, *J. Phys. D Appl. Phys.* **51**, 255301 (2018).
20. Y. Saito *et al*., *Nature* **432**, 84–87 (2004).
21. N. Hur *et al*., *Nature* **429**, 392–395 (2004).
22. B. Akkopru-Akgun, W. Zhu, M. T. Lanagan, S. Trolier-McKinstry, *J. Am. Ceram. Soc.* **102**, 5328–5341 (2019).
23. W. J. Merz, *Phys. Rev.* **88**, 421–422 (1952).
24. R. C. Bradt, G. S. Ansell, *J. Appl. Phys.* **38**, 5407–5408 (1967).
25. S. Y. Hu, Y. L. Li, L. Q. Chen, *J. Appl. Phys.* **94**, 2542–2547 (2003).
26. P. Gao *et al*., *Nat. Commun.* **2**, 591 (2011).
27. E. Burcsu, G. Ravichandran, K. Bhattacharya, *Appl. Phys. Lett.* **77**, 1698–1700 (2000).
28. G. L. Messing *et al*., *J. Mater. Res.* **32**, 3219–3241 (2017).
29. S. Kong, J. Daniels, *Appl. Phys. Lett.* **117**, 182901 (2020).
30. J. Li *et al*., *Nat. Mater.* **19**, 999–1005 (2020).
31. M. W. Chu *et al*., *Nat. Mater.* **3**, 87–90 (2004).
32. J. Cho *et al*., *Mater. Res. Lett.* **7**, 194–202 (2019).

ACKNOWLEDGMENTS

We thank S. Bauer and K. Ding for providing the reference indent in $BaTiO_3$, J. Lins for the acquisition of the NMR spectra, and G. Buntkowsky for access to the NMR spectrometer. B.X-X. and X.Z. thank the HHLR, Technical University of Darmstadt, for access to the Lichtenberg High-Performance Computer and for technical support. **Funding:** This work is supported by project no. 414179371 of the German Research Foundation (DFG). The TEM experiments, performed at the Sensitive Instrument Facility at Ames Laboratory, were supported by the U.S. National Science Foundation (NSF) through grant no. DMR-1700014. J.K. acknowledges financial support from the Athene Young Investigator program of TU Darmstadt. L.M.R. acknowledges financial support from the Swiss National Science Foundation (grant no. 200021 172525). P.B.G. and J.K. acknowledge the Profile Area "PMP" of TU Darmstadt for providing the goniometer NMR probe. The PF simulation work was partially funded by project no. 398072825 of DFG. **Author contributions:** M.H. prepared the samples. M.H. and L.M.R. performed the dielectric and electromechanical measurements. F.Z. provided confirmation on an independent orientation. J.R. and J.K. designed the experiments and, together with D.D., supervised the analysis of the data. X.Z. and B.X-X. performed the phase field simulations. B.L., L.Z., and X.T. performed the TEM analysis. P.B.G. planned and analyzed the NMR measurements and E.B. the ECCI measurements. D.D., J.R., J.K., M.H., and K.D. discussed the deformation mechanisms. M.H., J.K., and J.R. prepared the manuscript. All authors contributed and commented on the text. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/961/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Movie S1
References (*33*–*50*)

18 August 2020; accepted 26 April 2021
10.1126/science.abe3810

SPIN PHYSICS

Free coherent evolution of a coupled atomic spin system initialized by electron scattering

Lukas M. Veldman[1], Laëtitia Farinacci[1], Rasa Rejali[1], Rik Broekhoven[1], Jérémie Gobeil[1], David Coffey[1], Markus Ternes[2,3], Alexander F. Otte[1]*

Full insight into the dynamics of a coupled quantum system depends on the ability to follow the effect of a local excitation in real-time. Here, we trace the free coherent evolution of a pair of coupled atomic spins by means of scanning tunneling microscopy. Rather than using microwave pulses, we use a direct-current pump-probe scheme to detect the local magnetization after a current-induced excitation performed on one of the spins. By making use of magnetic interaction with the probe tip, we are able to tune the relative precession of the spins. We show that only if their Larmor frequencies match, the two spins can entangle, causing angular momentum to be swapped back and forth. These results provide insight into the locality of electron spin scattering and set the stage for controlled migration of a quantum state through an extended spin lattice.

One of the long-standing goals in experimental physics is the ability to create a "quantum simulator": an engineered system of coupled quantum degrees of freedom that can be initialized in an arbitrary state, allowing one to observe its intrinsic free evolution (*1*). In principle, scanning tunneling microscopy (STM) offers each of these ingredients. Individual magnetic atoms can be spatially arranged and studied by means of spin-polarized tunneling (*2*, *3*) and electron tunneling spectroscopy (*4*, *5*), allowing physicists to probe their local magnetization state and energy configuration, respectively. However, because of their slow time scales, these techniques have been able to observe the dynamic spin processes only indirectly (*6*–*8*).

In recent years, the STM toolbox was expanded to include pump-probe spectroscopy, which allows spin relaxation to be probed on the nanosecond time scale (*9*, *10*), as well as electron spin resonance performed locally at the probe tip (ESR-STM) (*11*). ESR-STM, combined with microwave ac pulsing schemes, enabled the observation of the coherent time evolution of a single atomic spin (*12*), which is on par with achievements in semiconductor

---

[1] Department of Quantum Nanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2628 CJ Delft, the Netherlands. [2] RWTH Aachen University, Institute of Physics, D-52074 Aachen, Germany. [3] Peter-Grünberg-Institute, Forschungszentrum Jülich, D-52425 Jülich, Germany.
*Corresponding author. Email: a.f.otte@tudelft.nl

spin qubits (*13*, *14*) and nitrogen-vacancy (NV) centers (*15*). However, to demonstrate the free evolution of a pair of entangled spins, the initial excitation has to be sufficiently fast to compete with the intrinsic dynamics set by the coupling strength. ESR-STM uses a Rabi flop process for initialization, the rate of which is limited by the radio frequency (rf) power available at the probe tip.

In this work, by sequentially combining the ESR-STM and dc pump-probe techniques, we demonstrate the detection of a free, coherent flip-flop evolution of two coupled spin-1/2 atoms resulting from an electron-induced spin excitation that is nearly instantaneous when compared with the typical duration of a Rabi rotation. Using the energy resolution of ESR-STM, we tuned the eigenstates of two coupled spin-1/2 particles from Zeeman states $|\uparrow\uparrow\rangle, |\downarrow\uparrow\rangle, |\uparrow\downarrow\rangle$, and $|\downarrow\downarrow\rangle$ to singlet-triplet states $|\uparrow\uparrow\rangle, |-\rangle, |+\rangle$, and $|\downarrow\downarrow\rangle$ by varying the tip height (*11*, *16*, *17*). Subsequently, by using a dc pump-probe scheme, we excited and read out the spin projection of one of the two spins with nanosecond resolution. This is in contrast to previous efforts, in which microwave ac pulses were used to control and read out spins (*12*). Rather than by means of a Rabi rotation, in this work the coherent evolution is initiated directly through an electron spin scattering event.

As we increased the delay between pump and probe pulses, we observed an oscillating magnetization for the spin underneath the tip, which we attributed to a flip-flop interaction between the two spins (*18*, *19*). This implies that the excitation process that is due to tunneling electron scattering is local; it only consists of a spin flip on the atom underneath the tip, irrespective of the energy eigenstates of the system. This is a noteworthy result in the light of previous works, in which it was deemed sufficient to consider electron-induced spin excitations as transitions between energy eigenstates (*20*–*22*).

We used a low-temperature STM to manipulate individual hydrogenated Ti atoms, henceforth referred to as TiH, on top of bilayer MgO islands on an Ag(100) crystal. To obtain spin polarization, we deposited Fe atoms and transferred them to the tip apex (*23*). The ESR and pump-probe experiments were performed by applying the rf voltage and pulse trains to the tip, at temperatures of 1.5 K and 400 mK, respectively. We used an external magnetic field B_{ext} = 450 mT in-plane at a 14° angle with respect to the MgO lattice to separate the energy levels by means of Zeeman splitting.

We studied TiH species without any observable nuclear spin that are adsorbed on bridge sites with different orientations with respect to the external magnetic field, as sketched in Fig. 1A. TiH on MgO has been shown to be an effective spin-1/2 particle with

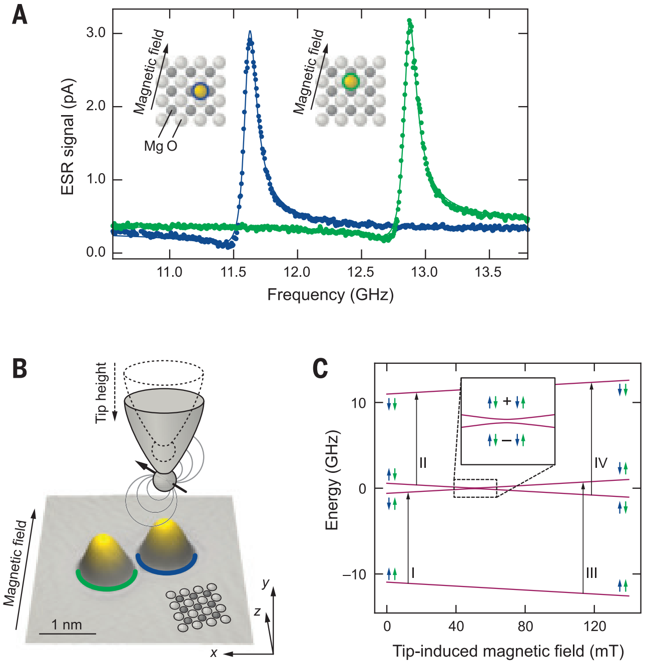


Fig. 1. Tuning the eigenstates of a TiH dimer by using a tip-induced magnetic field. (**A**) ESR measurements of single TiH adsorbed onto a vertical (blue) and a horizontal (green) bridge site [temperature (*T*) = 1.5 K, radio-frequency voltage (V_{rf}) = 30 mV, current (*I*) = 50 pA, V_{dc} = 60 mV, B_{ext} = 450 mT]. (**B**) STM topography of a TiH dimer with MgO lattice indication and schematic demonstrating tuning of the tip field. (**C**) Calculated energies and eigenstate compositions as function of tip field. An avoided crossing occurs at the point at which the tip field compensates the g-factor difference. Roman numerals indicate the four possible ESR transitions.

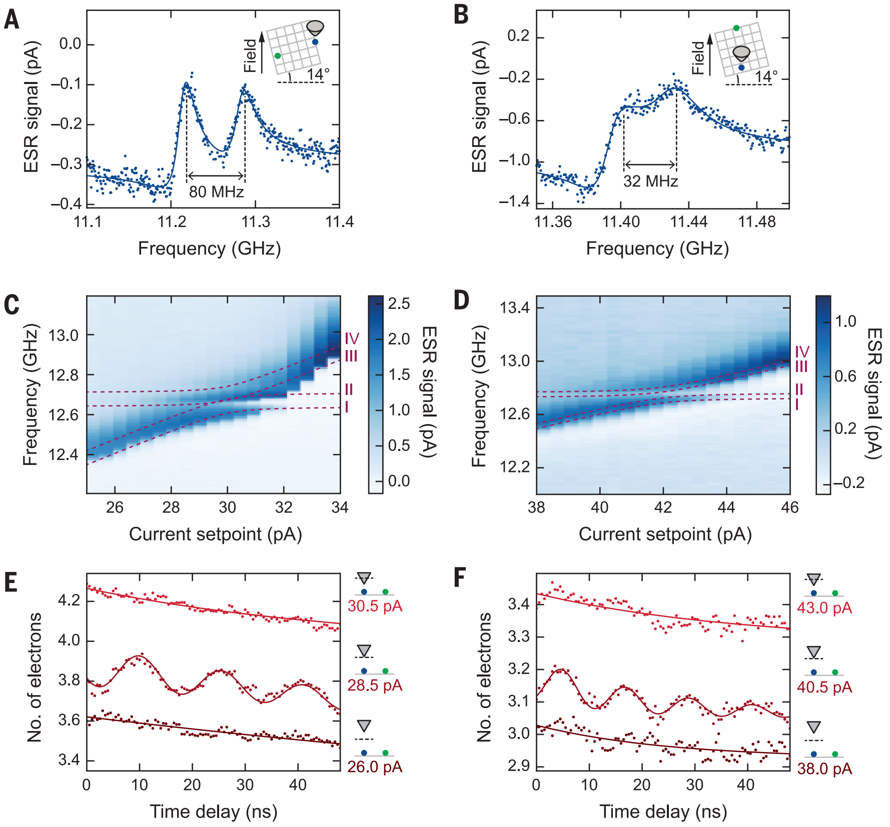


Fig. 2. Measurement of free coherent evolution at the tuning point. (**A** and **B**) ESR measurements on dimer A and dimer B (*T* = 1.5 K, V_{rf} = 50 mV, *I* = 10 pA, V_{dc} = 60 mV). (Insets) Schematic drawings of the dimer placement on the MgO lattice. (**C** and **D**) ESR measurements at various tip heights, showing an avoided crossing at the tuning point (*T* = 1.5 K, V_{rf} = 50 mV, V_{dc} = 60 mV). Dashed lines are guides to the eye that mark ESR transitions. (**E**) Pump-probe measurements on dimer A, above, below, and at the tuning point (*T* = 400 mK, setpoint voltage 60 mV, pulse width 7 ns). (**F**) Same as (E) for dimer B (pulse width 5 ns). All pump-probe experiments use +100-mV pump and −100-mV probe pulses.

an anisotropic g-factor (*24*, *25*). In agreement with these studies, we observed different ESR resonance frequencies for the two species; for the spin $\mathbf{S}_v$ of the vertically oriented TiH species (blue), we find a g-factor g_v = 1.75, whereas for the spin $\mathbf{S}_h$ of the horizontal TiH (green), we find g_h = 1.95 (supplementary text, sections S1 and S2).

Figure 1, B and C, demonstrates how we use the effective magnetic field that emanates from the tip on one of the two atoms to tune the level of entanglement between the spins (*10*, *17*). If the two spins experience the same effective Zeeman splitting, they precess at identical Larmor frequencies resulting in entangled states. Because we want to reach entanglement at a finite tip field, the two spins need to be inherently detuned in absence of the tip. For this reason, we built heterodimers out of vertically and horizontally oriented TiH species (Fig. 1B and supplementary text, section S3).

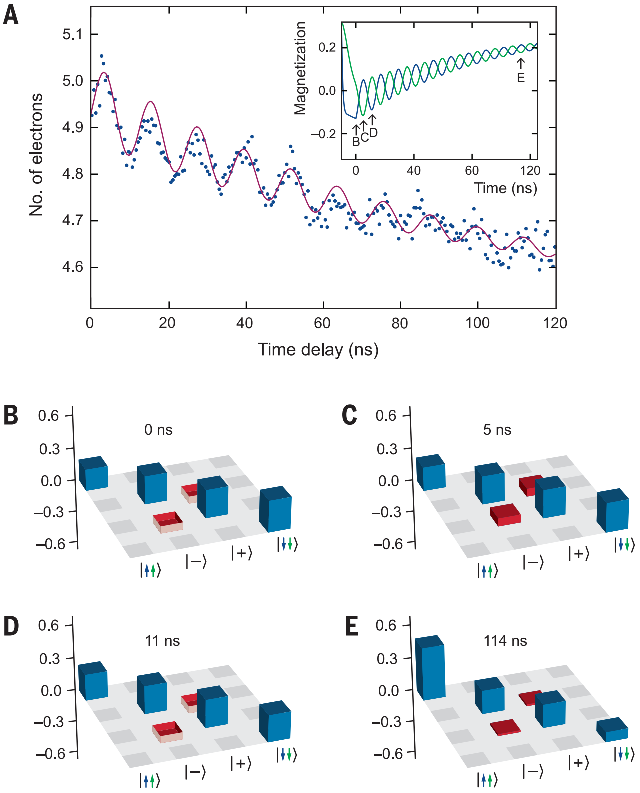


Fig. 3. Decoherence of the flip-flop oscillation. (**A**) Pump-probe measurement on dimer B showing the decay in amplitude of the oscillations (T = 400 mK, setpoint 40.5 pA, 60 mV, pulse width 7 ns). The solid line represents the calculated pump-probe signal. (Inset) Calculated magnetizations $\langle S_v^z \rangle$ (blue) and $\langle S_h^z \rangle$ (green); the origin of the time axis is set to coincide with the end of the pump pulse. (**B** to **E**) Density matrices at different times after the pump pulse [inset of (A)]. Off-diagonal elements are marked red for clarity.

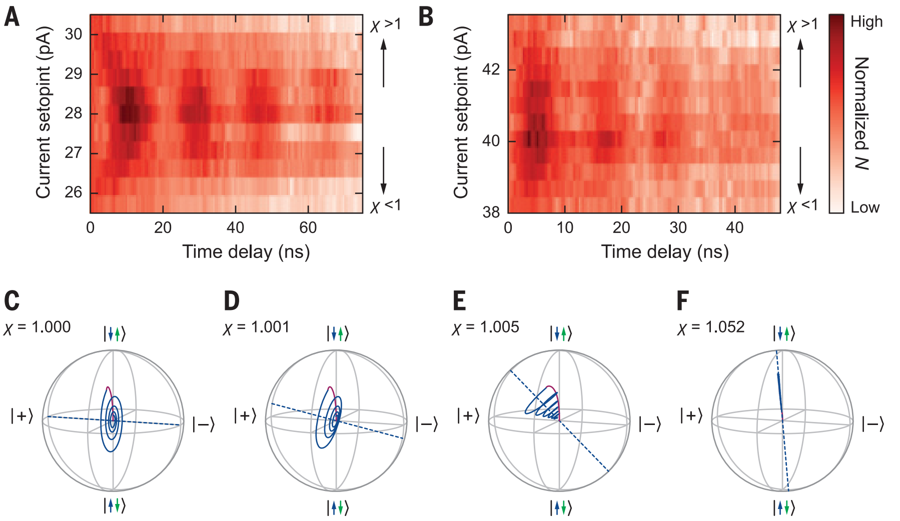


Fig. 4. Flip-flop oscillations as function of detuning. (**A** and **B**) Pump-probe measurements on dimers A and B at various tip heights (parameters same as those in Fig. 2). (**C** to **F**) Bloch sphere representations of the state evolution at increasing levels of detuning, as indicated with the detuning ratio $\chi = g_v(B_{ext} + B_{tip})/g_h B_{ext}$. The Bloch spheres show the reduced state space between the $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$ states. The excitation process due to the pump pulse is plotted in red, and the subsequent free evolution of the spin state is plotted in blue. For $\chi < 1$, the rotation axis (dashed blue line) tilts in the opposite direction.

The dimers were engineered to have a spacing of 1.3 nm, which corresponds to a coupling strength on the order of tens of megahertz. This coupling strength was chosen to ensure that the dynamics of the local magnetization are slow enough to be probed by our experimental setup, which is limited to ~5-ns pulses but still faster than the ~100-ns decoherence time of TiH dimers (*17*). At this distance, the atoms interact through both exchange and dipolar interactions. The Hamiltonian of the system can be written as (supplementary text, section S4)

$$\mathcal{H} = (J+2D)S_v^z S_h^z + (J-D)\left(S_v^x S_h^x + S_v^y S_h^y\right) - \mu_B B_{ext}\left(g_v S_v^z + g_h S_h^z\right) - \mu_B B_{tip} g_v S_v^z \qquad (1)$$

where μ_B is the Bohr magneton and J and D are the exchange and dipolar coupling parameters, respectively. The axes are defined in Fig. 1B. The two last terms account for the Zeeman splitting due to the external (B_{ext}) and effective tip (B_{tip}) fields, which, on the basis of their relative strengths, we assume to be aligned. In supplementary text, section S6, we discuss the justification and limits of this assumption.

We separated the exchange and dipolar contributions by performing the experiment on two heterodimers as sketched in Fig. 2, A (dimer A) and B (dimer B). The two dimers are equidistant, yielding identical exchange couplings. However, because they are oriented at different angles with respect to the external field, their dipolar coupling strengths differ. This was confirmed by ESR measurements performed on top of the vertically oriented TiH of each dimer with the tip well away from the tuning point (Fig. 2, A and B). In this situation, the $S_v^x S_h^x$ and $S_v^y S_h^y$ components of the coupling (Eq. 1) average out over time as the spins precess with different Larmor frequencies. The resulting coupling, being mediated through the $S_v^z S_h^z$ terms only, is effectively Ising-like. Because of the composition of the eigenstates and because ESR can only flip the spin underneath the tip (*26*), only transitions I and II were observed (Fig. 1C). The measured splitting between these two ESR resonances corresponds to $J + 2D$ and thus is different for the two heterodimers (Fig. 2, A and B).

To probe the full energy-level diagram of Fig. 1C and identify the exact tuning point for maximal entanglement, we performed ESR measurements at various tip heights for each dimer (Fig. 2, C and D). We observed two sets of peaks that, upon tip approach, shift together and broaden because of decoherence effects (*27*). These two sets of resonances can be assigned to transitions I and II and to III and IV in Fig. 1C. Away from the tuning point, only one of these pairs was observed: transition I and II before the tuning point and transition III and IV after it. As the energy eigenstates become more entangled near the tuning point, all four transitions become accessible. Because of the opposite signs of the dipolar coupling contributions, the two dimers show slightly different behaviors: transitions II and III intersect twice for dimer A, whereas they stay apart for dimer B. We find that $J = 67 \pm 2$ MHz and $D = 2 \pm 1$ MHz for dimer A and that $D = -15 \pm 1$ MHz for dimer B (supplementary text, section S5).

We then arrived at the second stage of the experiment, in which we measured the free time evolution of the spins. We used a pump-probe scheme to excite and measure the spin state of the atom underneath the tip for various degrees of entanglement. When the tip height is far away from the tuning point, the pump-probe experiments show the onset of an exponential decay that is similar to the decay signal of a single excited spin (Fig. 2, E and F, top and bottom curves) (*9*). By contrast, when tuned, we observed a clear oscillation with a frequency of 64 ± 1 MHz for dimer A and 84 ± 1 MHz for dimer B. We attribute these oscillations to the flip-flop interaction of strength $J - D$ between the two atoms in the dimer.

The dynamics of the flip-flop interaction can be well understood by describing the time evolution of the combined density matrix of the two spins within a dissipative Bloch-Redfield framework (*28*, *29*), which accounts for the uncorrelated electron baths in sample and tip (Fig. 3A and supplementary text, section S7). In Fig. 3, B to E, we show the density matrices in the energy basis obtained by numerical simulation for a perfectly tuned dimer at different moments in time after the pump pulse.

During the pump pulse, the system is pushed into a coherent superposition of its excited states (Fig. 3B). These add up to a net $|{\downarrow\uparrow}\rangle$ magnetization (with the left arrow corresponding to the spin underneath the tip) as a result of spin pumping (*23*). This net magnetization is reflected in the off-diagonal terms, which correspond to the coherence between the $|-\rangle$ and $|+\rangle$ states. Because of the finite thermal occupation of the excited states in the initial Boltzmann distribution, the pump pulse also populates the higher-energy $|{\downarrow\downarrow}\rangle$ state. Immediately after the pulse, the off-diagonals begin to oscillate between positive and negative values (Fig. 3, C and D), which gives rise to the observed periodicity in the magnetization (inset of Fig. 3A). Because of the interaction with the electron baths, the oscillations decay over an effective decoherence time, and eventually, the populations evolve back toward thermal equilibrium (Fig. 3E). We estimate the decoherence time to be 60 and 130 ns for the relaxation time (supplementary text, section S8).

We then proceeded to the effect of detuning on the flip-flop oscillations. For this purpose, we performed pump-probe experiments at different current setpoints around the tuning point: ~28 pA for dimer A and ~40 pA for dimer B (Fig. 4, A and B). As expected, the oscillations diminished rapidly as we tuned away from these values. Depending on the microscopic tip apex, we observed a small difference in tuning height between ESR (Fig. 2) and pump-probe measurements (Fig. 4). We attribute this discrepancy to the temperature dependence of the spin-polarized current, as the two different experiments were carried out at different temperatures (supplementary text, section S9).

To gain insight into the effect of detuning on the spin dynamics, we mapped the effective two-level system of the inner 2 by 2 matrix of the density matrix onto a Bloch sphere. For clarity, the axes of the sphere were fixed to be the energy eigenstates of the fully tuned case, and the projected spin state evolution is plotted for different levels of detuning (Fig. 4, C to F). As can be seen from the density matrices in Fig. 3, the spin state always has components outside the inner 2 by 2 matrix, meaning that the projection in Fig. 4 never reaches the surface of the sphere.

When the dimer is in tune (Fig. 4C), the state moves fully within the vertical plane of the Bloch sphere, making maximal flip-flops between $|{\uparrow\downarrow}\rangle$ and $|{\downarrow\uparrow}\rangle$. With increasing detuning, the axis that the state rotates around moves as the eigenstates of the system gradually tilts toward the vertical. The difference between the projected maxima and minima of the oscillation onto the vertical axis gets smaller, and thus, the oscillation amplitude decreases, which is consistent with our experimental observations.

The observed flip-flop frequency remains constant as a function of detuning. This is surprising, as the energy splitting is supposed to increase away from the tuning point, causing an increase in the frequency (supplementary text, section S10). Because the in-plane anisotropy of the g-factor indicates a partially unquenched orbital moment that results from the crystal-field symmetry of the bridge sites, as previously observed for the out-of-plane direction on TiH species on oxygen binding sites (*24*), the observed discrepancy may be related to this orbital moment.

By combining the energy resolution of ESR-STM and the time resolution of dc pump-probe spectroscopy, we have demonstrated an experimental procedure that enables the observation of the free coherent evolution of coupled atomic spins. Because the dynamic processes are initialized by means of a coherence-preserving pulse in the tunneling current, our method provides insight into the physics of electron spin scattering that could not be obtained by pulsed ESR-STM methods (*12*). Specifically, only the spin directly underneath the tip is affected by the spin excitation, irrespective of the global quantum state. In conjunction with the recent demonstration of pulsed ESR-STM, our technique offers pathways toward coherent manipulation of extended atomic spin arrays. The ability to perform a very local, nearly instantaneous, coherent spin flip inside an extended spin lattice constitutes an essential building block for advances in spintronic engineering as well as studies into the propagation of spin waves.

REFERENCES AND NOTES

1. J. I. Cirac, P. Zoller, *Nat. Phys.* **8**, 264–266 (2012).
2. S. Heinze *et al.*, *Science* **288**, 1805–1808 (2000).
3. F. Meier, L. Zhou, J. Wiebe, R. Wiesendanger, *Science* **320**, 82–86 (2008).
4. A. J. Heinrich, J. A. Gupta, C. P. Lutz, D. M. Eigler, *Science* **306**, 466–469 (2004).
5. C. F. Hirjibehedin, C. P. Lutz, A. J. Heinrich, *Science* **312**, 1021–1024 (2006).
6. M. Ternes, *New J. Phys.* **17**, 063016 (2015).
7. R. J. G. Elbertse, D. Coffey, J. Gobeil, A. F. Otte, *Commun. Phys.* **3**, 94 (2020).
8. L. Malavolti *et al.*, *Nanoscale* **12**, 11619–11626 (2020).
9. S. Loth, M. Etzkorn, C. P. Lutz, D. M. Eigler, A. J. Heinrich, *Science* **329**, 1628–1630 (2010).
10. S. Yan, D.-J. Choi, J. A. J. Burgess, S. Rolf-Pissarczyk, S. Loth, *Nat. Nanotechnol.* **10**, 40–45 (2015).
11. S. Baumann *et al.*, *Science* **350**, 417–420 (2015).
12. K. Yang *et al.*, *Science* **366**, 509–512 (2019).
13. J. J. Pla *et al.*, *Nature* **489**, 541–545 (2012).
14. E. Kawakami *et al.*, *Nat. Nanotechnol.* **9**, 666–670 (2014).
15. R. Hanson, V. V. Dobrovitski, A. E. Feiguin, O. Gywat, D. D. Awschalom, *Science* **320**, 352–355 (2008).
16. K. Yang *et al.*, *Phys. Rev. Lett.* **119**, 227206 (2017).
17. Y. Bae *et al.*, *Sci. Adv.* **4**, eaau4159 (2018).
18. J. R. Petta *et al.*, *Science* **309**, 2180–2184 (2005).
19. Y. He *et al.*, *Nature* **571**, 371–375 (2019).
20. C. F. Hirjibehedin *et al.*, *Science* **317**, 1199–1203 (2007).
21. A. F. Otte *et al.*, *Phys. Rev. Lett.* **103**, 107203 (2009).
22. A. Spinelli, B. Bryant, F. Delgado, J. Fernández-Rossier, A. F. Otte, *Nat. Mater.* **13**, 782–785 (2014).
23. S. Loth *et al.*, *Nat. Phys.* **6**, 340–344 (2010).
24. M. Steinbrecher *et al.*, *Phys. Rev. B* **103**, 155405 (2021).
25. P. Willke *et al.*, *Science* **362**, 336–339 (2018).
26. T. Choi *et al.*, *Nat. Nanotechnol.* **12**, 420–424 (2017).
27. P. Willke *et al.*, *Sci. Adv.* **4**, eaaq1543 (2018).
28. C. Cohen-Tannoudji, J. Dupont-Roc, G. Grynberg, *Atom-Photon Interaction* (Wiley, 1989).
29. F. Delgado, J. Fernández-Rossier, *Prog. Surf. Sci.* **92**, 40–82 (2017).
30. L. M. Veldman *et al.*, Free coherent evolution of a coupled atomic spin system initialized by electron scattering, Zenodo (2021); doi: 10.5281/zenodo.4467819.

ACKNOWLEDGMENTS

We thank V. Kornich and L. M. K. Vandersypen for scientific discussions. **Funding:** The authors acknowledge support from the Netherlands Organisation for Scientific Research (NWO Vici Grant VI.C.182.016) and from the European Research Council (ERC Starting Grant 676895 "SPINCAD"). M.T. acknowledges support by the Heisenberg Program (grant number TE 833/2-1) of the German Research Foundation. **Author contributions:** L.M.V., L.F., and R.R. performed the experiments. L.M.V., L.F., R.R., D.C., and J.G. implemented and optimized the experimental techniques. L.M.V. and L.F. analyzed the experimental data. R.B. and M.T. developed the dissipative Bloch-Redfield model and performed the simulations. L.M.V., R.B., and A.F.O. designed the experiment. L.M.V., L.F., R.B., M.T., and A.F.O. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data presented in this paper are publicly available through Zenodo (*30*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/964/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Table S1
References (*31*–*36*)

29 January 2021; resubmitted 19 February 2021
Accepted 23 April 2021
10.1126/science.abg8223

METABOLISM

Mitochondrial NADP(H) generation is essential for proline biosynthesis

Jiajun Zhu[1], Simon Schwörer[1]†, Mirela Berisa[2]†, Yeon Ju Kyung[1]†, Keun Woo Ryu[1], Junmei Yi[3], Xuejun Jiang[3], Justin R. Cross[2], Craig B. Thompson[1]*

The coenzyme nicotinamide adenine dinucleotide phosphate (NADP$^+$) and its reduced form (NADPH) regulate reductive metabolism in a subcellularly compartmentalized manner. Mitochondrial NADP(H) production depends on the phosphorylation of NAD(H) by NAD kinase 2 (NADK2). Deletion of *NADK2* in human cell lines did not alter mitochondrial folate pathway activity, tricarboxylic acid cycle activity, or mitochondrial oxidative stress, but rather led to impaired cell proliferation in minimal medium. This growth defect was rescued by proline supplementation. NADK2-mediated mitochondrial NADP(H) generation was required for the reduction of glutamate and hence proline biosynthesis. Furthermore, mitochondrial NADP(H) availability determined the production of collagen proteins by cells of mesenchymal lineage. Thus, a primary function of the mitochondrial NADP(H) pool is to support proline biosynthesis for use in cytosolic protein synthesis.

Mammalian cells depend on the interconversion of nicotinamide adenine dinucleotide phosphate molecules between the oxidized (NADP$^+$) and reduced (NADPH) forms to support reductive biosynthesis and to maintain cellular antioxidant defense. NADP$^+$ and NADPH molecules [NADP(H) hereafter] are unable to cross subcellular membranes (*1*, *2*). As a result, cellular pools of NADP(H) are compartmentalized. In the cytosol, NADP(H) is derived from nicotinamide adenine dinucleotide [NAD(H)] by NAD kinase (NADK, referred to as NADK1 hereafter). Cytosolic NADPH acts as a substrate in fatty acid biosynthesis and as the reducing equivalent required to regenerate reduced glutathione (GSH) and thioredoxin for antioxidant defense. Mitochondria host a number of biosynthetic activities critical for cellular metabolism but are also major sites for the generation of reactive oxygen species (ROS). Mammalian mitochondrial NAD kinase (NADK2) converts NAD(H) to NADP(H) through phosphorylation (*3*).

Using subcellular fractionation, we confirmed that NADK2 was purified in the membrane-associated fraction in cultured human cell lines (fig. S1, A to C). Mitochondria immunopurification (Mito-IP) (*4*, *5*) from DLD1 cells after CRISPR-Cas9 deletion of *NADK2* (fig. S1D) resulted in a metabolomic profile consistent with mitochondrial metabolism, and metabolites known to be excluded from the mitochondrial compartment were minimally detected (Fig. 1A; fig. S1, E to G; and table S1). We examined NADP(H) levels in immunopurified mitochondria using an adapted enzyme cycling assay (*6*). Although total NADP(H) abundance or the NADP$^+$-to-NADPH ratio were not changed at the whole-cell level upon NADK2 loss as previously reported (*6*, *7*), mitochondrial NADP(H) abundance was reduced by >80% ($P < 0.001$) in *NADK2* knockout cells (Fig. 1, B and C, and fig. S1, H to J). NAD(H) abundance or the NAD$^+$-to-NADH ratio were not altered by *NADK2* knockout in whole cells or in mitochondria (fig. S1, K to N).

Oncogenic mutant forms of isocitrate dehydrogenase 1 (IDH1) and IDH2 require cytosolic and mitochondrial NADPH, respectively, to produce 2-hydroxyglutarate (2HG) from α-ketoglutarate (αKG) (*8*) (fig. S1O). We deleted the *NADK2* gene in chondrosarcoma cell lines that had either an endogenous IDH1 R132 mutation (JJ012) or an IDH2 R172 mutation (CS1) (Fig. 1D). Loss of NADK2 resulted in reduced 2HG abundance ($P < 0.001$) in CS1 cells but not in JJ012 cells (Fig. 1, E and F). We further subjected control and *NADK2*-deleted CS1 cells to a xenograft tumor assay in vivo and observed similarly decreased 2HG abundance in tumors formed by *NADK2* knockout cells (Fig. 1G). These results confirmed that NADK2 is required to maintain the mitochondrial NADP(H) pool.

Methylenetetrahydrofolate dehydrogenase 2 (MTHFD2) and MTHFD2-like (MTHFD2L) use either NAD$^+$ or NADP$^+$ as electron acceptors in the mitochondrial folate pathway. Using [2,3,3-^{2}H3]serine isotope tracing, cells lacking *MTHFD2* or serine hydroxymethyltransferase 2 (*SHMT2*) both displayed an increase in doubly labeled thymidine triphosphate (TTP M+2) when compared with control cells (Fig. 2, A to C, and fig. S2, A and B), which suggests decreased mitochondrial folate pathway activity and increased cytosolic serine catabolism, as previously reported (*9*, *10*). By contrast, cells lacking *NADK2* maintained the fraction of singly labeled TTP (TTP M+1) derived from [2,3,3-^{2}H3]serine (Fig. 2, A to C, and fig. S2, A and B), which indicates that the mitochondrial folate pathway is not disrupted by NADK2 loss.

We performed isotope tracing experiments with uniformly labeled [U-^{13}C]glucose or [U-^{13}C]glutamine comparing control and *NADK2*-deleted cells to analyze tricarboxylic acid (TCA) cycle activity. We did not observe consistent changes in the TCA cycle intermediates derived from either glucose or glutamine (Fig. 2, D to G, and fig. S2, C to V). Additionally, *NADK2* deletion did not lead to changes in the mitochondrial basal oxygen consumption rate or uncoupled electron transport chain activity (fig. S2, W to Y).

Mitochondria are major sites of ROS generation in cells (*11*), and depletion of mitochondrial NADP(H) is thought to lead to oxidative stress. However, in all cell types that we tested, cells lacking *NADK2* did not display increased cellular ROS or mitochondrial superoxide (MitoSox) abundance (Fig. 2H and fig. S3, A to G). We used mitochondria-targeted redox-sensitive green fluorescent protein (roGFP2) constructs coupled to the yeast peroxidase Orp1 or human glutaredoxin-1 (Grx1) (*12*, *13*), and we measured similar amounts of mitochondrial H_2O_2 or GSH oxidation, respectively, in control and *NADK2* knockout cells (Fig. 2I and fig. S3, H to J). Treatment with

---

[1]Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [2]The Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [3]Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.
***Corresponding author. Email: thompsonc@mskcc.org**
†These authors contributed equally to this work.

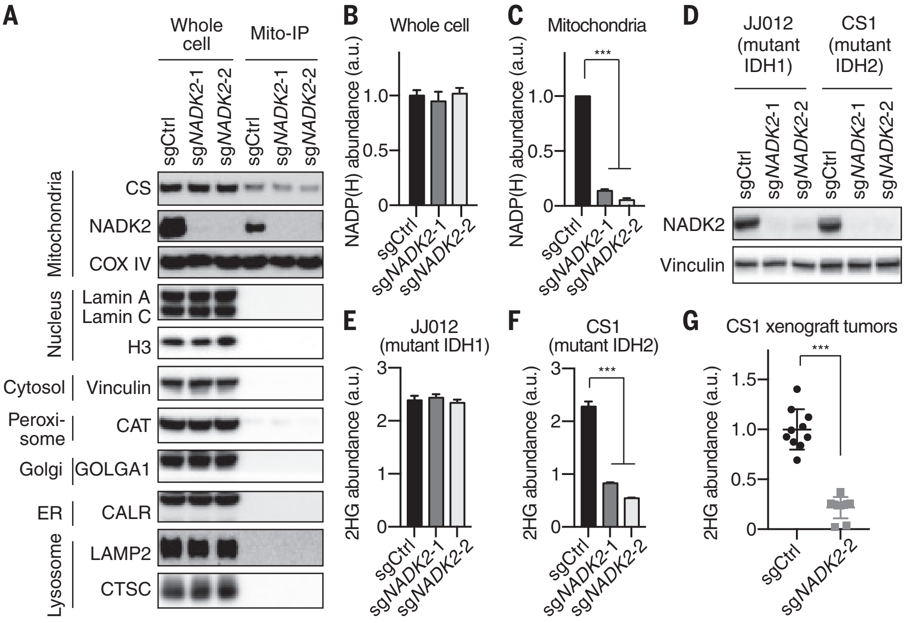


Fig. 1. NADK2 is required to maintain the mitochondrial NADP(H) pool. (**A**) DLD1 cells expressing HA-tagged OMP25 protein (DLD1-OMP25HA) were engineered to express control guide RNA (sgCtrl) or two independent guide RNA sequences targeting *NADK2* (sg*NADK2*-1 and sg*NADK2*-2) and then were subjected to Western blot of whole cells or anti-HA immunopurified mitochondria (Mito-IP). ER, endoplasmic reticulum. (**B** and **C**) Colorimetric enzyme-based measurement of total NADP(H) abundance in whole cells (B) or immunopurified mitochondria (C) of DLD1-OMP25HA cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2, cultured in DMEM/F12 medium. a.u., arbitrary units. (**D**) Western blot analysis of JJ012 (mutant IDH1) and CS1 (mutant IDH2) cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2. (**E** and **F**) 2HG abundance measured by gas chromatography–mass spectrometry (GC-MS) in JJ012 (E) and CS1 (F) cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2. (**G**) 2HG abundance measured by GC-MS in xenograft tumors formed by CS1 cells with sgCtrl or sg*NADK2*-2. Error bars in (B) represent means + SDs (*n* = 6 biological replicates); error bars in (C), (E), and (F) represent means + SDs (*n* = 3); and error bars in (G) represent means ± SDs (*n* = 10). In (C), one-way analysis of variance (ANOVA) was performed with matched measures. In (F), one-way ANOVA was performed. In (G), two-sided *t* test was performed with Welch's correction. ***$P < 0.001$.

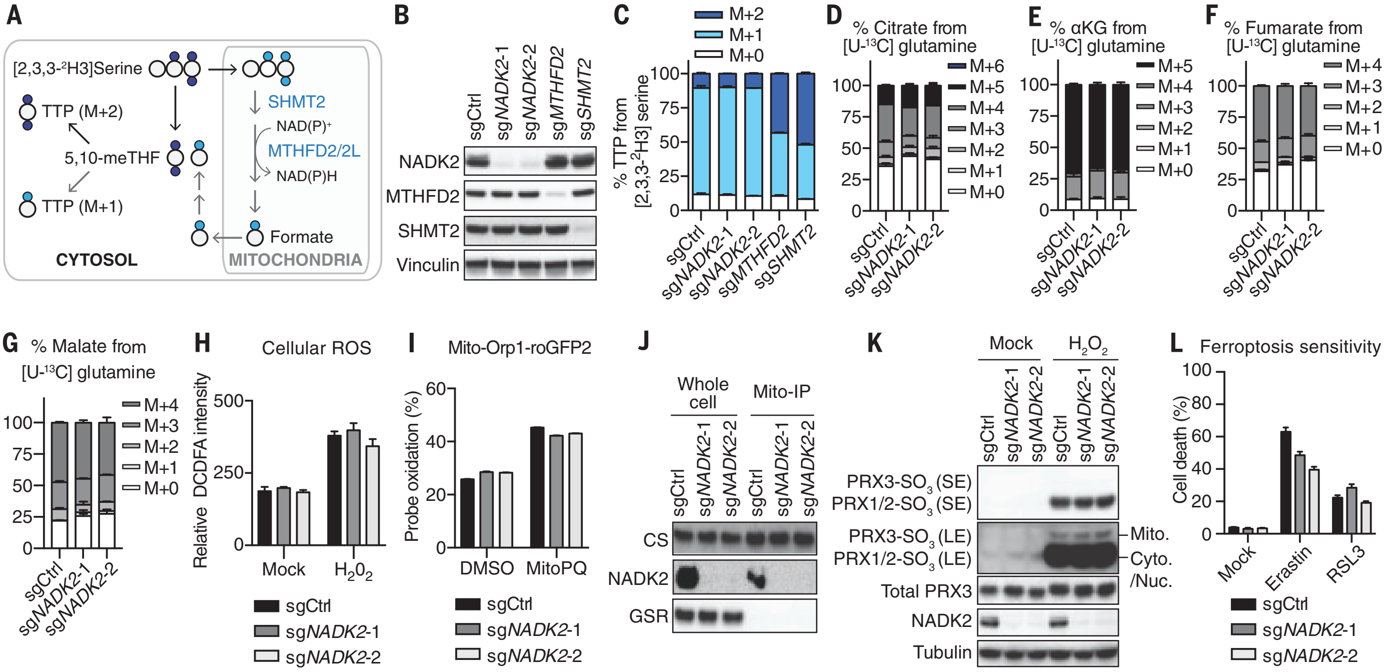


Fig. 2. Mitochondrial NADP(H) depletion does not have substantial effects on folate pathway, TCA cycle activity, or measures of oxidative stress. (**A**) Scheme of the tracing strategy, adapted from (*9*, *10*). Catabolism of [2,3,3-^{2}H3]serine in the mitochondrial or cytosolic folate pathway produces singly or doubly deuterated thymidine triphosphate (TTP M+1 or TTP M+2), respectively. (**B**) Western blot of DLD1 cells with sgCtrl, sg*NADK2*-1, sg*NADK2*-2, sg*MTHFD2*, or sg*SHMT2*. (**C**) Isotopologue distribution of TTP measured by liquid chromatography–mass spectrometry (LC-MS) in DLD1 cells denoted in (B), cultured in [2,3,3-^{2}H3]serine–containing medium for 8 hours. (**D** to **G**) Isotopologue distribution of the indicated metabolites measured by GC-MS in DLD1 cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2, cultured in [U-^{13}C]glutamine–containing medium for 6 hours. (**H**) Cellular ROS measured by CM-H$_2$DCFDA in the indicated DLD1 cells, mock treated or treated with 150 μM H$_2$O$_2$ for 4 hours. (**I**) DLD1 cells expressing Mito-Orp1-roGFP2 and the indicated guide RNA were treated with vehicle (DMSO) or 100 μM MitoPQ for 24 hours. Oxidation status was expressed as a percentage of maximal oxidation, which was determined by treating cells with 5 mM H$_2$O$_2$ for 5 min before harvest. (**J**) Western blot analysis of whole cells or immunopurified mitochondria of DLD1-OMP25HA cells expressing the indicated guide RNA. (**K**) Western blot of the indicated DLD1 cells mock treated or treated with 500 μM H$_2$O$_2$ for 6 hours. SE, short exposure; LE, long exposure. (**L**) Ferroptosis sensitivity of the indicated DLD1 cells, measured as a percentage of cell death upon mock, erastin (5 μM), or RSL3 (0.5 μM) treatment for 24 hours. All error bars in this figure represent means + SDs (*n* = 3).

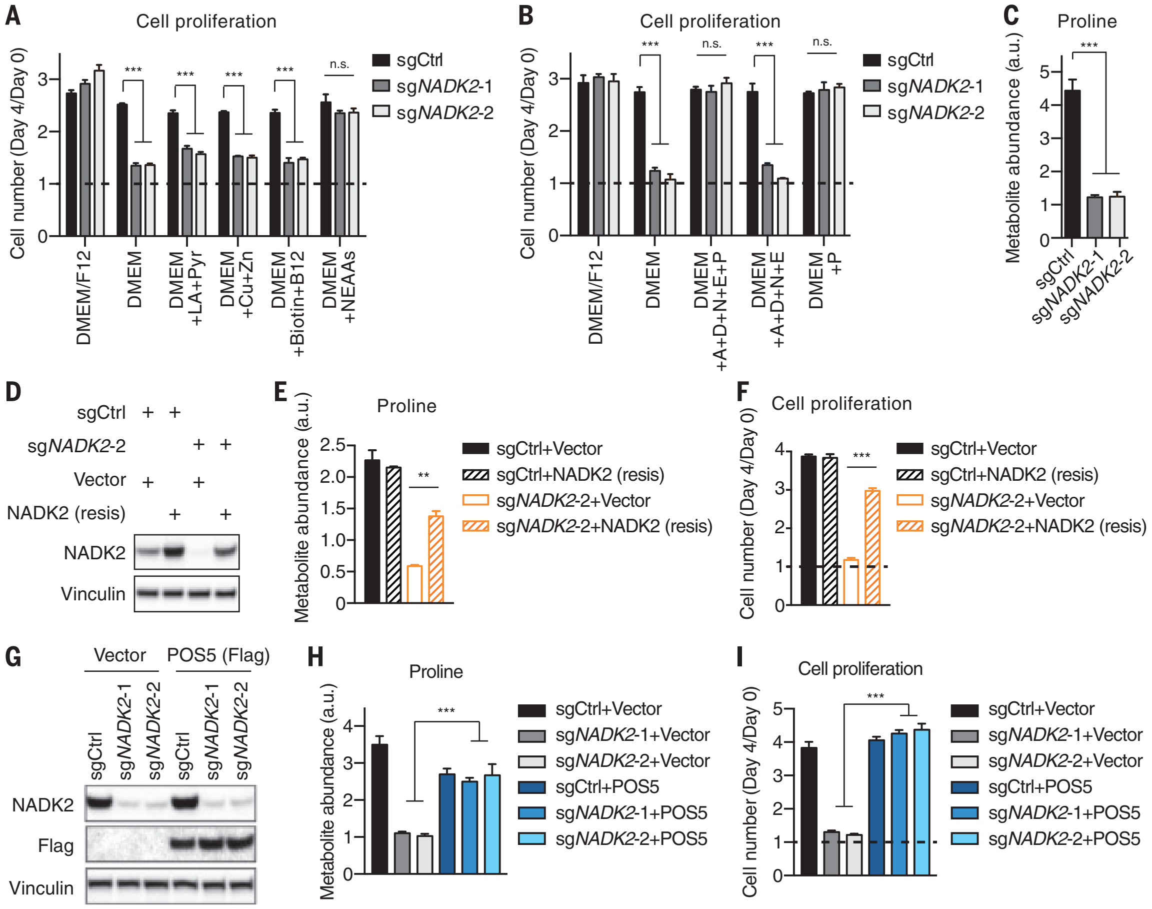


Fig. 3. Mitochondrial NADP(H) depletion results in proline auxotrophy. (**A** and **B**) Cell proliferation measured as cell number fold change (day 4/day 0) of T47D cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2, cultured in the indicated medium and supplementation. LA, lipoic acid; Pyr, pyruvate; Cu, cupric sulfate; Zn, zinc sulfate; B12, vitamin B12; A, alanine; D, aspartate; N, asparagine; E, glutamate; P, proline. All supplements are added at the concentrations present in DMEM/F12. (**C**) Proline abundance measured by GC-MS in the indicated T47D cells cultured in DMEM. (**D** to **F**) Western blot (D), proline abundance measured by GC-MS (E), and cell proliferation (F) of DMEM-cultured T47D cells with sgCtrl or sg*NADK2*-2 and ectopically expressing vector or *NADK2* cDNA resistant (resis) to sg*NADK2*-2–mediated CRISPR-Cas9 genome editing. (**G** to **I**) Western blot (G), proline abundance measured by GC-MS (H), and cell proliferation (I) of DMEM-cultured T47D cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2 and ectopically expressing vector or the POS5 cDNA. All error bars in this figure represent means + SDs ($n = 3$). In (A), (B), (C), (H), and (I), one-way ANOVA was performed. In (E) and (F), two-sided *t* test was performed with Welch's correction. $^{**}P < 0.01$; $^{***}P < 0.001$; not significant (n.s.), $P > 0.05$.

MitoParaquat (MitoPQ) increased the expression of enzymes involved in GSH synthesis to a similar extent in cells lacking *NADK2* as that observed in the control cells (*14*) (fig. S3, K and L). In agreement, loss of NADK2 did not alter cellular or mitochondrial GSH abundance or the ratio of GSH to its oxidized form, GSSG (GSH/GSSG) (fig. S3, M to P). [U-^{13}C]glutamine tracing revealed no substantial changes in the fraction of GSH or GSSG derived from glutamine upon NADK2 loss (fig. S3, Q and R). These results are consistent with the cytosolic NADP(H) pool, but not mitochondrial NADP(H), being critical for maintaining cellular GSH levels to prevent oxidative damage (*7*). Glutathione reductase (GSR) expression was absent in the mitochondrial fraction (Fig. 2J); thus, the NADPH-dependent GSH reduction appears not to take place in mitochondria.

Hyperoxidation of peroxiredoxins (PRXs-SO_3) indicates oxidative stress of the cellular thioredoxin system. We observed similar amounts of mitochondrial (PRX3) as well as cytosolic and nuclear (PRX1 and PRX2) peroxiredoxin oxidation when comparing cells lacking *NADK2* with control cells (Fig. 2K and fig. S3, S and T). Cellular and mitochondrial oxidative stress can lead to ferroptotic cell death (*15*, *16*). When treated with erastin or RSL3, chemicals that induce ferroptosis, cells lacking *NADK2* showed no increase in cell death (Fig. 2L and fig. S3U). Similarly, *Nadk2* knockout did not increase sensitivity to ferroptosis in contact-inhibited, nonproliferative mouse embryonic fibroblasts (MEFs) (fig. S3, V and W). Thus, the loss of NADK2—and the depletion of mitochondrial NADP(H)—did not increase oxidative stress under the experimental conditions we examined, although it remains possible that mitochondrial NADP(H) generation might play a role in antioxidant defense in response to other physiological perturbations.

We observed that proliferation of cells lacking *NADK2* was not perturbed compared with that of control cells when cultured in a nutrient-rich medium [Dulbecco's modified Eagle's medium and Ham's F-12 medium (DMEM/F12)] (fig. S4, A to D). However, our studies of IDH2 mutant cells indicated that NADK2 could have a role in NADPH-dependent biosynthesis (Fig. 1, F and G). To test whether mitochondrial NADP(H) supports biosynthetic reactions in general, we subjected control and *NADK2* knockout cells to culture medium composed of minimal essential nutrients (DMEM) and found that the growth of *NADK2*-deleted cells was compromised (fig. S4, A to D). Apparently, mitochondrial NADP(H) promotes the synthesis of one or more nutrients required to sustain cell proliferation.

Growth of cells lacking *NADK2* was restored in DMEM by supplementing nonessential amino acids (NEAAs), but not by other nutrients present in DMEM/F12 (Fig. 3A and fig. S4, E and F). Supplementing individual amino acids revealed that proline was both necessary and sufficient to restore proliferation of *NADK2* knockout cells in DMEM (Fig. 3B and fig. S4, G to J). In agreement, cells lacking *NADK2* showed reduced intracellular proline abundance (Fig. 3C). Similar results were obtained under hypoxia (0.5% O_2) (fig. S4, K to M). To

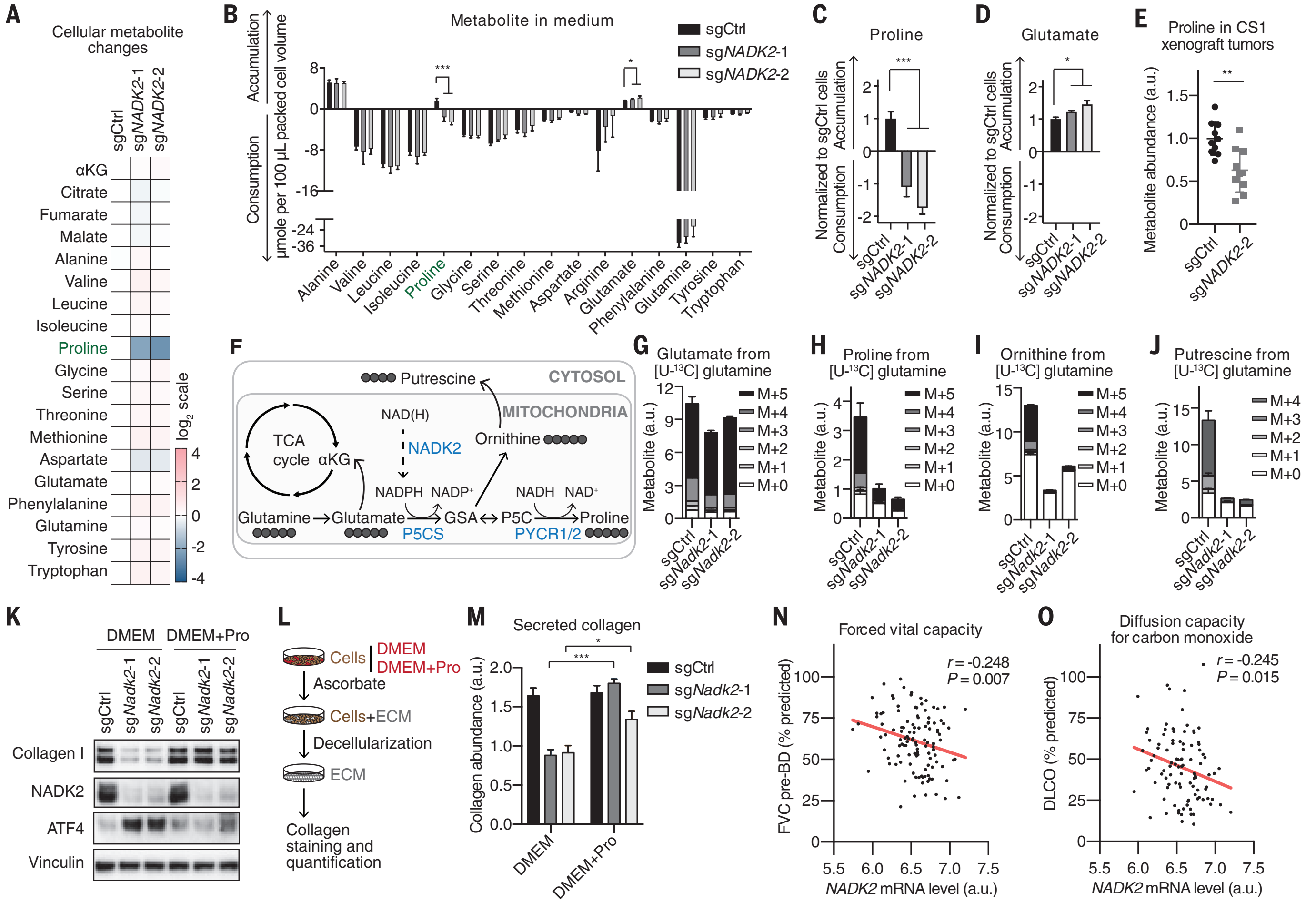


Fig. 4. The mitochondrial NADP(H) pool is required to support proline biosynthesis and collagen production. (**A**) Heatmap representing changes of metabolite levels measured by GC-MS in T47D cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2 cultured in DMEM for 48 hours. The average of three biological replicates is shown. For each metabolite, values of sg*NADK2*-1 and sg*NADK2*-2 cells are shown as $\log_2$ (fold change) relative to the value of sgCtrl cells. (**B**) Changes of metabolite levels measured by GC-MS in DMEM/F12 medium used to culture T47D cells with sgCtrl, sg*NADK2*-1, or sg*NADK2*-2 for 48 hours. (**C** and **D**) Proline (C) and glutamate (D) data from (B), replotted as normalized values to sgCtrl cells. (**E**) Proline abundance measured by GC-MS in xenograft tumors formed by CS1 cells with sgCtrl or sg*NADK2*-2. (**F**) Scheme of proline biosynthesis pathway in the mitochondria. (**G** to **J**) Relative total level and isotopologue distribution of the indicated metabolites measured by LC-MS in MEFs with sgCtrl, sg*Nadk2*-1, or sg*Nadk2*-2, cultured in DMEM containing [U-^{13}C]glutamine for 8 hours. (**K**) Western blot of the indicated MEFs, cultured in DMEM or DMEM supplemented with 300 μM proline. (**L**) Scheme of extracellular matrix (ECM) extraction and collagen staining in cells and under conditions described in (M). (**M**) Secreted collagen levels quantified by picro sirius red staining in ECM derived from MEFs with sgCtrl, sg*Nadk2*-1, or sg*Nadk2*-2, cultured for 48 hours in DMEM or DMEM supplemented with 300 μM proline, in the presence of 50 μM ascorbate. (**N**) Pearson correlation (r) of *NADK2* mRNA level and forced vital capacity (FVC) before bronchodilator (pre-BD) as percentage of what was predicted for each patient. Data are from GSE32537. (**O**) Pearson correlation of *NADK2* mRNA level and diffusing capacity for carbon monoxide (DLCO) as percentage of what was predicted for each patient. Data are from GSE32537. Error bars in (E) represent means ± SDs (n = 10). All other error bars in this figure represent means + SDs (n = 3). In (B) to (D), one-way ANOVA was performed. In (E) and (M), two-sided t test was performed with Welch's correction. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

validate that the proline-dependent growth phenotype was the result of NADK2 loss, we introduced *NADK2* cDNA resistant to CRISPR-Cas9–mediated genome editing into the *NADK2* knockout cells, which restored both intracellular proline abundance and cell growth (Fig. 3, D to F, and fig. S5, A to C). Similar results were observed when the yeast mitochondrial NAD(H) kinase, POS5 (*17*), was reconstituted in NADK2-deficient cells (Fig. 3, G to I, and fig. S5, D to F).

We performed metabolite profiling of cells lacking *NADK2* cultured in DMEM and confirmed the depletion of intracellular proline, whereas amounts of many other amino acids were slightly increased (Fig. 4A and fig. S6, A and B). Loss of NADK2 also reduced proline abundance in nonproliferating (contact-inhibited) MEFs (fig. S6, C and D). By contrast, loss of cytosolic NADK1 did not decrease proline abundance (fig. S6, E and F). Likewise, the oxygen-dependent NADPH oxidase, TPNOX (*18*), reduced proline amounts when expressed in mitochondria (mitoTPNOX) but not in cytosol (cytoTPNOX) (fig. S6, G to J). To extend these observations, we examined the consumption of nutrients from the proline-containing DMEM/F12 medium. Although we observed net proline accumulation in medium conditioned by control cells, proline was consumed by cells lacking *NADK2* (Fig. 4, B and C, and fig. S7, A to D). Additionally, glutamate accumulation was found in medium conditioned by cells lacking *NADK2* (Fig. 4, B and D, and fig. S7, A, B, E, and F), which might result from compensatory accumulation of carbon and nitrogen in the form of glutamate instead of proline. We performed similar analyses in xenograft tumors formed by CS1 cells (Fig. 1). We found that across a panel of amino acids, proline amount was reduced in tumors formed by CS1 cells lacking *NADK2* (Fig. 4E and fig. S7G), which correlated with a slower growth rate of these tumors compared with those formed by control cells (fig. S7H). Mice grafted with control or *NADK2* knockout cells displayed similar plasma levels of proline as well as other amino acids at the time of tumor resection (fig. S7I). Thus, the loss of NADK2, and the consequent depletion of mitochondrial NADP(H), results in proline auxotrophy.

Proline biosynthesis takes place in the mitochondria, where glutamine-derived glutamate is converted to pyrroline-5-carboxylate (P5C) by pyrroline-5-carboxylate synthase (P5CS). P5C is further reduced to proline by mitochondrial pyrroline-5-carboxylate reductases (PYCR1 and PYCR2) (Fig. 4F). [U-^{13}C]glutamine tracing revealed that most cellular glutamate and proline were derived from glutamine and that glutamine-derived proline was reduced upon NADK2 loss (Fig. 4, G and H, and fig. S8, A and B). By contrast, proline abundance was not perturbed when the cytosolic pyrroline-5-carboxylate reductase (PYCRL) was deleted (fig. S8, C and D).

P5CS is an NADPH-dependent enzyme, whereas PYCR1 and PYCR2 have higher affinities for NADH than for NADPH (*19*–*21*). To test whether the loss of NADK2 impairs the conversion of glutamate to P5C by P5CS, we took advantage of the fact that cellular P5C is in equilibrium with glutamate-5-semialdehyde (GSA), which can be diverted to produce ornithine for polyamine biosynthesis (Fig. 4F). Intracellular arginine can also contribute to ornithine and polyamines. Isotope tracing using [U-^{13}C] glutamine and [U-^{13}C]arginine allowed us to assess the relative contribution of these pathways to polyamine production (fig. S8E). The fraction of ornithine and putrescine derived from [U-^{13}C]glutamine decreased in cells lacking *NADK2*, which indicates that P5CS flux from glutamate to P5C and GSA was diminished (Fig. 4, I and J). This also resulted in a reciprocal increase in the proportional contribution of arginine to ornithine and putrescine (fig. S8, F to I). Because ornithine transcarbamylase expression is restricted to the liver and small intestine, loss of NADK2 did not change glutamine or arginine contribution to cellular citrulline (fig. S8, J and K). Thus, loss of NADK2 and the resulting decrease in mitochondrial NADP(H) blocks the reduction of glutamate to P5C required for proline biosynthesis.

Incorporation of the proline pyrrolidine ring slows protein translation (*22*, *23*) but endows proline-containing polypeptides with conformational rigidity. As a result, proline and its posttranslationally modified form, hydroxyproline, are abundant in collagen proteins (*24*), so a consequence of decreased mitochondrial NADP(H) generation could be impaired collagen production. Cultured mouse fibroblasts lacking *Nadk2* had decreased expression of collagen when grown in DMEM (Fig. 4K and fig. S9A). These cells accumulated activating transcription factor 4 (ATF4), indicative of amino acid shortage. The addition of 300 μM proline to the culture medium restored collagen expression and blunted ATF4 accumulation in cells lacking *Nadk2* (Fig. 4K and fig. S9, A and B). Similar results were obtained in osteosarcoma and chondrosarcoma cells that produce collagens (fig. S9, C and D). Fibroblasts lacking *Nadk2* showed decreased collagen secretion, which was rescued by proline supplementation to the medium (Fig. 4, L and M). In patients with idiopathic pulmonary fibrosis (IPF) (*25*), higher *NADK2* expression in the lung correlated with lower forced vital capacity (FVC) (P = 0.007) and diffusion capacity for carbon monoxide (DLCO) (P = 0.015)—parameters that measure maximum air exhalation and the ability of the lung to transfer air into the blood, respectively (Fig. 4, N and O). Similarly, IPF patients with both high *NADK2* and high *P5CS* expression in the lung had reduced FVC and DLCO values compared with those with low *NADK2* and low *P5CS* expression (fig. S9, E and F). Thus, increased expression of *NADK2* correlated with enhanced fibrotic diseases characterized by excessive collagen deposition.

These findings provide insights into the regulation of intracellular metabolism. In endosymbiosis with the host cell, mitochondria produce NADP(H), which supplies biosynthetic precursors to their host, and appear not to use the NADP(H) for antioxidant defense in support of their own homeostasis. Compartmentalization of cellular metabolism thus has important roles in eukaryotic cells beyond the well-known collaborative production of adenosine 5′-triphosphate (ATP).

REFERENCES AND NOTES

1. R. P. Goodman, S. E. Calvo, V. K. Mootha, *J. Biol. Chem.* **293**, 7508–7516 (2018).
2. C. A. Lewis *et al.*, *Mol. Cell* **55**, 253–263 (2014).
3. K. Ohashi, S. Kawai, K. Murata, *Nat. Commun.* **3**, 1248 (2012).
4. W. W. Chen, E. Freinkman, D. M. Sabatini, *Nat. Protoc.* **12**, 2215–2231 (2017).
5. W. W. Chen, E. Freinkman, T. Wang, K. Birsoy, D. M. Sabatini, *Cell* **166**, 1324–1337.e11 (2016).
6. G. Hoxhaj *et al.*, *Science* **363**, 1088–1092 (2019).
7. C. C. Ding *et al.*, *Nat. Metab.* **2**, 270–277 (2020).
8. A. H. Shih, O. Abdel-Wahab, J. P. Patel, R. L. Levine, *Nat. Rev. Cancer* **12**, 599–612 (2012).
9. G. S. Ducker *et al.*, *Cell Metab.* **23**, 1140–1153 (2016).
10. N. Kory *et al.*, *Science* **362**, eaat9528 (2018).
11. H. Kong, N. S. Chandel, *J. Biol. Chem.* **293**, 7499–7507 (2018).
12. M. Gutscher *et al.*, *Nat. Methods* **5**, 553–559 (2008).
13. M. Gutscher *et al.*, *J. Biol. Chem.* **284**, 31532–31540 (2009).
14. E. L. Robb *et al.*, *Free Radic. Biol. Med.* **89**, 883–894 (2015).
15. M. Gao *et al.*, *Mol. Cell* **73**, 354–363.e3 (2019).
16. B. R. Stockwell *et al.*, *Cell* **171**, 273–285 (2017).
17. C. E. Outten, V. C. Culotta, *EMBO J.* **22**, 2015–2024 (2003).
18. V. Cracan, D. V. Titov, H. Shen, Z. Grabarek, V. K. Mootha, *Nat. Chem. Biol.* **13**, 1088–1095 (2017).
19. J. De Ingeniis *et al.*, *PLOS ONE* **7**, e45190 (2012).
20. J. J. Kramer, R. C. Gooding, M. E. Jones, *Anal. Biochem.* **168**, 380–386 (1988).
21. J. M. Phang, *Antioxid. Redox Signal.* **30**, 635–649 (2019).
22. M. Y. Pavlov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 50–54 (2009).
23. F. Loayza-Puch *et al.*, *Nature* **530**, 490–494 (2016).
24. S. Schwörer *et al.*, *EMBO J.* **39**, e103334 (2020).
25. I. V. Yang *et al.*, *Thorax* **68**, 1114–1121 (2013).

ACKNOWLEDGMENTS

We thank members of the Thompson laboratory for critical discussions during manuscript preparation. We thank E. de Stanchina for assistance with tumor xenograft experiments and S. Violante for assistance with LC-MS experiments. **Funding:** J.Z. is supported by the Leukemia and Lymphoma Society (5460-18) and the NCI (1K99CA248711). S.S. is supported by the Human Frontier Science Program (LT000854/2018-L), the European Molecular Biology Organization (ALTF 467-2018), and the Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center. K.W.R. is supported by the Hunter Douglas Fellowship in Breast Cancer Research (13459) and the BRIA Postdoctoral Researcher Innovation Grant (18057). J.R.C. is supported by grants from the NIAID (R25 training grant AI140472-01A1) and the Donald B. and Catherine C. Marron Cancer Metabolism Center. X.J. is supported by grants from the NIH (R01CA204232 and R01CA166413). This work is supported by grants from the NCI (to C.B.T.) and by the cancer center support grant (P30 CA008748) to Memorial Sloan Kettering Cancer Center. **Author contributions:** J.Z. and C.B.T. conceived the study. J.Z. performed most experiments and analyzed the data, with assistance from Y.J.K. S.S. performed some of the experiments in mouse fibroblasts, analyzed IPF datasets, and optimized experimental methods used in this study. M.B. performed LC-MS measurements and analyses under the supervision of J.R.C. K.W.R. contributed key reagents and experimental methods. J.Y. and X.J. performed ferroptosis assays. J.Z. and C.B.T. interpreted the results and wrote the manuscript. All authors participated in discussing and finalizing the manuscript. **Competing interests:** C.B.T. is a founder of Agios Pharmaceuticals and a member of its scientific advisory board. He is also a former member of the Board of Directors and a former stockholder of Merck and Charles River Laboratories. He holds patents related to cellular metabolism. X.J. holds patents related to autophagy and cell death. The remaining authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/968/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 and S2
References (*26*–*28*)
MDAR Reproducibility Checklist

27 June 2020; accepted 7 April 2021
Published online 22 April 2021
10.1126/science.abd5491

QUANTUM CRITICALITY

Nematic quantum criticality in an Fe-based superconductor revealed by strain-tuning

Thanapat Worasaran[1,2]*†, Matthias S. Ikeda[1,2]†, Johanna C. Palmstrom[1,2]‡, Joshua A. W. Straquadine[1,2]§, Steven A. Kivelson[1,3], Ian R. Fisher[1,2]*

Quantum criticality may be essential to understanding a wide range of exotic electronic behavior; however, conclusive evidence of quantum critical fluctuations has been elusive in many materials of current interest. An expected characteristic feature of quantum criticality is power-law behavior of thermodynamic quantities as a function of a nonthermal tuning parameter close to the quantum critical point (QCP). Here, we observed power-law behavior of the critical temperature of the coupled nematic/structural phase transition as a function of uniaxial stress in a representative family of iron-based superconductors, providing direct evidence of quantum critical nematic fluctuations in this material. These quantum critical fluctuations are not confined within a narrow regime around the QCP but rather extend over a wide range of temperatures and compositions.

Long-range electronic nematic order, defined as electronic order that breaks only point group symmetries, is a ubiquitous feature of Fe-based superconductors [(*1*, *2*) and references therein]. For cuprate superconductors, mounting evidence points toward a generic in-plane electronic anisotropy for underdoped compositions, implying the presence of at least a nematic component to an electronic ordered state, and possibly even a vestigial nematic state [(*3*, *4*) and references therein]. From a theoretical perspective, several lines of reasoning suggest that nematic fluctuations can provide a pairing interaction (*5*–*7*) and thus that the presence of nematic order in the phase diagrams of these high-temperature superconductors may not be coincidental. In particular, nematic fluctuations enhance superconductivity in any symmetry channel (*8*) and thus could be a key element for increasing the critical temperature even if the dominant pairing interaction arises from spin fluctuations. A key open question is whether quantum critical nematic fluctuations are present and, if so, over how much of the phase diagram. Here, we directly address this latter question for a representative family of Fe-based superconductors, revealing the presence of quantum critical nematic fluctuations by power-law variation of the critical temperature of the nematic phase transition with respect to nonthermal tuning parameters.

Direct evidence for a quantum critical regime in Fe-based superconductors has been limited thus far. A divergence of the effective mass in $BaFe_2(As_{1-x}P_x)_2$ has been inferred from penetration depth, quantum oscillation, heat capacity, and resistivity measurements (*9*–*13*), suggesting the presence of a quantum critical point (QCP). These measurements cannot, however, establish the character of the fluctuating order, specifically whether it is nematic or magnetic. In comparison, nuclear magnetic resonance measurements in $NaFe_{1-x}Co_xAs$ indicate the possible presence of well-separated magnetic and nematic QCPs in that system (*14*). Measurements of the doping and temperature dependence of the nematic susceptibility for a wide variety of Fe-based

---

[1]Stanford Institute for Materials and Energy Science, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [2]Department of Applied Physics and Geballe Laboratory of Advanced Materials, Stanford University, Stanford, CA 94305, USA. [3]Department of Physics and Geballe Laboratory of Advanced Materials, Stanford University, Stanford, CA 94305, USA.

***Corresponding author. E-mail: tworasar@stanford.edu (T.W.); irfisher@stanford.edu (I.R.F.)**

†These authors contributed equally to this work. ‡Present address: National High Magnetic Field Laboratory, Los Alamos National Laboratory, Los Alamos, NM, USA. §Present address: Laboratory of Quantum Materials (QMAT), Institute of Materials (IMX), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

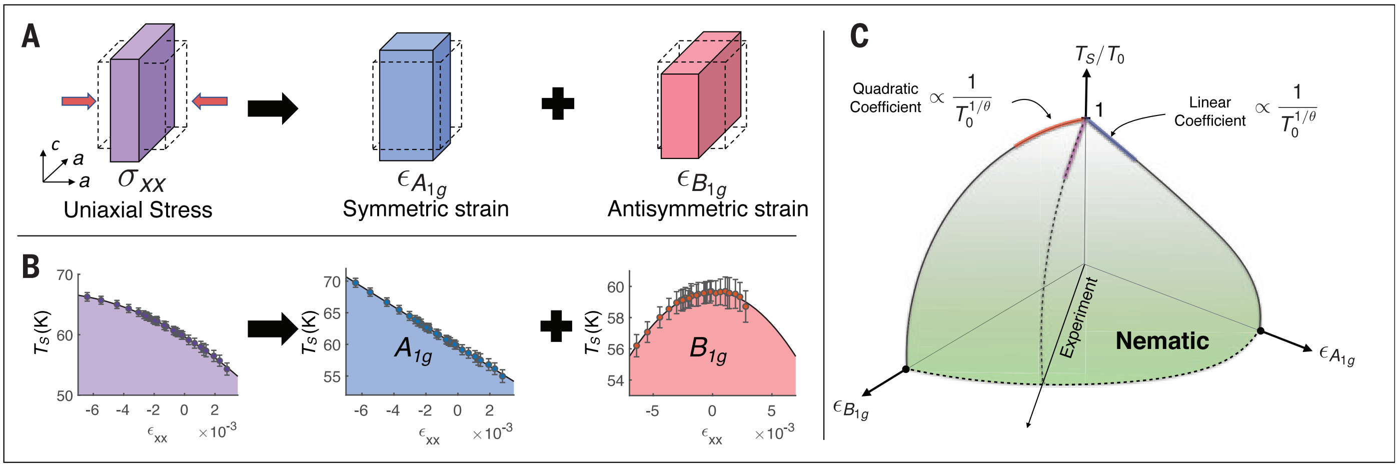


Fig. 1. Consequences of uniaxial stress along [100]. (**A**) In a tetragonal system, deformation caused by the uniaxial stress applied along the [100] direction can be expressed as the combination of the symmetry-preserving strain $\varepsilon_{A_{1g}}$ and the orthogonal antisymmetric strain $\varepsilon_{B_{1g}}$. The ratio of these strain components depends on the elastic moduli of the material. Note that the orthogonal antisymmetric B_{1g} strain is defined as $\varepsilon_{B_{1g}} = (\varepsilon_{xx} - \varepsilon_{yy})/2$, whereas the symmetry-preserving A_{1g} strain in this scenario is a combination of the in-plane $\varepsilon_{A_{1g,1}} = (\varepsilon_{xx} + \varepsilon_{yy})/2$ and the out-of-plane $\varepsilon_{A_{1g,2}} = \varepsilon_{zz}$. Dashed lines indicate the undeformed tetragonal unit cell. (**B**) Variation of coupled nematic/structural phase transition temperature as a function of the measured strain along [100], ε_{xx}, for a representative sample with Co concentration $x = 4.8 \pm 0.2\%$. The leftmost graph (purple) shows the (linear + quadratic) variation caused by uniaxial stress, the center graph (blue) shows the linear contribution from the A_{1g} strain, and the rightmost graph (red) shows the quadratic contribution from the B_{1g} strain. (**C**) Schematic $T_S - \varepsilon_{A_{1g}} - \varepsilon_{B_{1g}}$ phase diagram. The experimental path lies along the straight line between the $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$ axes. The effects of these two tuning parameters are generally uncorrelated. However, in the presence of strong quantum critical fluctuations, the coefficients of both the linear and quadratic strain responses (resulting from the response to the A_{1g} and B_{1g} strains, respectively) are both related to the zero-strain transition temperature as $1/T_0^{1/\theta}$.

superconductors, obtained through elastoresistivity (*15*, *16*), elastic constant (*17*), Raman scattering (*18*–*20*), and nuclear magnetic resonance measurements (*21*), reveal the presence of strong nematic fluctuations and thus are suggestive of the presence of a nematic QCP beneath the superconducting "dome" in these materials. Recent elastoresistivity measurements in $Ba(Fe_{1-x}Co_x)_2As_2$ for a fine comb of overdoped compositions approaching the critical doping (x_c) are consistent with a power-law divergence of the nematic susceptibility with respect to $(x - x_c)$; however, the temperature dependence of the nematic susceptibility is not currently understood in detail, so the regime over which quantum critical fluctuations extends is unknown (*22*).

Attempts to observe power-law behavior in thermodynamic quantities as a response to nonthermal tuning parameters upon approaching the putative QCP face daunting challenges. These challenges are associated either with difficulties in accurately determining the magnitude of the tuning parameters as the material is tuned infinitesimally close to the putative QCP (which is the case for chemical substitution) or in obtaining sufficient fine-tuning of the tuning parameter in that regime (which can be the case for hydrostatic pressure, another common tuning parameter). Here, we bypass these difficulties by using symmetric strain ($\varepsilon_{A_{1g}}$) and orthogonal antisymmetric strain ($\varepsilon_{B_{1g}}$) induced by in-plane uniaxial stress as essentially continuously variable tuning parameters. By doing so, we are able to show for underdoped compositions of the representative Fe-based superconductor $Ba(Fe_{1-x}Co_x)_2As_2$ that a single power law governs the rate of variation of the critical temperature of the coupled nematic/structural phase transition (T_S) with respect to both of these symmetry-inequivalent tuning parameters. This provides direct evidence of quantum critical power-law behavior in this representative material. Moreover, the perspective is widely applicable with a variety of tuning parameters and phase transitions, assuming the condition of uniform tuning parameter is achievable, establishing a useful framework to observe quantum criticality in other materials in this class and beyond.

Strain induced by external stresses breaks only point symmetries and consequently has a special role to play in the study of electronic nematicity. In recent years, considerable progress has been made by exploiting the fact that strain with the same symmetry as the nematic order acts as an effective conjugate field because of the bilinear coupling between the strain and the nematic order parameter (*15*, *23*, *24*). Here, we explore how strains of different symmetries couple to Ising nematic order and determine the shape of the phase boundary in temperature-strain space. These ideas are not specific to nematicity because they do not rely on bilinear coupling, and they could be applied to an even wider set of strain-tuned phase transitions.

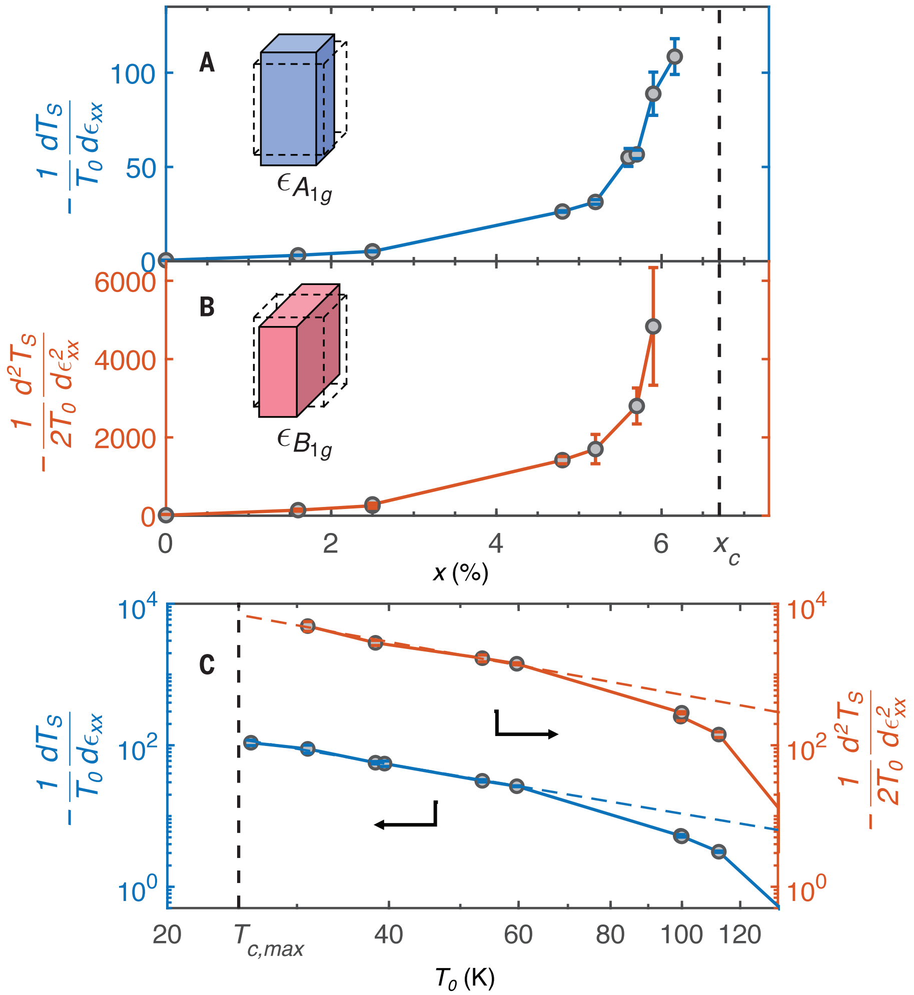


Fig. 2. Power-law divergence of $dT_S/d\varepsilon_{xx}$ and $d^2T_S/d\varepsilon_{xx}^2$ as a function of Co concentration *x* in $Ba(Fe_{1-x}Co_x)_2As_2$. (**A** and **B**) The linear coefficient of the variation of T_S caused by the symmetric A_{1g} strain (A) and the quadratic coefficient caused by the antisymmetric B_{1g} strain (B), each normalized by the zero-strain critical temperature T_0. The schematic insets illustrate the appropriate irreducible representations of strain for the two cases. Both the linear and the quadratic coefficient diverge as the Co concentration approaches the critical doping of x_C ~ 6.7% (vertical dashed line). (**C**) The same normalized coefficients as a function of zero-strain critical temperature T_0 shown in a log-log scale. The linear behavior seen in this plot at low temperatures indicates similar power-law behavior for both symmetry channels. The slopes of the linear regimes are equal to the inverse critical exponent $1/\theta$, which is found to be $1/\theta = 1.75 \pm 0.2$ and $1/\theta = 1.87 \pm 0.7$ for the A_{1g} and B_{1g} strain, respectively; these values agree within a standard deviation. Note that extracting statistically significant quadratic coefficients requires data extending to relatively large strains. For two experiments (two extra data points in the blue traces), this strain regime could not be reached before the mechanical failure of the sample, so only the linear coefficients are shown.

The representative material, $Ba(Fe_{1-x}Co_x)_2As_2$, belongs to the D_{4h} point group (corresponding to the tetragonal system) and undergoes a phase transition to an electronic nematic state with B_{2g} symmetry. Because of the coupling between the electronic and lattice degrees of freedom, the phase is also characterized by a spontaneous B_{2g} strain, $\varepsilon_{B_{2g}} \equiv \varepsilon_{xy}$. We consider the effects of strains belonging to two other irreducible representations of the point group, namely the symmetry-preserving strain $\varepsilon_{A_{1g}}$ and the orthogonal antisymmetric strain $\varepsilon_{B_{1g}}$ (Fig. 1A, center and right panels). Here, $\varepsilon_{A_{1g}}$ is a combination of in-plane ($\varepsilon_{A_{1g,1}} \equiv (\varepsilon_{xx} + \varepsilon_{yy})/2$) and out-of-plane ($\varepsilon_{A_{1g,2}} \equiv \varepsilon_{zz}$) symmetric strains, and $\varepsilon_{B_{1g}}$ is defined as $\varepsilon_{B_{1g}} \equiv (\varepsilon_{xx} - \varepsilon_{yy})/2$ (*25*). Note that the coordinates are defined in the two-iron unit cell, i.e., ε_{xx} and ε_{xy} are the deformation along the Fe–As bond and the Fe–Fe bond, corresponding to the [100] and [110] crystallographic directions, respectively. The

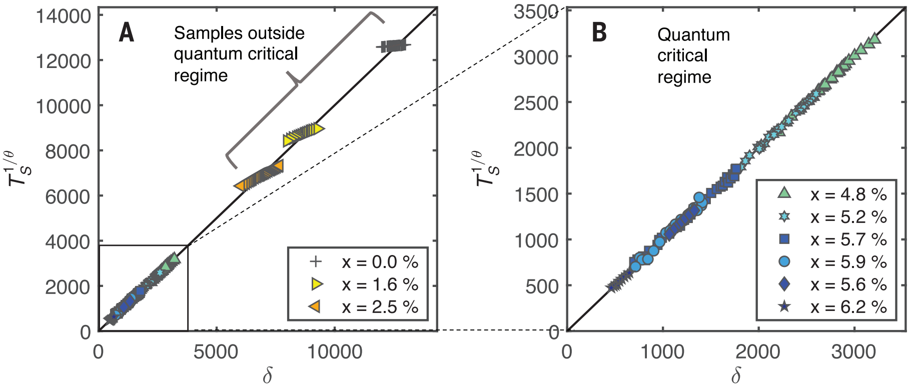


Fig. 3. Scaling collapse of nonthermal tuning parameters. (**A**) The scaling collapse described in Eq. 5, in which $T_S^{1/\theta}$ is plotted versus δ. The best-fit value θ = 0.52 ± 0.02 is extracted using the method explained in (*25*). Symbols represent samples with differing dopings. The data are expected to collapse onto a single line in the quantum critical regime, which is observed for samples within the range of Co concentrations between *x* = 4.8% and *x* = 6.2%. This regime is magnified in the plot in (**B**), which shows a collapse over almost a decade of δ.

tuning effects of the A_{1g} and B_{1g} strains are different, as has been shown previously (*26*). In the small-strain regime (which applies to this work), A_{1g} strain tunes the critical temperature T_S linearly to leading order. The orthogonal antisymmetric strain $\varepsilon_{B_{1g}}$, however, can vary the critical temperature T_S only quadratically to leading order; linear variation with respect to $\varepsilon_{B_{1g}}$ is prohibited by symmetry (*27*). Therefore, up to quadratic order, the variation of the critical temperature T_S in the presence of the $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$ strains is as follows:

$$T_S\left(\varepsilon_{A_{1g}}, \varepsilon_{B_{1g}}\right) = T_0 + a\varepsilon_{A_{1g}} + a'\varepsilon^2_{A_{1g}} + b\varepsilon^2_{B_{1g}} \qquad (1)$$

where $T_0 \equiv T_S(0,0)$ is the free-standing critical temperature and a, a', and b are coefficients to be determined. Of particular relevance to the present work, measurements performed under hydrostatic pressure conditions reveal that the coefficient a' is negligibly small for all compositions studied (*25*). Therefore, the tuning effect of $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$ can be unambiguously disentangled even in the presence of both symmetry strains. Because the A_{1g} and B_{1g} strains belong to different irreducible representations, the coefficients a and b are ordinarily anticipated to be completely independent, i.e., unrelated by any symmetry operations. As we will demonstrate below, this is no longer the case in the presence of strong quantum critical fluctuations.

Uniaxial stress was applied to $Ba(Fe_{1-x}Co_x)_2As_2$ single crystals using a commercially available strain cell (*25*) and varied in situ in an almost continuous fashion. Bar-shaped single-crystalline samples (with typical dimensions of 2000 × 500 × 35 μm, cut along the [100] direction) are glued onto two mounting plates that can be pushed or pulled by varying the voltage to lead-zirconium-titanate (PZT) piezoelectric stacks (*25*). The strain cell is designed to compensate for thermal expansion of the PZT (*28*), and because the differential thermal expansions of the cell body and the sample are negligible, the strain is almost perfectly temperature independent at a fixed voltage (*26*). Strain along the [100] crystallographic direction, ε_{xx}, can be inferred from the change in the mounting plate separation by measuring the change in capacitance of the capacitive sensor inside the cell body using a capacitance bridge (*25*). Through finite element simulations, the strain relaxation through the glue can be estimated, and the strain experienced by the sample is found to be ~70% of the measured strain (*26*). The strains $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$ are related to ε_{xx} through the elastic moduli c_{ijkl} of the samples. It can be shown (*25*) that a variation in c_{ijkl} within the composition and temperature range investigated here has no effect on the conclusions of this work. The critical temperature T_S is determined using elastoresistivity (*29*) outlined in (*25*) or, in some cases, the elastocaloric effect described in detail elsewhere (*30*).

Figure 1A illustrates schematically how the deformation stemming from the [100] uniaxial stress can be expressed as a linear combination of the symmetric and antisymmetric strains $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$. Because $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$ are both linearly proportional to ε_{xx} (*25*) and the coefficient a' is vanishingly small, the variation of T_S caused by [100] uniaxial stress is as follows:

$$T_S(\varepsilon_{xx}) = T_0 + \alpha\varepsilon_{xx} + \beta\varepsilon^2_{xx} \qquad (2)$$

where $\alpha = dT_S/d\varepsilon_{xx} \propto \partial T_S/\partial\varepsilon_{A_{1g}}$ and $\beta = 1/2 \times d^2T_S/d\varepsilon^2_{xx} \propto \partial^2 T_S/\partial\varepsilon^2_{B_{1g}}$. Figure 1B shows representative data for a sample with composition x = 4.8 ± 0.2% revealing the linear and quadratic behavior.

Further investigation of the coefficients α and β within the doping series of $Ba(Fe_{1-x}Co_x)_2As_2$ reveals a notable result. Figure 2A shows the normalized coefficients α/T_0 and β/T_0 as a function of Co concentration x. Both normalized coefficients grow monotonically and appear to diverge as the Co concentration approaches critical doping (x_C). Here, x_C is defined in the absence of superconductivity, and it was recently measured to be x_C ~ 6.7 ± 0.2% using high-magnetic fields to suppress the superconducting phase (*22*).

In what follows, we argue that a power law describing the divergence of α/T_0 and β/T_0 is a hallmark of quantum critical fluctuations. We start from the ansatz:

$$T_S \sim (g_C - g)^\theta \qquad (3)$$

This suggests that within a regime dominated by quantum critical fluctuations, the finite temperature phase transition is governed by a power law of some (relevant) nonthermal tuning parameter, g (*31*). Here, g_C is the value of g at the QCP and θ is the critical exponent governing the phase boundary in the immediate vicinity of the QCP (*32*). In the context of this work, g can be varied by varying either $\varepsilon_{A_{1g}}$ or $\varepsilon^2_{B_{1g}}$ (*33*), and we reiterate that these are distinct tuning parameters belonging to orthogonal symmetry channels.

Taylor expanding Eq. 3 in a small-strain region around zero strain and substituting $g_C \sim T_0^{1/\theta}$, we arrive at a simple expression for the leading order effect of the tuning parameter g on T_S in the quantum critical regime:

$$\frac{T_S}{T_0} = 1 - \frac{C\theta}{T_0^{1/\theta}} g + O\left(g^2\right) \qquad (4)$$

where C is a constant. Substituting $\varepsilon_{A_{1g}}$ or $\varepsilon^2_{B_{1g}}$ for g gives $\alpha/T_0 \sim -T_0^{-1/\theta}$ and $\beta/T_0 \sim -T_0^{-1/\theta}$, respectively. This is the power law that governs the morphology of the strain-tuned phase diagram in the presence of quantum critical fluctuations. Notice in particular that in the quantum critical regime, the same power law governs the behavior of both symmetry channels, even though these two strains belong to distinct irreducible representations of the point group (Fig. 1C). Notice also how the composition x does not enter into Eq. 4. All that is needed is a measurement of T_0 and of the rate of suppression of T_S with respect to $\varepsilon_{A_{1g}}$ and $\varepsilon^2_{B_{1g}}$. If the material is in the quantum critical regime tuned by x, then these two coefficients will follow the same power-law behavior.

The distinct advantages of this perspective lie in (i) the fine tunability of ε_{xx}, allowing

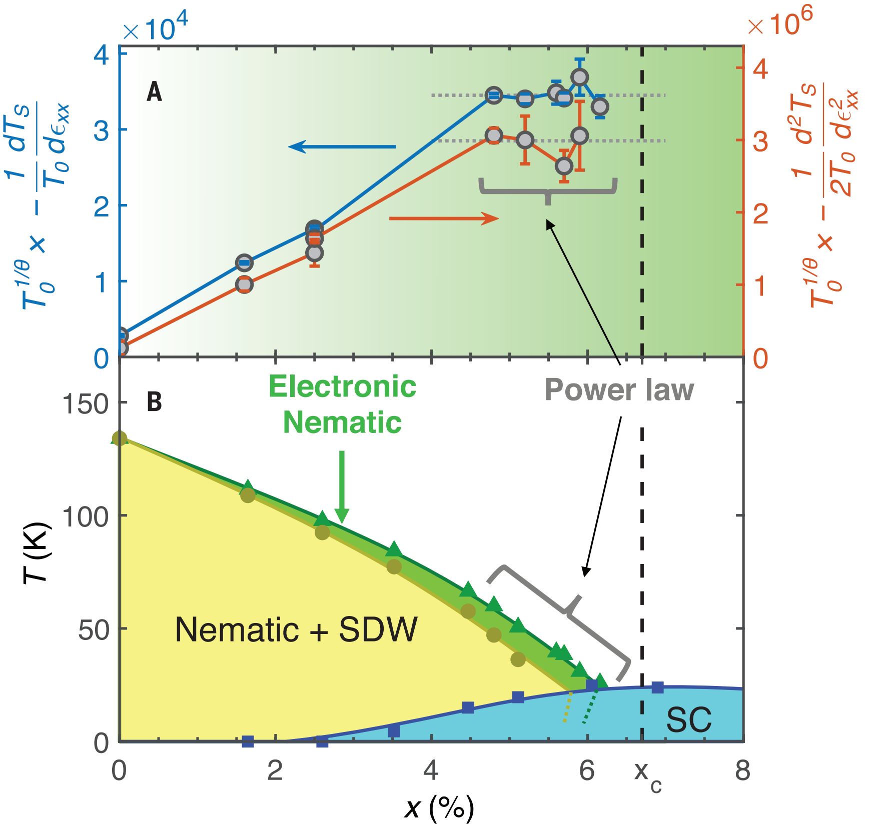


Fig. 4. Quantum criticality in Ba(Fe$_{1-x}$Co$_x$)$_2$As$_2$ phase diagram. (**A**) Normalized linear α/T_0 (left axis) and quadratic coefficients β/T_0 (right axis) scaled by $T_0^{-1/\theta}$. These quantities are expected to be constant in the quantum critical regime, which spans the higher dopings, as suggested by the dotted lines. The green shading represents schematically the ratio of these scaled coefficients with their saturated values. (**B**) Phase diagram of Ba(Fe$_{1-x}$Co$_x$)$_2$As$_2$ (*35*). Power-law behavior is observed for materials with a Co concentration between x = 4.8 ± 0.2% all the way to x = 6.2 ± 0.2%, above which T_S is unobservable because of the presence of superconductivity. This power-law behavior signifies the presence of strong nematic fluctuation of quantum origin caused by the presence of the nematic QCP at x_C. The dotted lines in (B) denote the phase transition lines that presumably backbend inside the superconducting state (*36*).

for the accurate determination of $dT_S/d\varepsilon_{xx}$ and $d^2T_S/d\varepsilon_{xx}^2$; (ii) the simultaneous determination of the effect of two tuning parameters (symmetric and antisymmetric strains) within a single experiment; (iii) the circumvention of the large uncertainties in determining the chemical composition, i.e., the determination of the chemical composition is unnecessary; and (iv) the fact that our approach completely bypasses the need to determine the critical value of the tuning parameter g_C (with respect to composition or strain) because the analysis investigates the variation of T_S with respect to the tuning parameter in the limit of small strains (this is especially useful in the present context because of the presence of the superconducting dome, but more generally it eliminates all errors that one would ordinarily incur from considering decades of variation in the quantity $\delta = g - g_C$ when both g and g_C suffer from large uncertainty).

The power-law behavior in α/T_0 and β/T_0 is visualized best as a log-log plot for which a straight line with a slope of $1/\theta$ is expected for both coefficients. This is precisely what we observed, as shown in Fig. 2C. As the zero-strain critical temperature (T_0) decreases (tuned by composition, x), there is an apparent tendency for the strain coefficients toward a linear behavior in the log-log plot, indicated by the dashed lines. The crossover regime appears at a similar value of T_0 for both α/T_0 and β/T_0. Moreover, the slopes of the best fits that represent the inverse critical exponents appear to be similar in both cases, with the value $1/\theta$ = 1.75 ± 0.2 for α/T_0 and $1/\theta$ = 1.87 ± 0.7 for β/T_0, agreeing within experimental error. Averaging the two exponents obtained from the temperature dependence of the linear and quadratic coefficients, we obtain θ = 0.56 ± 0.07.

Equivalently, this can also be illustrated by a scaling collapse with respect to the nonthermal tuning parameters. As a function of the three tuning parameters x, $\varepsilon_{A_{1g}}$, and $\varepsilon_{B_{1g}}$, the critical temperature T_S varies in the quantum critical regime as follows:

$$T_S\left(x, \varepsilon_{A_{1g}}, \varepsilon_{B_{1g}}\right) = \delta^\theta = \left[T_0(x)^{1/\theta} + A_1\varepsilon_{A_{1g}} + A_2\varepsilon_{B_{1g}}^2\right]^\theta \qquad (5)$$

where $T_0(x) \equiv T_S(x,0,0)$ and A_1 and A_2 are constants. With appropriate fit values of A_1, A_2, and θ, the data points should collapse into a single line inside the quantum critical regime. This is shown in Fig. 3, where a linear relationship of δ and $T_S^{1/\theta}$ can be seen in strain-tuned samples with T_S ranging from ~25 K up to ~70 K. Using this method, θ = 0.52 ± 0.02 provides the best collapse, agreeing with the value obtained from the power-law analysis [see (*25*) for details].

This value of θ implies that the x derivative of T_S diverges as x approaches x_C, which is consistent with the shape of the phase diagram (Fig. 4B). Presumably, this value of θ relates to universal exponents of a universality class appropriate for metallic Ising nematic systems with disorder and as such provides a test for future theoretical treatments.

It is worth noting that the measured value of θ is equal within experimental error to the theoretically predicted value (θ = 0.5) for a zero-temperature structural transition in a (Fermi liquid) metal (*34*). However, although considerations of universality suggest that this observation may be pertinent asymptotically close to the putative QCP, in most or all of the parameter range covered by our experiments, the nematic fluctuations are primarily a property of the metallic degrees of freedom. In particular, a variety of experiments, including elastoresistivity (*15*, *16*), elastic stiffness measurements (*17*), and Raman scattering (*18*–*20*), have shown that the driving force for the nematic transition comes primarily from the low-energy metallic degrees of freedom and that the response of the lattice is mostly parasitic.

Ideally, power-law behavior is established over decades of a tuning parameter. In the present case, however, the presence of superconductivity at $T_{C,\max}$ ~ 25 K limits the accessible parameter range for our investigation. Nevertheless, the observation of such closely similar behavior for the two fully independent tuning parameters $\varepsilon_{A_{1g}}$ and $\varepsilon_{B_{1g}}$, and the scaling collapse shown in Fig. 3, provide compelling evidence that the apparent power-law behavior is driven by critical fluctuations.

The power-law variation of $\alpha/T_0 \sim -1/T^{1/\theta}$ and $\beta/T_0 \sim -1/T^{1/\theta}$ implies that the normalized quantities $T^{1/\theta} \times \alpha/T_0$ and $T^{1/\theta} \times \beta/T_0$ should be constant so long as the material is in the quantum critical regime. This behavior is verified in Fig. 4A, which shows the variation of the normalized quantities as a function of composition, x.

This figure makes clear two very important points. First, the composition x does not need to be accurately determined to deduce that the material is in the quantum critical regime; it is sufficient solely that the normalized quantities have constant values, and the absolute value of the x-coordinates in Fig. 4A does not matter in reaching this conclusion. Second, the regime of power law exists over a substantially wide regime of composition and temperature (gray brackets in Fig. 4, A and B), extending over most of the superconducting dome for these underdoped compositions. The immediate implications for superconductivity are unknown. However, these experimental results empirically establish that the superconductor is not only born out of a metal that hosts strong nematic fluctuations [as has been previously inferred from a wide variety of measurements (*15*, *16*, *18*–*20*)] but also more specifically, born from a metal that exhibits quantum critical nematic fluctuations. Put another way, for a wide range of compositions in the $x - T$ plane, the material "knows" how far it is from the nematic QCP tuned by chemical composition, and this approximately correlates with the range of compositions over which the material superconducts.

REFERENCES AND NOTES

1. R. M. Fernandes, J. Schmalian, *Supercond. Sci. Technol.* **25**, 084005 (2012).
2. Q. Si, R. Yu, E. Abrahams, *Nat. Rev. Mater.* **1**, 16017 (2016).
3. L. Nie, G. Tarjus, S. A. Kivelson, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7980–7985 (2014).
4. S. Mukhopadhyay *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13249–13254 (2019).
5. M. A. Metlitski, D. F. Mross, S. Sachdev, T. Senthil, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 115111 (2015).
6. T. A. Maier, D. J. Scalapino, *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 174510 (2014).
7. S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4905–4910 (2017).
8. S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, *Phys. Rev. Lett.* **114**, 097001 (2015).
9. K. Hashimoto *et al.*, *Science* **336**, 1554–1557 (2012).
10. H. Shishido *et al.*, *Phys. Rev. Lett.* **104**, 057008 (2010).
11. P. Walmsley *et al.*, *Phys. Rev. Lett.* **110**, 257002 (2013).
12. J. G. Analytis *et al.*, *Nat. Phys.* **10**, 194–197 (2014).
13. T. Shibauchi, A. Carrington, Y. Matsuda, *Annu. Rev. Condens. Matter Phys.* **5**, 113–135 (2014).
14. C. G. Wang *et al.*, *Phys. Rev. Lett.* **121**, 167004 (2018).
15. J.-H. Chu, H.-H. Kuo, J. G. Analytis, I. R. Fisher, *Science* **337**, 710–712 (2012).
16. H.-H. Kuo, J.-H. Chu, J. C. Palmstrom, S. A. Kivelson, I. R. Fisher, *Science* **352**, 958–962 (2016).
17. A. E. Böhmer *et al.*, *Phys. Rev. Lett.* **112**, 047001 (2014).
18. P. Massat *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9177–9181 (2016).
19. Y. Gallais *et al.*, *Phys. Rev. Lett.* **111**, 267001 (2013).
20. V. K. Thorsmølle *et al.*, *Phys. Rev. B* **93**, 054515 (2016).
21. T. Kissikov *et al.*, *Phys. Rev. B* **96**, 241108 (2017).
22. J. C. Palmstrom *et al.*, Comparison of temperature and doping dependence of nematic susceptibility near a putative nematic quantum critical point. arXiv:1912.07574 [cond-mat-str-el] (16 December 2019).
23. M. C. Shapiro, P. Hlobil, A. T. Hristov, A. V. Maharaj, I. R. Fisher, *Phys. Rev. B Condens. Matter Mater. Phys.* **92**, 235147 (2015).
24. A. E. Böhmer, C. Meingast. *C. R. Phys.* **17**, 90–112 (2016).
25. Materials and methods are available as supplementary materials.
26. M. S. Ikeda *et al.*, *Phys. Rev. B* **98**, 245133 (2018).
27. In a tetragonal system, positive and negative B_{1g} strain have the same effect on the material; therefore, T_S can only be an even function of the B_{1g} strain and thus is quadratic to leading order.
28. C. W. Hicks, M. E. Barber, S. D. Edkins, D. O. Brodsky, A. P. Mackenzie, . *Rev. Sci. Instrum.* **85**, 065003 (2014).
29. A. T. Hristov *et al.*, *Rev. Sci. Instrum.* **89**, 103901 (2018).
30. M. S. Ikeda *et al.*, *Rev. Sci. Instrum.* **90**, 083902 (2019).
31. S. Sachdev, *Quantum Phase Transitions* (Cambridge Univ. Press, ed. 2, 2011).
32. In cases where a scaling hypothesis holds, the critical temperature is set by an energy scale that is inversely proportional to the critical correlation time $\tau_C \sim (g_C - g)^{-\nu z}$, where ν is the critical exponent for the correlation length and z is the dynamical scaling exponent. In such a case, $\theta = \nu z$ (*31*), but we persist in writing the exponent simply as θ to preserve generality.
33. Positive and negative B_{1g} strains suppress the critical temperature equally, hence the square on $\varepsilon_{B_{1g}}$.
34. M. Zacharias, I. Paul, M. Garst, *Phys. Rev. Lett.* **115**, 025703 (2015).
35. J.-H. Chu, J. G. Analytis, C. Kucharczyk, I. R. Fisher, *Phys. Rev. B* **79**, 014506 (2009).
36. S. Nandi *et al.*, *Phys. Rev. Lett.* **104**, 057006 (2010).
37. Data for: T. Worasaran *et al.*, Nematic quantum criticality in an Fe-based superconductor revealed by strain-tuning. Stanford Digital Repository (2021). https://doi.org/10.25740/wg432jy0214.

ACKNOWLEDGMENTS

We thank S. Raghu and P. Walmsley for many insightful discussions. **Funding:** This work was supported by the Department of Energy, Office of Basic Energy Sciences, under contract no. DEAC02-76SF00515. M.S.I. was supported in part by the Gordon and Betty Moore Foundation's EPiQS Initiative through grant no. GBMF4414. J.C.P. was supported by a Gabilan Stanford Graduate Fellowship and a Stanford Lieberman Fellowship. J.A.W.S. acknowledges support as an ABB Stanford Graduate Fellow. **Author contribution:** T.W., M.S.I., S.A.K., and I.R.F. designed the overall concept for the experiment. T.W., M.S.I., and I.R.F. developed the experimental methodology. M.S.I. and J.C.P. grew the samples. T.W., M.S.I., and J.A.W.S. developed the instrumentation. T.W. and M.S.I. performed the measurements. T.W., M.S.I, I.R.F., and S.A.K. interpreted the data and formalized the data analysis. T.W. prepared the original draft. All authors contributed extensively to reviewing and editing the manuscript. The work was supervised by S.A.K. and I.R.F. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data for this report are available at the Stanford Digital Repository (*37*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/973/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
References (*38*–*42*)

31 March 2020; accepted 17 April 2021
10.1126/science.abb9280

CORAL REEFS

Local conditions magnify coral loss after marine heatwaves

Mary K. Donovan[1,2]*, Deron E. Burkepile[2,3], Chelsey Kratochwill[4], Tom Shlesinger[4], Shannon Sully[4], Thomas A. Oliver[5], Gregor Hodgson[6], Jan Freiwald[6,7], Robert van Woesik[4]

Climate change threatens coral reefs by causing heat stress events that lead to widespread coral bleaching and mortality. Given the global nature of these mass coral mortality events, recent studies argue that mitigating climate change is the only path to conserve coral reefs. Using a global analysis of 223 sites, we show that local stressors act synergistically with climate change to kill corals. Local factors such as high abundance of macroalgae or urchins magnified coral loss in the year after bleaching. Notably, the combined effects of increasing heat stress and macroalgae intensified coral loss. Our results offer an optimistic premise that effective local management, alongside global efforts to mitigate climate change, can help coral reefs survive the Anthropocene.

Climate change is increasing the intensity and frequency of disturbances such as droughts and heatwaves (*1*), which are occurring along a backdrop of local stressors. From forests to coral reefs, these climate-driven disturbances often result in mass mortalities of foundation species (*2*). Alongside these global stressors, local stressors can also have persistent negative effects on ecosystems and can further exacerbate the mortality caused by climate-driven disturbances. For example, outbreaks of insect pests in forests can compound drought-related mortality in trees (*3*), and overfishing of predators in salt marshes can lead to population explosions of herbivorous snails, resulting in runaway consumption of marsh plants during heatwaves (*4*). The interaction between these global and local stressors suggests that effective management of local factors could lead to less-consequential impacts to ecosystems from climate-driven disturbances.

Coral reefs are sentinel ecosystems that are sensitive to both global and local disturbances (*5*, *6*). Climate change is causing marine heatwaves—periods of anomalously high seawater temperature—that induce corals to bleach. Coral bleaching is a result of the breakdown of the mutualism between corals

---

[1]Center for Global Discovery and Conservation Science and School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85281, USA. [2]Marine Science Institute, University of California, Santa Barbara, CA 93106, USA. [3]Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106, USA. [4]Institute for Global Ecology, Florida Institute of Technology, Melbourne, FL 32901, USA. [5]Ecosystem Sciences Division, Pacific Islands Fisheries Science Center, NOAA Fisheries, Honolulu, HI 96818, USA. [6]Reef Check Foundation, Marina del Rey, CA 90292, USA. [7]Institute of Marine Sciences, University of California, Santa Cruz, CA 95064, USA.
***Corresponding author. Email: marydonovan@asu.edu**

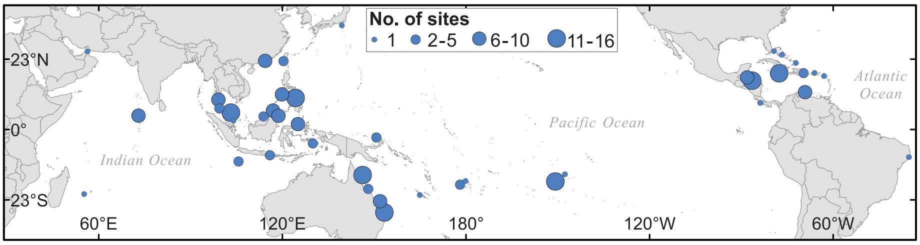


Fig. 1. Location of 223 study sites. Observations included sites from the global Reef Check dataset with coral cover both during a coral bleaching event and in the year after the bleaching event. Sites within 100 km of each other are plotted as one point, with the size of the point correlated to the number of sites.

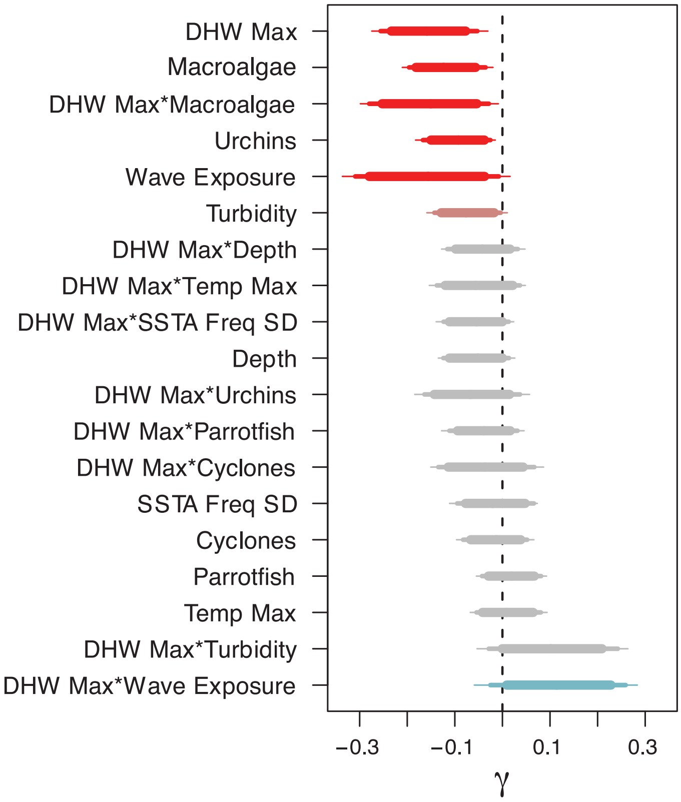


Fig. 2. Coefficients (γ) from a Bayesian hierarchical model estimating predictors of change in coral cover in the year after bleaching. Thin, medium, and thick lines represent 95, 90, and 80% credible intervals, respectively. Coefficients with negative effects are colored dark red when 90% intervals do not overlap zero, or light red when 80% intervals do not overlap zero. Coefficients with positive effects are colored light blue when 80% intervals do not overlap zero. All other effects are colored gray. Posterior distributions are plotted in fig. S2. DHW Max is the maximum degree heating weeks (a key metric of heat stress that is known to predict coral bleaching) in the 365 days preceding bleaching. Temp Max is the maximum temperature over the climatology from 1982 to 2017. SSTA Freq SD is the standard deviation of the frequency of temperature anomalies over the same climatology.

and their endosymbiotic algae of the family Symbiodiniaceae (*7*). Extreme heatwaves cause extensive coral bleaching, resulting in widespread coral mortality and changes in reef communities (*8*). These mass coral mortality events have damaged reefs globally, and many studies have concluded that reducing global carbon emissions is the only route to conserving coral reefs (*5*, *9*, *10*). Yet local factors such as nutrient pollution (*11*) can exacerbate coral mortality associated with marine heatwaves. Identifying the local stressors that magnify coral mortality after marine heatwaves may help facilitate management strategies that boost the resistance and resilience of coral reefs to climate change (*6*, *12*).

We used a Bayesian hierarchical model to analyze the trajectories of coral cover in the year after heatwaves at 223 sites worldwide (Fig. 1). Using a suite of data on abiotic (e.g., depth, wave exposure) and biotic factors (e.g., macroalgal cover, sea urchin abundance), we investigated the effect of each factor, and its interaction with heat stress, on change in absolute coral cover. We show that, along with heat stress, key abiotic factors such as wave exposure and turbidity influenced coral mortality. We also show that increasing abundance of either sea urchins or macroalgae increased coral mortality after coral bleaching.

Not surprisingly, our results show that higher heat stress increased coral loss in the year after bleaching events (Fig. 2), consistent with other studies showing that greater heat stress results in extensive coral mortality (*5*, *13*). Our analyses show that this pattern of increasing coral mortality with increasing heat stress [ranging from 0 to 18.7 degree heating weeks (DHW), a key metric of heat stress that is known to predict coral bleaching] was robust across 26 geographic regions spanning the globe. This relationship has previously been shown at regional scales, such as within the Great Barrier Reef or Caribbean Sea (*5*, *13*), and in association with El Niño events (*14*). However, we also found interactions between heat stress and local conditions, such as macroalgae (Fig. 2), highlighting the context dependency of increasing heat stress on coral mortality.

We found that reefs with higher initial macroalgal cover experienced greater coral mortality after coral bleaching (Fig. 2). The relationship between macroalgal cover and coral loss was more negative at higher heat stress (Fig. 3A). Even at similar levels of heat stress (e.g., 4 DHW), reefs with more macroalgae experienced 10 times higher coral mortality (fig. S1). For reefs with low levels of macroalgae (<5%), there was minimal mortality even after relatively extreme heat stress (>12 DHW). There are several nonexclusive mechanisms whereby abundant macroalgae could directly lead to increased coral loss when combined with heat stress. First, macroalgae can exude dissolved organic carbon (DOC) into their environment, and exposure to excess DOC can induce coral bleaching (*15*). Algal DOC also causes areas of hypoxia on corals during coral-algal competition, often causing mortality of coral tissue (*16*). Furthermore, the microbiomes of corals exposed to macroalgae have fewer beneficial bacteria and more virulent pathogenic bacteria (*17*). These effects on the coral microbiome may contribute to a rise in coral diseases after heatwaves, leading to further coral mortality (*17*, *18*). Additionally, direct contact with macroalgae can cause coral mortality via mechanisms such as allelopathy, in the form of chemically mediated bleaching of coral tissue (*19*). Macroalgae can hinder coral recovery by limiting reproductive output (*20*), reducing coral recruitment (*21*), and increasing post-settlement mortality (*22*), thereby intensifying initial declines in coral. Higher ocean temperatures likely exacerbate many of these mechanisms, which may have contributed to the stronger relationship between the abundance of macroalgae and coral loss at higher levels of heat stress (Fig. 3A).

Macroalgae can become abundant on coral reefs for many reasons, among which abundant nutrients and reduced herbivory are important drivers (*23*, *24*). Thus, some of the signal we show relating increased macroalgal abundance with increased coral mortality in the wake of heat stress may come as a consequence of other stressors associated with increased macroalgae. For example, increased nitrogen availability can impair the physiological relationship between corals and their symbiotic dinoflagellates (*25*), increase the severity of coral bleaching (*26*), and increase coral mortality after heatwaves (*11*). Abundant macroalgae may also be a signal of reduced herbivory from exploitation of herbivorous fishes (*6*). Not only do herbivorous fishes directly reduce the cover of algae, but they also play a role in facilitating coral recruitment (*21*) and thereby contribute to recovery after coral mortality. We did not find a direct relationship between change in coral cover and herbivorous parrotfish abundance (Fig. 2), which could be due to the indirect nature of the relationship between fishes and corals. Ultimately, there are multiple candidate drivers of macroalgal abundance that cannot be differentiated in our analyses that may have direct or indirect effects on the mortality and recovery of corals after heat stress events.

A higher density of urchins was also associated with increased coral mortality after bleaching (Figs. 2 and 3B). This result may seem surprising given that on many reefs an increasing abundance of urchins, which are important herbivores and bioeroders, is associated with a decrease in macroalgae and an increase in corals (*27*). However, urchins can become so abundant on some reefs, especially those that are heavily fished (*28*), that they suppress corals via direct predation and bioerosion of the reef matrix (*29*). In our dataset, urchin abundance varied by nearly four orders of magnitude. Reefs having higher densities of urchins, up to 1000 urchins per 100 m^2, often had negative coral trajectories in the year after coral bleaching (Fig. 3B). However, reefs with more-modest urchin densities (<18 per 100 m^2) often had positive trajectories in the year after bleaching events. Notably, although urchins often determine macroalgal cover (which also affects coral mortality after bleaching) at the local scale, they were not correlated at the global scale of our analyses (fig. S3). Thus, the context-dependent role of urchins on coral mortality after bleaching suggests that local intervention to prevent extreme urchin densities, such as reducing fishing on their predators, may help mitigate coral mortality associated with heat stress.

In addition to heat stress, we found modest evidence that several other abiotic factors were related to change in coral cover (Fig. 2), including wave exposure and turbidity. Reefs more exposed to waves lost more coral after bleaching events than more-sheltered reefs, except during the highest levels of heat stress (Fig. 3C). Bleaching and mortality may be lower on sheltered reefs owing to lower physical disturbance, the abundance of more-resistant species (*30*), or acclimation to higher temperatures (*31*). Sheltered reefs may also experience more-variable temperatures than exposed reefs, leading to higher thermal tolerance (*32*). We also found modest evidence for a negative effect of turbidity on coral mortality after bleaching (Fig. 2) and a tendency toward a positive effect of heat stress and turbidity together, which agrees with results from a recent study showing that corals may bleach less overall on turbid reefs during heat stress (*33*).

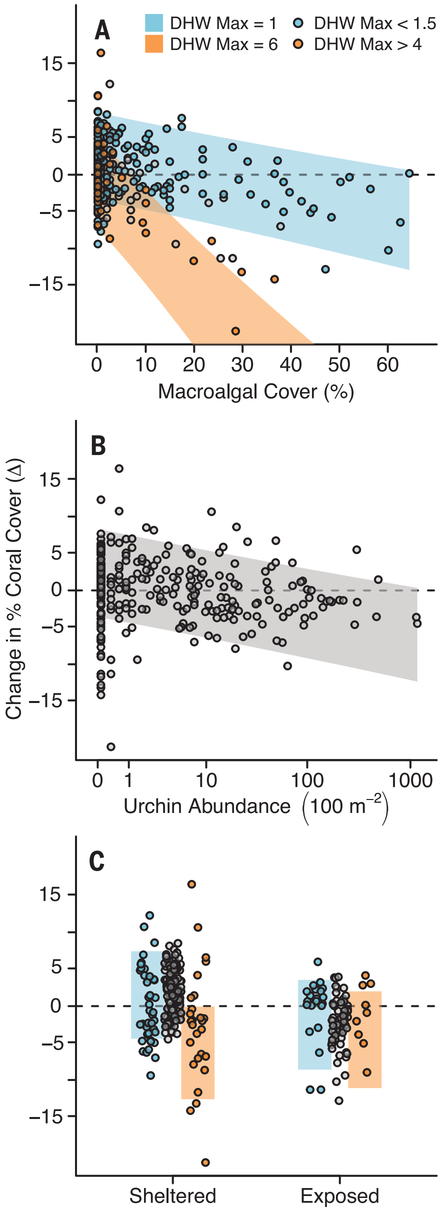


Fig. 3. Relationship between change in coral cover (Δ) over a 1-year time interval and predictors of change. (**A**) Macroalgal cover, (**B**) urchin abundance (note the log axis), and (**C**) wave exposure. Points are posterior estimates of β from equation 5 in the materials and methods (see supplementary materials). *Y* axis was back-transformed by regressing β from equation 5 against Δ in equation 4 (see supplementary materials) to reflect units of absolute change in coral cover. Shaded areas are 50% credible intervals for the predicted relationships across all regions. Given evidence for interactions with DHW Max (the 80% interval did not overlap zero), points and predicted relationships in (A) and (C) are colored. To represent mild heat stress, points are colored blue where DHW Max is <1.5, and shaded areas are blue for the relationship when DHW Max is held at 1. To represent high heat stress, points are colored orange where DHW Max is >4, and the shaded area is orange for DHW Max held at 6. All other points are colored gray.

Understanding the factors that shape coral mortality in response to heatwaves will help inform management strategies that may improve the future trajectories of coral reefs. Our results suggest that abundant macroalgae are a strong predictor of whether corals survive bleaching events and that this effect is even stronger when heat stress is moderate to high. Therefore, given that local stressors such as nutrient pollution and overfishing of herbivorous fishes can lead to increases in macroalgae on reefs (*23*, *24*), reducing these stressors may provide an opportunity to reduce coral morality, especially during moderate levels of heat stress. For example, reductions in fishing of herbivorous fishes or in nutrient pollution can lead to declines in macroalgal abundance of 10 to 40% (*34*, *35*). Our analyses suggest that such reductions in macroalgal abundance could result in substantial mitigation of coral mortality under moderate heat stress (4 DHW). These types of interventions to alleviate local stressors may be key to the persistence of coral reefs given that mild-to-moderate heatwaves may occur annually in the near future (*36*). Clearly, some heatwaves will be extreme, such as what the Great Barrier Reef experienced in 2017 (*37*), overwhelming any effect of local interventions. Ultimately, combining our results with predictions of where and when both moderate and extreme heat stress events are likely to occur could prioritize decisive action at local scales (*36*, *38*).

Our results suggest that an opportunity exists to facilitate coral persistence in the face of climate-driven heatwaves by making effective local management decisions. Of course, local management is no substitute for concerted global action to limit greenhouse gas emissions and slow climate change. Yet we can take a both/and approach, strategically using local conservation actions that help ecosystems resist and recover from climate-driven disturbances while also working toward global reductions in carbon emissions.

REFERENCES AND NOTES

1. E. M. Fischer, U. Beyerle, R. Knutti, *Nat. Clim. Chang.* **3**, 1033–1038 (2013).
2. A. M. Ellison *et al.*, *Front. Ecol. Environ.* **3**, 479–486 (2005).
3. J. D. Lauder, E. V. Moran, S. C. Hart, *Tree Physiol.* **39**, 1071–1085 (2019).
4. B. R. Silliman, J. van de Koppel, M. D. Bertness, L. E. Stanton, I. A. Mendelssohn, *Science* **310**, 1803–1806 (2005).
5. T. P. Hughes *et al.*, *Nature* **543**, 373–377 (2017).
6. R. Steneck *et al.*, *Front. Mar. Sci.* **6**, 265 (2019).
7. P. W. Glynn, *Coral Reefs* **12**, 1–17 (1993).
8. Y. Loya *et al.*, *Ecol. Lett.* **4**, 122–131 (2001).
9. J. F. Bruno, I. M. Côté, L. T. Toth, *Ann. Rev. Mar. Sci.* **11**, 307–334 (2019).

10. C. M. Eakin, H. P. A. Sweatman, R. E. Brainard, *Coral Reefs* **38**, 539–545 (2019).
11. D. E. Burkepile *et al*., *Ecosystems* **23**, 798–811 (2020).
12. A. Abelson, *ICES J. Mar. Sci.* **77**, 40–45 (2020).
13. C. M. Eakin *et al*., *PLOS ONE* **5**, e13969 (2010).
14. D. C. Claar, L. Szostek, J. M. McDevitt-Irwin, J. J. Schanze, J. K. Baum, *PLOS ONE* **13**, e0190957 (2018).
15. C. Pogoreutz *et al*., *Glob. Change Biol.* **23**, 3838–3848 (2017).
16. K. Barott *et al*., *PLOS ONE* **4**, e8043 (2009).
17. J. R. Zaneveld *et al*., *Nat. Commun.* **7**, 11833 (2016).
18. M. E. Brandt, J. W. McManus, *Ecology* **90**, 2859–2867 (2009).
19. D. B. Rasher, M. E. Hay, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9683–9688 (2010).
20. J. E. Tanner, *J. Exp. Mar. Biol. Ecol.* **190**, 151–168 (1995).
21. S. N. Arnold, R. S. Steneck, P. J. Mumby, *Mar. Ecol. Prog. Ser.* **414**, 91–105 (2010).
22. I. B. Kuffner *et al*., *Mar. Ecol. Prog. Ser.* **323**, 107–117 (2006).
23. K. E. Fabricius, *Mar. Pollut. Bull.* **50**, 125–146 (2005).
24. D. E. Burkepile, M. E. Hay, *Ecology* **87**, 3128–3139 (2006).
25. L. Ezzat, J.-F. Maguer, R. Grover, C. Ferrier-Pagès, *Proc. Biol. Sci.* **282**, 20150610 (2015).
26. M. K. Donovan *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5351–5357 (2020).
27. P. J. Edmunds, R. C. Carpenter, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5067–5071 (2001).
28. T. R. McClanahan, N. A. Muthiga, *Hydrobiologia* **166**, 269–276 (1988).
29. C. Birkeland, *Echinoderm Stud.* **3**, 1–79 (1989).
30. C. R. C. Sheppard, *Mar. Ecol. Prog. Ser.* **7**, 83–115 (1982).
31. S. R. Palumbi, D. J. Barshis, N. Traylor-Knowles, R. A. Bay, *Science* **344**, 895–898 (2014).
32. A. Safaie *et al*., *Nat. Commun.* **9**, 1671 (2018).
33. S. Sully, R. van Woesik, *Glob. Change Biol.* **26**, 1367–1373 (2020).
34. B. Stockwell, C. R. L. Jadloc, R. A. Abesamis, A. C. Alcala, G. R. Russ, *Mar. Ecol. Prog. Ser.* **389**, 1–15 (2009).
35. C. L. Hunter, C. W. Evans, *Bull. Mar. Sci.* **57**, 501–515 (1995).
36. R. van Hooidonk *et al*., *Sci. Rep.* **6**, 39666 (2016).
37. T. P. Hughes *et al*., *Nature* **556**, 492–496 (2018).
38. Q. He, B. R. Silliman, *Curr. Biol.* **29**, R1021–R1035 (2019).
39. M. K. Donovan, fishymary/global_coral_change: Revised release of coral change codes, Version 1.1.0, Zenodo (2021); http://doi.org/10.5281/zenodo.4619461.

ACKNOWLEDGMENTS

We thank J. Mihaly and the more than 10,000 volunteer scientists and citizen scientists who have collected Reef Check data since 1997. We thank C. Cacciapaglia for providing access to hurricane data and J. Lecky for help with Fig. 1. We thank B. Silliman, R. Vega Thurber, T. Adam, J. Lauder, and J. Lecky for helpful reviews and comments. T.A.O. acknowledges his affiliation with the University of Hawaiʻi at Mānoa. **Funding:** This research was funded by National Science Foundation grants OCE-1829393 and OCE 1657633 and a grant from the Zegar Family Foundation. **Author contributions:** Formulation of ideas: M.K.D., D.E.B., T.A.O., and R.v.W. Development of methodology and models: M.K.D. Programming and/or implementation of code: M.K.D. Verification of replicability of results: M.K.D., D.E.B., S.S., T.A.O., and R.v.W. Data analysis: M.K.D. Data collection: G.H. and J.F. Data stewardship: M.K.D., C.K., S.S., and G.H. Writing and revisions: All authors. Figures: M.K.D. Oversight: D.E.B. and R.v.W. Coordination: M.K.D., D.E.B., and R.v.W. Funding acquisition: D.E.B. and R.v.W. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and scripts used in the analysis are available in Zenodo (*39*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/977/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 and S2
References (*40*–*49*)

27 July 2020; accepted 25 March 2021
10.1126/science.abd9464

ECOSYSTEM SENTINELS

Hemispheric asymmetry in ocean change and the productivity of ecosystem sentinels

W. J. Sydeman[1]*, D. S. Schoeman[2,3], S. A. Thompson[1], B. A. Hoover[4], M. García-Reyes[1], F. Daunt[5], P. Agnew[6], T. Anker-Nilssen[7], C. Barbraud[8], R. Barrett[9], P. H. Becker[10], E. Bell[11], P. D. Boersma[12], S. Bouwhuis[10], B. Cannell[13], R. J. M. Crawford[14], P. Dann[15], K. Delord[8], G. Elliott[16], K. E. Erikstad[17], E. Flint[18], R. W. Furness[19], M. P. Harris[5], S. Hatch[20], K. Hilwig[21], J. T. Hinke[22], J. Jahncke[23], J. A. Mills[24], T. K. Reiertsen[25], H. Renner[21], R. B. Sherley[26], C. Surman[27], G. Taylor[16], J. A. Thayer[1], P. N. Trathan[28], E. Velarde[29], K. Walker[16], S. Wanless[5], P. Warzybok[23], Y. Watanuki[30]

Climate change and other human activities are causing profound effects on marine ecosystem productivity. We show that the breeding success of seabirds is tracking hemispheric differences in ocean warming and human impacts, with the strongest effects on fish-eating, surface-foraging species in the north. Hemispheric asymmetry suggests the need for ocean management at hemispheric scales. For the north, tactical, climate-based recovery plans for forage fish resources are needed to recover seabird breeding productivity. In the south, lower-magnitude change in seabird productivity presents opportunities for strategic management approaches such as large marine protected areas to sustain food webs and maintain predator productivity. Global monitoring of seabird productivity enables the detection of ecosystem change in remote regions and contributes to our understanding of marine climate impacts on ecosystems.

Earth's environments and biological systems are changing at unprecedented rates. An underappreciated emergent property of global change is differences, or asymmetries, in the responses of marine ecosystems in the Northern and Southern Hemispheres to anthropogenic influences. In the Northern Hemisphere, ecosystem change is thought to be more pronounced because humans have been exploiting marine resources at industrial levels there over longer periods of time (*1*). Further, greater land mass in the north may amplify rates of anthropogenic global warming (*2*). By contrast, the vast oceanic domains of the Southern Hemisphere are believed to more efficiently buffer the effects of greenhouse gas emissions on ocean temperatures. The remoteness of marine systems in the south has also limited human access and some associated impacts (whaling being one major exception), until recently (*3*).

Analysis of temperature trends, velocity of ocean warming (*4*), and indices of marine heatwaves confirms hemispheric differences in climate change impacts for the upper ocean (Fig. 1). The variety and scope of human impacts on marine ecosystems (*3*) are greater in the Northern Hemisphere but have expanded faster in the Southern Hemisphere over the period 2003–2013 [Fig. 1, I to L (*3*, *5*)].

To date, global syntheses of marine ecosystem change have not explicitly considered hemispheric variation in ocean climate change (*6*). Moreover, by necessity, most assessments have been based on disparate taxon-specific response variables, such as calcification rates in corals, range and distributional shifts of fish, timing and intensity of plankton blooms, or vital rates of marine vertebrates (*6*, *7*). The use of dissimilar response variables across species and trophic levels limits synthesis to simplified, often binary metrics of effects (e.g.,

---

[1]Farallon Institute, Petaluma, CA, USA. [2]Global-Change Ecology Research Group, School of Science, Technology and Engineering, University of the Sunshine Coast, Sippy Downs, Queensland, Australia. [3]Centre for African Conservation Ecology, Department of Zoology, Nelson Mandela University, Gqeberha, South Africa. [4]Chapman University, Orange, CA, USA. [5]UK Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian, UK. [6]Oamaru Blue Penguin Colony, Oamaru, New Zealand. [7]Norwegian Institute for Nature Research (NINA), Trondheim, Norway. [8]Centre d'Etudes Biologiques de Chizé, CNRS UMR7372, Villiers en Bois, France. [9]UiT The Arctic University of Norway, Tromsø, Norway. [10]Institute of Avian Research, Wilhelmshaven, Germany. [11]Wildlife Management International, Blenheim, New Zealand. [12]Center for Ecosystem Sentinels, Department of Biology, University of Washington, Seattle, WA, USA. [13]Murdoch University, Murdoch, Western Australia, and University of Western Australia, Perth, Western Australia. [14]Department of Environment, Forestry and Fisheries, Cape Town, South Africa. [15]Phillip Island Nature Parks, Cowes, Victoria, Australia. [16]New Zealand Department of Conservation, Wellington, New Zealand. [17]Norwegian Institute for Nature Research (NINA), FRAM Centre, Tromsø, Norway and Centre for Biodiversity Dynamics (CBD), Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [18]U.S. Fish and Wildlife Service, Honolulu, HI, USA. [19]University of Glasgow, Glasgow, Scotland, UK. [20]Institute for Seabird Research and Conservation, Anchorage, AK, USA. [21]U.S. Fish and Wildlife Service, Anchorage, AK, USA. [22]Antarctic Ecosystem Research Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, USA. [23]Point Blue Conservation Science, Petaluma, CA, USA. [24]Kaikoura, New Zealand. [25]Norwegian Institute for Nature Research (NINA), FRAM Centre, Tromsø, Norway. [26]Centre for Ecology and Conservation, University of Exeter, Cornwall, UK. [27]Halfmoon Biosciences, Ocean Beach, Western Australia, Australia. [28]British Antarctic Survey, Cambridge, UK. [29]Universidad Veracruzana, Veracruz, Mexico. [30]Hokkaido University, Hakodate, Hokkaido, Japan.
***Corresponding author. Email: wsydeman@faralloninstitute.org**

"consistency" with predictions of climate change), which hampers comparisons across marine ecosystems (*8*, *9*). To avoid these issues, a unified approach is required in which analogous and interpretable variables are considered at the scale of large marine ecosystems (*10*). A good example of an overarching metric to assess marine ecosystem change has recently emerged in the form of estimates of "breeding productivity" of marine predators, i.e., the number of young produced per female per year (*11*, *12*).

Here, we test the hypothesis that the breeding productivity of seabirds is tracking hemispheric asymmetry in ocean climate change and human use. Because of the availability of global-scale data on breeding productivity, seabirds stand out among marine vertebrates, with numerous multidecadal datasets in both hemispheres [e.g., the Crozet, Pribilof, and Farallon islands (*13, 14*); see tables S1 to S3 and figs. S1 and S2). The accumulation of long-term datasets among hemispheres is markedly similar, although there are fewer data overall in the south (fig. S1). During breeding, seabirds provision themselves and their offspring on a wide variety of food resources spanning copepods to small pelagic fish and thereby provide an integrated response to climate change across trophic levels (*7*). Seabird breeding productivity is known to reflect nonlinear numerical responses to mesozooplankton and small fish availability in the epipelagic zone (*15*–*18*). Seabirds, which breed in colonies but forage at sea during reproduction, may be particularly vulnerable to ocean change because their breeding sites are static in space, whereas the availability of their food resources is spatially and temporally dynamic.

We predicted greater declines in Northern Hemisphere seabird breeding productivity than Southern. To test this prediction, we compiled 122 time series of annual breeding productivity (proportionate change from the long-term mean) for 66 seabird species, representing 3586 annual data points across the globe over the period 1964–2018 (tables S1 to S3 and fig. S2). We used these data to conduct an analysis of seabird breeding success in relation to hemispheric asymmetry in ocean warming (*19*).

To evaluate possible hemispheric variation in reproductive trends associated with seabird ecology, we categorized each species' trophic level based on their primary diet during the breeding season on the resolution of decades (*19*). Trophic level is a key ecological characteristic because the effect of climate change on marine top predators often acts mechanistically through food resources (*7*). For this investigation, we categorized species as (i) planktivores that primarily consume mesozooplankton and larval fishes,

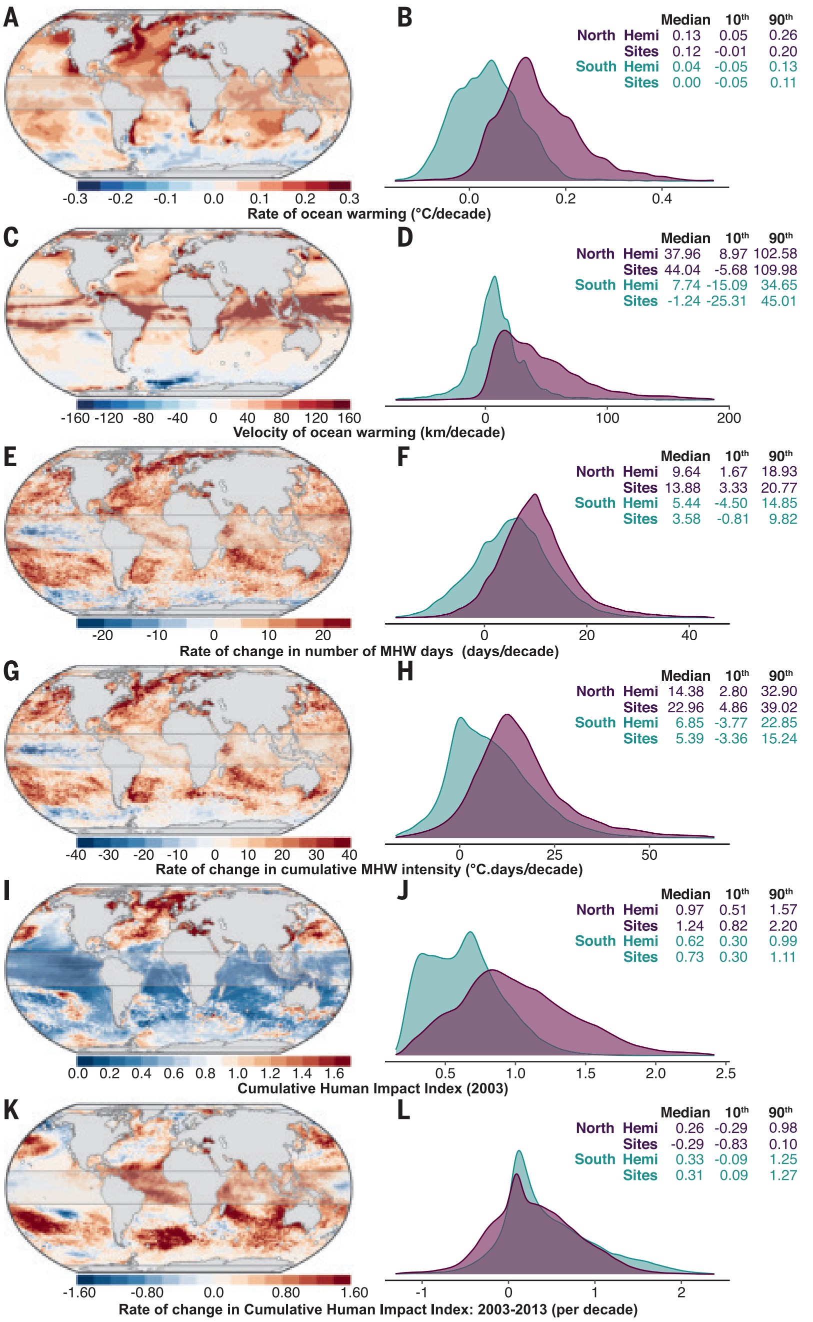


Fig. 1. Maps and kernel density plots showing hemispheric differences in indices of marine climate change and human use in the upper ocean. (**A** and **B**) Rate of warming and (**C** and **D**) velocity of ocean warming based on HadISST1 data over the 50-year period 1968–2019. Also shown are trends over the period 1968–2017 for cumulative number of marine heatwave (MHW) days (**E** and **F**), trends in cumulative marine heatwave intensity (**G** and **H**) based on National Oceanic and Atmospheric Administration Optimum Interpolation Sea Surface Temperature (OISST), cumulative human impacts in 2003 [e.g., fishing, shipping, contamination as defined by Halpern *et al*. (*3*)] (**I** and **J**), and finally, rate of change in human impacts from 2003 to 2013 (**K** and **L**). All maps are overlaid with 46 locations across the world where seabird breeding biology was studied (white circles). For kernel density plots (and summary statistics), data equatorward of 15° and poleward of 75° in both hemispheres (lightly shaded polygons on the maps) were excluded to avoid bias by extreme values or seasonally missing data. None of our sample sites was within these areas. All data presented were regridded onto equal-area hexagons (~0.5° at the equator) for computation and visual representation to avoid latitudinal bias in grid area. Resulting data are summarized for each kernel density plot [tables in (B), (D), (F), (J), and (L)] by median, 10th and 90th percentiles by hemisphere (Hemi), and sites within hemisphere (Sites).

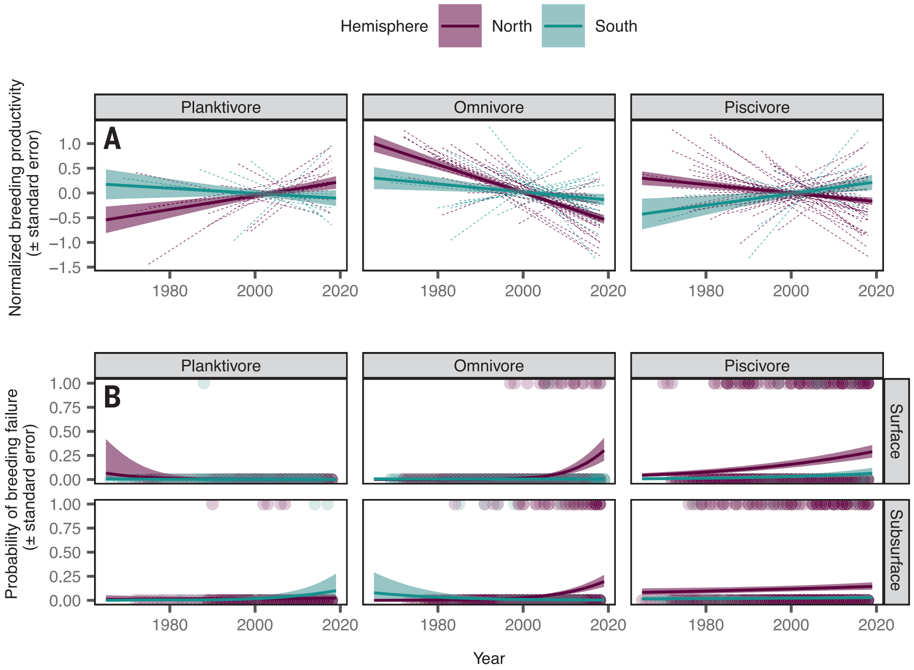


Fig. 2. Modeled trends of seabird breeding productivity and probability of breeding failure. (**A**) Normalized breeding productivity of seabirds by trophic level and hemisphere (north, purple; south, turquoise; values are ± SE) as a function of time. Trends in individual time series (dashed lines) are shown as background. See the supplementary materials and methods (*19*) for model details. (**B**) Modeled trends in (±SE) the probability of breeding failure by trophic level, foraging depth (surface or subsurface), and hemisphere. Observations of breeding success or failure are provided as background points, colored by hemisphere.

(ii) piscivores that primarily consume small pelagic fish, and (iii) omnivores that consume both plankton and fish. After considering trophic level, we further dissected species' foraging characteristics by investigating whether species feeding in the upper water column ("surface foraging") were more vulnerable to ocean changes than species foraging at depths >10 m ("subsurface foraging"). Feeding depth covaries with other life-history traits in seabirds, such as body size and foraging range (*20*). Whereas >47% of the planktivores and 59% of the piscivorous species included in our study were subsurface foragers, only nine (33%) of the 27 omnivorous species foraged at greater depths (table S2).

Trends in seabird breeding productivity varied by trophic level; within each trophic level, they also varied by hemisphere, but we found no overall effect of foraging depth on productivity (Fig. 2A, table S4, and fig. S3). Omnivorous species, many of which provision young with small pelagic fish (table S2), showed the most substantial changes in both hemispheres, with larger decreasing trends in normalized breeding productivity in the north [1.00 ± 0.17 to −0.53 ± 0.09 (point estimates from the start to the end of the study period, respectively, ± SE)] than in the south (0.3 ± 0.22 to −0.13 ± 0.11). Breeding productivity of piscivorous species declined in the north (0.30 ± 0.14 to −0.17 ± 0.07) but increased in the south (−0.43 ± 0.31 to 0.21 ± 0.16). By contrast, planktivorous seabirds showed increasing productivity trends in the north (−0.54 ± 0.27 to 0.21 ± 0.13) and stable productivity in the south (0.18 ± 0.30 to −0.1 ± 0.16).

To ascertain whether decreasing productivity is related to an increasing rate of breeding failure that is potentially associated with the increasing frequency of marine heatwaves in both hemispheres (Fig. 1, C and D), we examined the probability of breeding failure, defined as breeding success <10% relative to the mean for each time series (*19*). Trends in the probability of breeding failure corroborated observations of normalized breeding success, with significant effects varying by hemisphere, trophic level, and foraging depth (Fig. 2B and table S5). Overall, the probability of breeding failure increased for piscivores in both hemispheres and for omnivores particularly in the north. For piscivores, probability of breeding failure was significantly higher in the Northern Hemisphere, and it was also elevated for surface-foraging species in both hemispheres, especially in recent years. Surface-feeding omnivores in the north followed a similar pattern, with the probability of breeding failure escalating rapidly after the year 2000. Breeding failure was relatively uncommon for planktivores and omnivores in the Southern Hemisphere, where trends were weak.

Thus far, global analyses of seabirds have shown inconsistent responses to climate change in terms of their phenology [i.e., timing of reproduction (*21*)], although trends in vital rates have been more consistently negative (*6*, *7*). By using seabird breeding productivity as a unified variable to sense change at the global scale, we observed greater consistency in identified responses to ocean warming (Figs. 1 and 2), although hemispheric variation in the magnitude and rate of the warming apparently affected fish-eating species the most. We could not include temperature change directly in our models because temperature increased with time and would confound the effect of hemisphere. Moreover, because rates of change in temperature, velocity of ocean warming, and marine heatwaves vary strongly by hemisphere (Fig. 1), using hemisphere as a covariate in models is a proxy for these metrics of anthropogenic climate change. We nevertheless conducted sensitivity tests that omitted hemisphere as a fixed effect, in which we found evidence of declining breeding success with an increasing rate of ocean warming or velocity of ocean warming (*19*). The variance explained by our models was low, so other variables that may play a role in determining productivity, such as short-term local weather events (*22*) or other factors that may affect food resources (e.g., fisheries) or density-dependent mechanisms [e.g., (*23*)], should be considered in future analyses. Even with the substantial global dataset that we compiled, we were unable to address all competing factors that drive variation in seabird breeding productivity.

Our study provides important insights for ecosystem monitoring and management. First, because seabirds accumulate and integrate, in a statistical sense, climatic, oceanographic, and food-web variation, they provide immediate signals of changes in ecosystems that are difficult to observe directly, particularly in remote regions of the world (*24*). The signals provided by seabird breeding productivity could easily be used to assess global change in marine ecosystems on an annual time frame with relatively simple coordination and data sharing of governmental monitoring programs. Second, most global climate models predict increasing ocean stratification caused by ocean warming (*9*, *25*), which may limit nutrient input into the epipelagic zone and thereby affect mesozooplankton and forage fish populations (*26*). The dwindling productivity of seabirds across the north, with its greater rate of warming, suggests that increasing stratification may already be affecting marine ecosystems there (*27*, *28*), although other confounding human impacts on surface-foraging species [e.g., plastics pollution (*29*)] may also partly explain increases in probability of breeding failures. Correspondingly, the significant effect of foraging depth on trends in probability of breeding failure across trophic levels suggests that access to subsurface foraging habitats, regardless of trophic level or hemisphere, confers some resilience to subsurface-feeding seabirds such as penguins and puffins. By contrast, surface-feeding albatrosses, petrels, and terns may be the most susceptible to warming-related changes in food resources. Third, whereas the rate of change in human use and impacts is increasing more rapidly in the Southern Hemisphere, the overall impacts of humanity on marine ecosystems, including the combined effects of climate change, fisheries and other forms of marine resource exploitation, and pollution, are considerably greater in the Northern Hemisphere (*3*, *9*).

Our study indicates that the prognosis for sustained breeding productivity of Northern Hemisphere fish-eating and omnivorous breeding seabirds is poor unless the availability of food resources is improved. One approach to increasing functional seabird predator-prey interactions could include enhancing food web redundancies and connectivity through management targeting prey diversity, which would promote a greater portfolio of forage fish populations. An obvious factor that needs continuing attention is the temporal or spatial management of fisheries that target small pelagic fish or large zooplankton (e.g., krill) and may compete with seabirds for food, especially near colonies during the reproductive period (*30*). Time-area fisheries closures may be an effective strategy for improving seabird productivity in the Northern Hemisphere (*31*).

Implications of our study for Southern Hemisphere seabirds and their ecosystems demonstrate less urgency, although there are regional exceptions, especially in the Southern Ocean (*32*). Generally, however, our results portend opportunity in the Southern Hemisphere, where implementation of longer-term, ecosystem-based approaches could be effective in mitigating impending human impacts (*3*, *32*) and those predicted by the latest suite of Earth system models (*9*). Although compelling effects on seabird breeding productivity have been realized for fisheries closures near seabird colonies (*31*), the establishment of large marine protected areas (*33*, *34*) could enhance seabird and other predator foraging opportunities and communities year-round. If foraging opportunities can be managed, even relatively small changes in breeding productivity over the long term could enhance population stability and recovery (*31*).

In conclusion, the disparity between central-place foraging seabirds breeding at fixed points in space relative to spatially and temporally dynamic ocean habitats and prey resources (*35*, *36*) places seabirds at particular risk from ocean climate change, especially in combination with other human-induced perturbations such as pollution and fisheries. The changes in seabird breeding productivity related to hemispheric variation in ocean warming and human uses documented in this study calls out the need to sustain long-term monitoring programs (some of which are threatened), illustrates the critical role that seabirds play as sentinels of global marine change, and highlights the need for policies that reduce climate change impacts on the world's marine ecosystems.

REFERENCES AND NOTES

1. P. M. Vitousek, H. A. Mooney, J. Lubchenco, J. M. Melillo, *Science* **277**, 494–499 (1997).
2. M. T. Burrows *et al*., *Science* **334**, 652–655 (2011).
3. B. S. Halpern *et al*., *Sci. Rep.* **9**, 11609 (2019).
4. S. R. Loarie *et al*., *Nature* **462**, 1052–1055 (2009).
5. B. S. Halpern *et al*., *PLOS ONE* **10**, e0117863 (2015).
6. E. S. Poloczanska *et al*., *Nat. Clim. Chang.* **3**, 919–925 (2013).
7. W. J. Sydeman, E. Poloczanska, T. E. Reed, S. A. Thompson, *Science* **350**, 772–777 (2015).
8. C. J. Brown *et al*., *Glob. Chang. Biol.* **22**, 1548–1560 (2016).
9. N. L. Bindoff, W. W. L. Cheung, J. G. Kairo, "Changing ocean, marine ecosystems, and dependent communities," in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N. M. Weyer, Eds. (IPCC, 2019), pp. 447–587.
10. J. C. Rice, M.-J. Rochet, *ICES J. Mar. Sci.* **62**, 516–527 (2005).
11. E. L. Hazen *et al*., *Front. Ecol. Evol.* **17**, 565–574 (2019).
12. E. Velarde, D. W. Anderson, E. Ezcurra, *Science* **365**, 116–117 (2019).
13. R. D. Wooller, J. S. Bradley, J. P. Croxall, *Trends Ecol. Evol.* **7**, 111–114 (1992).
14. J. P. Croxall, P. N. Trathan, E. J. Murphy, *Science* **297**, 1510–1514 (2002).
15. D. K. Cairns, *Biol. Oceanogr.* **5**, 261–271 (1988).
16. J. F. Piatt *et al*., *Mar. Ecol. Prog. Ser.* **352**, 221–234 (2007).
17. P. M. Cury *et al*., *Science* **334**, 1703–1706 (2011).
18. J. A. Santora, I. D. Schroeder, J. C. Field, B. K. Wells, W. J. Sydeman, *Ecol. Appl.* **24**, 1730–1747 (2014).
19. Materials and methods are available as supplementary materials.
20. R. W. Furness, M. L. Tasker, *Mar. Ecol. Prog. Ser.* **202**, 253–264 (2000).
21. K. Keogan *et al*., *Nat. Clim. Chang.* **8**, 313–318 (2018).
22. E. A. Schreiber, "Climate and weather effects on seabirds," in *Biology of Marine Birds*, E. A. Schreiber, J. Burger, Eds. (CRC Press, 2002), pp. 179–215.
23. G. L. Hunt Jr., Z. A. Eppley, D. C. Schneider, *Auk* **103**, 306–317 (1986).
24. A. J. Constable, W. K. de la Mare, D. J. Agnew, I. Everson, D. Miller, *ICES J. Mar. Sci.* **57**, 778–791 (2000).
25. Intergovernmental Panel on Climate Change, *Climate Change 2013 – The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
26. C. M. Free *et al*., *Science* **363**, 979–983 (2019).
27. M. J. Carroll *et al*., *Clim. Res.* **66**, 75–89 (2015).
28. F. Ramírez *et al*., *Proc. Biol. Sci.* **283**, 20152287 (2016).
29. C. Wilcox, E. Van Sebille, B. D. Hardesty, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11899–11904 (2015).
30. W. J. Sydeman *et al*., *Fish. Res.* **194**, 209–221 (2017).
31. R. B. Sherley *et al*., *Proc. Biol. Sci.* **285**, 20172443 (2018).
32. S. Bestley *et al*., *Front. Ecol. Evol.* **8**, 566936 (2020).
33. P. N. Trathan *et al*., *Adv. Mar. Biol.* **69**, 15–78 (2014).
34. C. M. Brooks, L. B. Crowder, H. Osterblom, A. L. Strong, *Conserv. Lett.* **13**, e12676 (2019).
35. I. Brito-Morales *et al*., *Trends Ecol. Evol.* **33**, 441–457 (2018).
36. M. T. Burrows *et al*., *Nat. Clim. Chang.* **9**, 959–963 (2019).
37. Data and code for: W. J. Sydeman, D. S. Schoeman, S. A. Thompson, B. A. Hoover, M. García-Reyes, F. Daunt, P. Agnew, T. Anker-Nilssen, C. Barbraud, R. Barrett, P. H. Becker, E. Bell, P. D. Boersma, S. Bouwhuis, B. Cannell, R. J. M. Crawford, P. Dann, K. Delord, G. Elliott, K. E. Erikstad, E. Flint, R. W. Furness, M. P. Harris, S. Hatch, K. Hilwig, J. T. Hinke, J. Jahncke, J. A. Mills, T. K. Reiertsen, H. Renner, R. B. Sherley, C. Surman, G. Taylor, J. A. Thayer, P. N. Trathan, E. Velarde, K. Walker, S. Wanless, P. Warzybok, Y. Watanuki, Hemispheric asymmetry in ocean change and the productivity of ecosystem sentinels, Zenodo (2021); https://doi.org/10.5281/zenodo.4667747.

ACKNOWLEDGMENTS

This study results from the dedication and commitment to long-term seabird population studies by many governmental and nongovernmental agencies and numerous self-funded individuals across the globe. Various ideas for analyses and prereview of results and other material were provided by colleagues B. A. Black, G. L. Hunt, C. Kroeger, J. F. Piatt, J. A. Santora, and G. Kegley. **Funding:** No specific funding sources contributed to the analyses and write-up of the data reported in this paper. Acknowledgments for data collection efforts and funding are provided in the supplementary text. **Author contributions:** W.J.S., D.S., and S.A.T. designed the study. W.J.S., D.S., S.A.T., B.A.H., M.G.R., and F.D. designed and implemented analyses. W.J.S., D.S., S.A.T., and B.A.H. wrote the initial draft of the manuscript. P.A., T.A.N., C.B., R.B., P.H.B., E.B., P.D.B., S.B., B.C., R.J.M.C., P.D., F.D., K.D., G.E., K.E.E., E.F., R.W.F., M.G.R., M.P.H., S.H., K.H., J.T.H., J.J., J.A.M., T.K.R., H.R., R.B.S., C.S., G.T., J.A.T., P.N.T., E.V., K.W., S.W., P.W., and Y.W. contributed data and reviewed and edited the manuscript for accuracy, clarity, and brevity. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Active versions of the code and data are available on GitHub (https://github.com/DavidSchoeman/sydeman_et_al_seabirds) and are archived on Zenodo (*37*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/980/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Tables S1 to S5
References (*38*–*55*)

8 October 2020; accepted 20 April 2021
10.1126/science.abf1772

GENOMIC EVOLUTION

3D genomics across the tree of life reveals condensin II as a determinant of architecture type

Claire Hoencamp[1]†, Olga Dudchenko[2,3,4]†, Ahmed M. O. Elbatsh[1]†‡, Sumitabha Brahmachari[4]†, Jonne A. Raaijmakers[5]§, Tom van Schaik[6]§, Ángela Sedeño Cacciatore[1]§, Vinícius G. Contessoto[4,7]§, Roy G. H. P. van Heesbeen[5]§¶, Bram van den Broek[8], Aditya N. Mhaskar[1]#, Hans Teunissen[6], Brian Glenn St Hilaire[2,3], David Weisz[2,3], Arina D. Omer[2], Melanie Pham[2], Zane Colaric[2], Zhenzhen Yang[9], Suhas S. P. Rao[2,3,10], Namita Mitra[2,3], Christopher Lui[2], Weijie Yao[2], Ruqayya Khan[2,3], Leonid L. Moroz[11], Andrea Kohn[11], Judy St. Leger[12], Alexandria Mena[13], Karen Holcroft[14], Maria Cristina Gambetta[15], Fabian Lim[16], Emma Farley[16], Nils Stein[17,18,19], Alexander Haddad[2], Daniel Chauss[20], Ayse Sena Mutlu[3], Meng C. Wang[3,21,22], Neil D. Young[23], Evin Hildebrandt[24], Hans H. Cheng[24], Christopher J. Knight[25], Theresa L. U. Burnham[26,27], Kevin A. Hovel[27], Andrew J. Beel[10], Pierre-Jean Mattei[10], Roger D. Kornberg[10], Wesley C. Warren[28], Gregory Cary[29], José Luis Gómez-Skarmeta[30]††, Veronica Hinman[31], Kerstin Lindblad-Toh[32,33], Federica Di Palma[34], Kazuhiro Maeshima[35,36], Asha S. Multani[37], Sen Pathak[37], Liesl Nel-Themaat[37]‡‡, Richard R. Behringer[37], Parwinder Kaur[19], René H. Medema[5], Bas van Steensel[6], Elzo de Wit[6], José N. Onuchic[4,38], Michele Di Pierro[4,39], Erez Lieberman Aiden[2,3,4,9,19,32]*, Benjamin D. Rowland[1]***

We investigated genome folding across the eukaryotic tree of life. We find two types of three-dimensional (3D) genome architectures at the chromosome scale. Each type appears and disappears repeatedly during eukaryotic evolution. The type of genome architecture that an organism exhibits correlates with the absence of condensin II subunits. Moreover, condensin II depletion converts the architecture of the human genome to a state resembling that seen in organisms such as fungi or mosquitoes. In this state, centromeres cluster together at nucleoli, and heterochromatin domains merge. We propose a physical model in which lengthwise compaction of chromosomes by condensin II during mitosis determines chromosome-scale genome architecture, with effects that are retained during the subsequent interphase. This mechanism likely has been conserved since the last common ancestor of all eukaryotes.

The mechanisms controlling nuclear architecture at the scale of whole chromosomes remain poorly understood. To investigate principles of genome folding, we performed in situ Hi-C (*1*) on 24 species, representing all subphyla of chordates, all seven extant vertebrate classes, seven of nine major animal phyla, as well as plants and fungi (Fig. 1, figs. S1 and S2, and table S1). For 14 species, there was no existing chromosome-length reference genome assembly. For these, we upgraded existing genome assemblies or assembled a reference genome entirely from scratch (*2*) (table S2). Together, these species offer a comprehensive overview of nuclear organization since the last common ancestor of all eukaryotes.

The resulting maps reveal four features of nuclear architecture at the scale of whole chromosomes (Fig. 1 and fig. S1). First, some species, such as the red piranha, exhibit enhanced contact frequency between loci on the same chromosome. This is consistent with, though not necessarily identical to, classical chromosome territories as traditionally observed by cytogenetics—when a chromosome occupies a discrete subvolume of the nucleus, excluding other chromosomes (*3*). Second, species like the yellow fever and southern domestic mosquitoes exhibit prominent contacts between centromeres. Third, species like the ground peanut exhibit prominent contacts between telomeres. Finally, species like bread wheat exhibit an X-shape on the chromosomal map (Fig. 1 and figs. S1, S2, S3, and S4). We refer to these last three features as Rabl-like, because they are reminiscent of the Rabl chromosome configuration (*4*), in which centromeres cluster and chromosome arms are arranged in parallel.

To identify these architectural features in an unbiased fashion, we developed aggregate chromosome analysis (ACA), whereby contact maps for each chromosome are rescaled and summed and then used to score each feature (*2*) (figs. S3 and S6 and table S3). All species that are not holocentric exhibit at least one feature. The architectural features can be divided into two clusters, type-I and type-II, on the basis of how likely the features are to co-occur (fig. S7 and table S4). Type-I includes the three Rabl-like features: centromere clustering, telomere clustering, and a telomere-to-centromere axis. Type-II includes only chromosome territories. Consequently, species can also be subdivided depending on which feature cluster is more strongly exhibited (table S3).

Homologs tend to be separated or paired depending on the species. We found that type-II species typically exhibit homolog separation, whereas this is less frequent among type-I species (figs. S8 and S9 and table S5). We developed an algorithm, dubbed 3D-DNA Phaser, that exploits this separation, when present, to

---

[1]Division of Gene Regulation, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [2]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA. [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [4]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA. [5]Division of Cell Biology, Oncode Institute, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [6]Division of Gene Regulation, Oncode Institute, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [7]Department of Physics, Institute of Biosciences, Letters and Exact Sciences, São Paulo State University (UNESP), São José do Rio Preto – SP, 15054-000, Brazil. [8]BioImaging Facility, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [9]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech, Pudong 201210, China. [10]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [11]Whitney Laboratory and Department of Neuroscience, University of Florida, Gainesville, FL 32611, USA. [12]Department of Biosciences, Cornell University College of Veterinary Medicine, Ithaca, NY 14853, USA. [13]SeaWorld San Diego, San Diego, CA 92109, USA. [14]Moody Gardens, Galveston, TX 77554, USA. [15]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. [16]Department of Medicine and Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA. [17]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben), 06466 Seeland, Germany. [18]Center of Integrated Breeding Research (CiBreed), Department of Crop Sciences, Georg-August-University Göttingen, 37075 Göttingen, Germany. [19]UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA 6009, Australia. [20]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [21]Huffington Center on Aging, Baylor College of Medicine, Houston, TX 77030, USA. [22]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA. [23]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC 3010, Australia. [24]Avian Diseases and Oncology Laboratory, US Department of Agriculture, Agricultural Research Service, East Lansing, MI 48823, USA. [25]Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA. [26]Department of Wildlife, Fish, and Conservation Biology, University of California, Davis, Davis, CA 95616, USA. [27]Coastal and Marine Institute and Department of Biology, San Diego State University, San Diego, CA 92106, USA. [28]Department of Animal Sciences, University of Missouri, Columbia, MO 65211, USA. [29]The Jackson Laboratory, Bar Harbor, ME 04609, USA. [30]Centro Andaluz de Biología del Desarrollo CSIC, Universidad Pablo de Olavide, 41013 Sevilla, Spain. [31]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [32]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [33]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden. [34]Department of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [35]Genome Dynamics Laboratory, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [36]Department of Genetics, Sokendai (Graduate University for Advanced Studies), Mishima, Shizuoka 411-8540, Japan. [37]Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [38]Departments of Physics and Astronomy, Chemistry, and Biosciences, Rice University, Houston, TX 77005, USA. [39]Department of Physics, Northeastern University, Boston, MA 02115, USA.

***Corresponding author. Email: b.rowland@nki.nl (B.D.R.); erez@erez.com (E.L.A.)**

†These authors contributed equally to this work. ‡Present address: Novartis Institutes for Biomedical Research, 4002 Basel, Switzerland. §These authors contributed equally to this work. ¶Present address: Janssen Vaccines & Prevention BV, 2333 CN Leiden, Netherlands. #Present address: Department of Developmental Biology, Erasmus MC, University Medical Centre Rotterdam, 3015 GD Rotterdam, Netherlands. **Present address: Zoetis (VMRD Global Biologics Research), Kalamazoo, MI 49007, USA. ††Deceased. ‡‡Present address: University of Colorado Advanced Reproductive Medicine, Denver, CO 80238, USA.

assign variants to individual homologs, producing chromosome-length haploblocks for multiple species. When homologs are not separated, as in *Drosophila melanogaster*, we show that this approach cannot be used. Taken together, these data are consistent with a model in which features of genome architecture appeared and disappeared over billions of years, as lineages switched between Rabl-like and territorial architectures.

Next, we sought to understand the mechanism underlying this switching behavior. When investigating the transition between the two architectures, we noted that mosquitoes, which display type-I features (Fig. 1), also lack a subunit of the condensin II complex (*5*), which promotes mitotic chromosome compaction (*6*). We therefore searched for condensin II

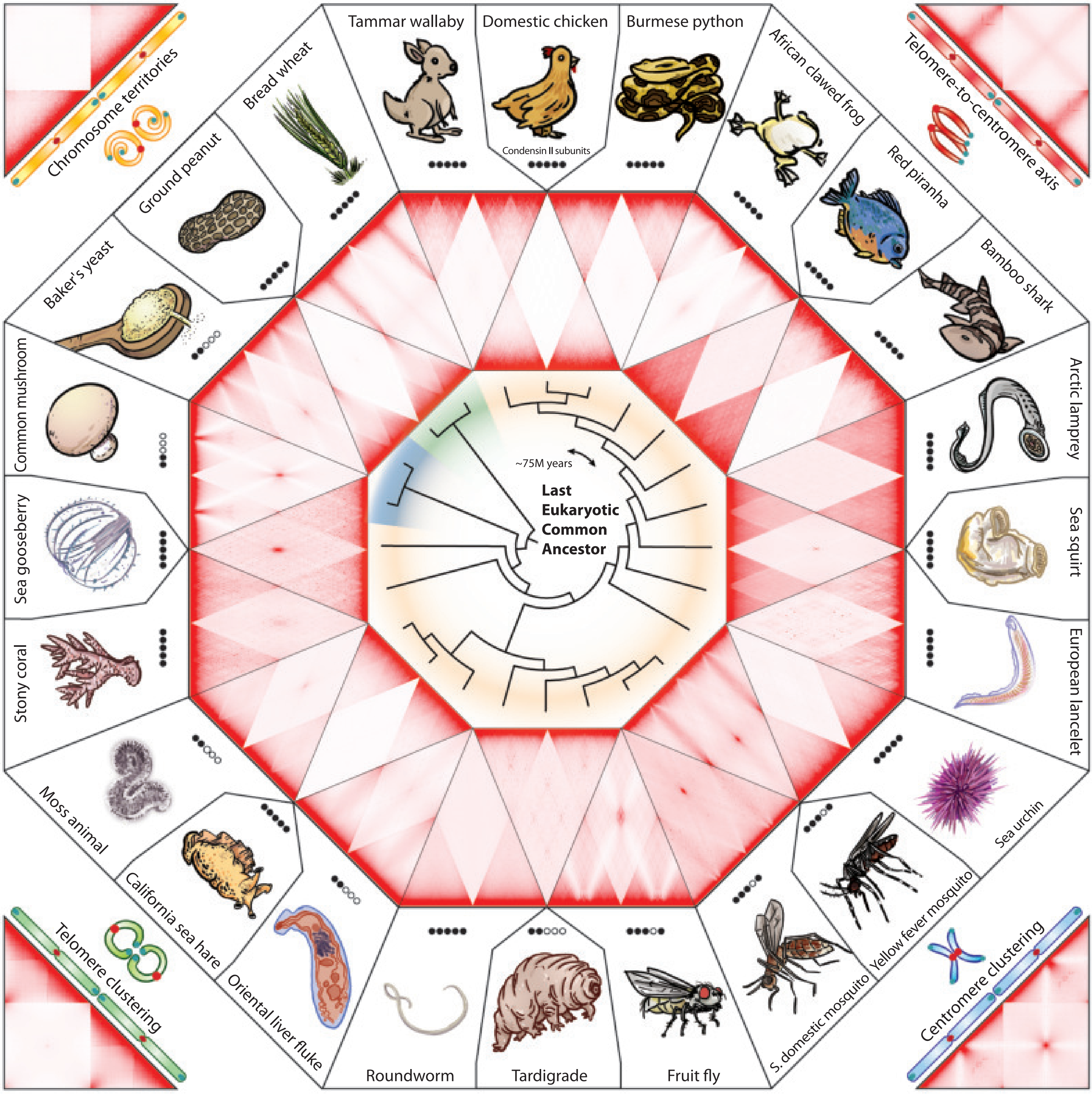


Fig. 1. A comprehensive overview of nuclear architecture across evolution. Aggregate chromosome analysis (ACA) on in situ Hi-C maps of 24 species. In ACA, chromosome arms are rescaled to a uniform length and then the signal of all intra- and interchromosomal contacts is aggregated. This yields an aggregate portrait of genome folding in a species at the scale of whole chromosomes. The 24 ACA plots are rescaled to fit into an octagon, with a depiction of the corresponding species flanking each ACA plot. The species span three kingdoms: animals (yellow), fungi (blue), and plants (green); their evolutionary relationship is represented with a cladogram (*2*). Each corner shows an example ACA map and a schematic drawing of one of the four chromosome-scale features. The location of these example maps does not correspond to the architecture type of the closest species in the figure. Presence of the condensin II subunits in each species is indicated by solid black circles (left to right: SMC2, SMC4, CAP-H2, CAP-G2, and CAP-D3).

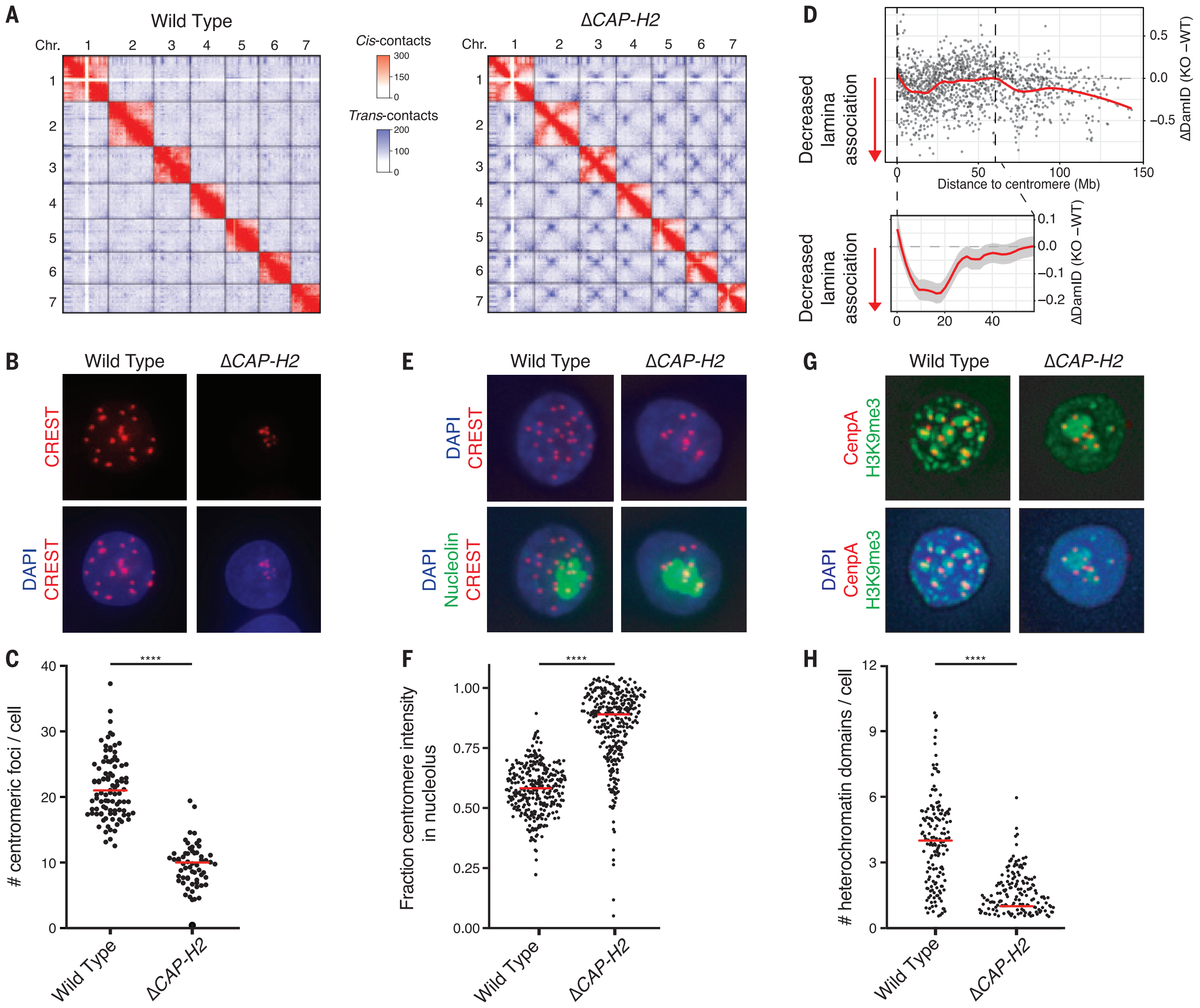


Fig. 2. Condensin II prevents centromeric clustering and keeps apart heterochromatin domains. (**A**) Hi-C matrices of the depicted genotypes in Hap1 cells. Chr., chromosome. (**B** and **C**) Immunofluorescence of centromeres (CREST) and DNA [4′,6-diamidino-2-phenylindole (DAPI)] (B), as quantified in (C). (**D**) Difference in DamID score relative to distance to centromere. Zoom-in includes 95% confidence interval of the mean in gray. KO, knockout; WT, wild type. (**E**) Immunofluorescence of centromeres (CREST), nucleoli (nucleolin), and DNA (DAPI). (**F**) Quantification of the fraction of centromere intensity within 0.4 μm of nucleoli, as shown in (E). (**G** and **H**) Immunofluorescence of centromeres (CenpA), heterochromatin (H3K9me3), and DNA (DAPI) (G), as quantified in (H). $^{****}P < 0.0001$.

subunits in the genomes of all 24 species. Eight species lacked one or more condensin II subunit(s) (table S6) and exhibited Rabl-like features (table S3). Because these organisms lie far apart on the evolutionary tree, type-I architectural features and the loss of condensin II subunits appear to have coevolved repeatedly. This could indicate that condensin II strengthens chromosome territories or counteracts Rabl-like features.

Notably, of the eight species, five lacked all condensin II subunits, whereas the other three species only lacked CAP-G2. Previous work has shown that condensin complexes lacking the G-subunit still localize to DNA but yield elongated chromosomes (*7*). Condensin complexes in these species may thus be impaired, at least partially, in their ability to shorten chromosomes.

Humans exhibit type-II genome architecture, with strong chromosomal territories and no Rabl-like features (Fig. 2A). Moreover, human genomes contain all condensin II subunits. Would disruption of condensin II in human cells then interfere with chromosome territories and enhance the strength of type-I features? To test this, we performed in situ Hi-C on Hap1 cells lacking the condensin II subunit CAP-H2 (Fig. 2A, figs. S14 and S15, and table S7). Disruption of this core condensin II subunit prevents recruitment of the CAP-D3 and CAP-G2 subunits to the complex and renders the complex fully nonfunctional.

ΔCAP-H2 cells exhibited weaker chromosome territories and much stronger contacts between centromeres in trans (Fig. 2A; fig. S15, B and C; and table S8). Immunofluorescence microscopy revealed that in *ΔCAP-H2* cells the centromeres are clustered together. Disruption of condensin II thus transforms the folding of the human genome into a type-I–like configuration (Fig. 2, B and C, and fig. S16).

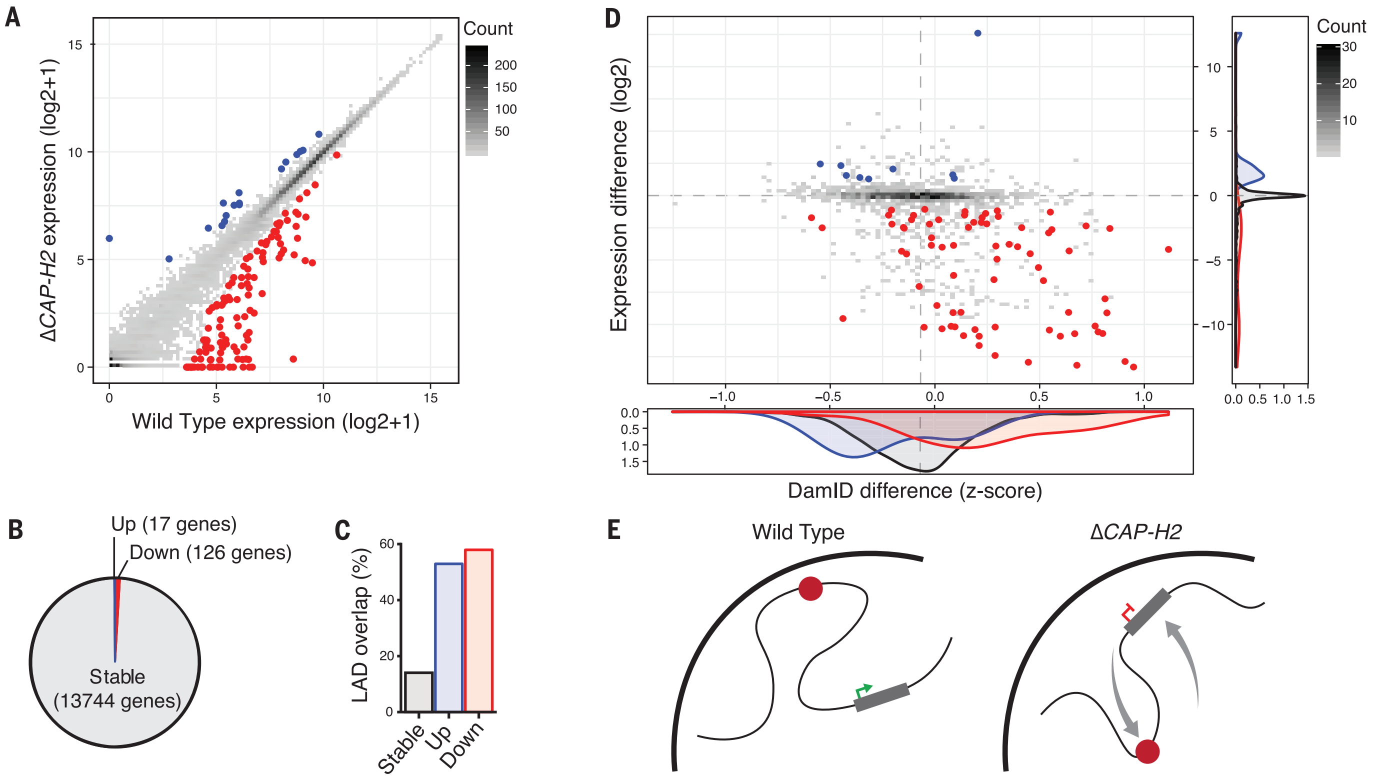


Fig. 3. Massive 3D genome changes hardly affect gene expression. (**A**) Gene expression of wild type relative to Δ*CAP-H2*. Unaffected genes are depicted in gray, up-regulated genes in blue, and down-regulated in red. (**B**) Number of genes in each category. (**C**) Percentage of active genes overlapping with LADs. (**D**) Intersection of differences in gene expression with differences in lamina association, depicting active genes within LADs. (**E**) Schematic model of centromeres (red) moving to the inner nucleus and silenced genes that now localize to the lamina.

Results previously obtained in other species support the model that condensin II plays a major role in three-dimensional (3D) genome organization. In *Arabidopsis*, condensin II regulates the spatial relationship between ribosomal DNAs (rDNAs) and centromeric regions (*8*, *9*), whereas in mouse cells, condensin II regulates the distribution of chromocenters (*10*). Fruit flies lack a condensin II subunit and exhibit centromeric clustering (Fig. 1). Additional depletion of the remaining condensin II subunits in flies affects the spatial distribution of pericentromeric heterochromatin and leads to intermixing of chromosome territories, further strengthening the existing Rabl-like features (*11*, *12*).

Next, we investigated the effects of condensin II loss on human genome architecture in greater detail. To identify DNA segments associated with the nuclear lamina [lamina-associated domains (LADs)], we performed DamID of LaminB1 (*13*) (fig. S17A). LADs localizing up to 25 Mb from the centromeres appeared to move away from the lamina (Fig. 2D and fig. S17, B and C). Centromere repositioning in absence of condensin II thus also moderately affects the lamina association of the regions flanking the centromeres.

In fruit flies, centromeres cluster and localize to the nucleolus (*14*). In Δ*CAP-H2* human cells, centromeres also cluster in or around the nucleolus (Fig. 2, E and F). However, disrupting nucleolar structure did not affect centromeric clustering (fig. S18, A and B). The clustering of centromeres at the human nucleolus is likely because rDNA sequences, which are the genomic component of the nucleolus, often lie near centromeres in the human genome (on the short arm of acrocentric chromosomes) (fig. S18C).

Regions surrounding centromeres are enriched for heterochromatin and cluster upon condensin II depletion in mice and fruit flies (*10*, *11*). Similarly, in Δ*CAP-H2* cells, condensin II deficiency led to clustering of H3K9me3-containing heterochromatin (Fig. 2, G and H), which indicates that condensin II plays a conserved role in the spatial organization of this repressive epigenetic mark. Condensin II deficiency did not affect smaller-scale 3D genome organization at the level of chromatin loops (fig. S19, A and B). Also, compartmentalization was only mildly affected, specifically in regions surrounding the centromeres (fig. S19, C and D). Thus, large-scale reorganization does not necessarily bring about major changes in smaller-scale structures.

RNA sequencing revealed that condensin II deficiency affected the expression of only a fraction of genes (Fig. 3, A and B), which were enriched within LADs (Fig. 3C) and near LAD borders (fig. S20, B and C). The down-regulated genes moved toward the lamina (Fig. 3D). Genes that are near or within LADs could potentially occupy the space that is vacated by the centromeres moving to the nuclear interior upon condensin II loss. The increased lamina association of these genes may, in turn, lead to their transcriptional repression, although the gain in lamina interactions could also be the consequence of the reduced expression of these genes (*15*, *16*) (Fig. 3E).

Thus, condensin II controls the architecture of the interphase genome, but whether it does so by acting in interphase remained unclear. We therefore acutely depleted condensin II in HCT116 cells (*17*) at the G_1-S cell cycle phase transition and either halted the cells before mitotic entry or allowed the cells to progress through mitosis (Fig. 4, A and B, and fig. S21A). When condensin II–depleted cells were halted before mitosis, centromeres did not cluster, which is consistent with condensin II depletion in postmitotic cells not changing the 3D genome (*18*). By contrast, progression through mitosis led to clear centromeric clustering in the subsequent G_1 phase. This suggests that condensin II acts in mitosis, or directly thereafter, to establish 3D genome organization for the next interphase (fig. S21B).

In mitosis, condensin II extrudes loops to compact chromosomes in a lengthwise manner (*19–21*). We used physical simulations to investigate whether this activity of condensin II can affect centromere clustering. In these simulations, chromosomes are polymers bisected by a centromere. These chromosomes are shaped by two forces: (i) the ideal chromosome potential that models lengthwise compaction by condensin II (*22*, *23*) and (ii) centromeric self-adhesion, which models heterochromatin's tendency to cluster (*24–26*) and stabilizes intercentromeric contacts in our setup. We simulated 10 chromosomes with fixed centromere self-adhesion and decreased lengthwise compaction to model condensin II depletion (Fig. 4, C to G; fig. S22; and table S9).

Under high lengthwise compaction (i.e., intact condensin II), chromosomes form nonoverlapping entities and hinder the spatial clustering of centromeres. Correspondingly, lower lengthwise compaction (i.e., impaired condensin II) leads to chromosome intermingling and centromere clustering. This physical model illustrates how the loss of lengthwise compaction might explain the observed clustering of centromeres.

Condensin I and condensin II together drive mitotic chromosome condensation (fig. S23, A and B). In contrast to condensin II, condensin I primarily decreases the width of the

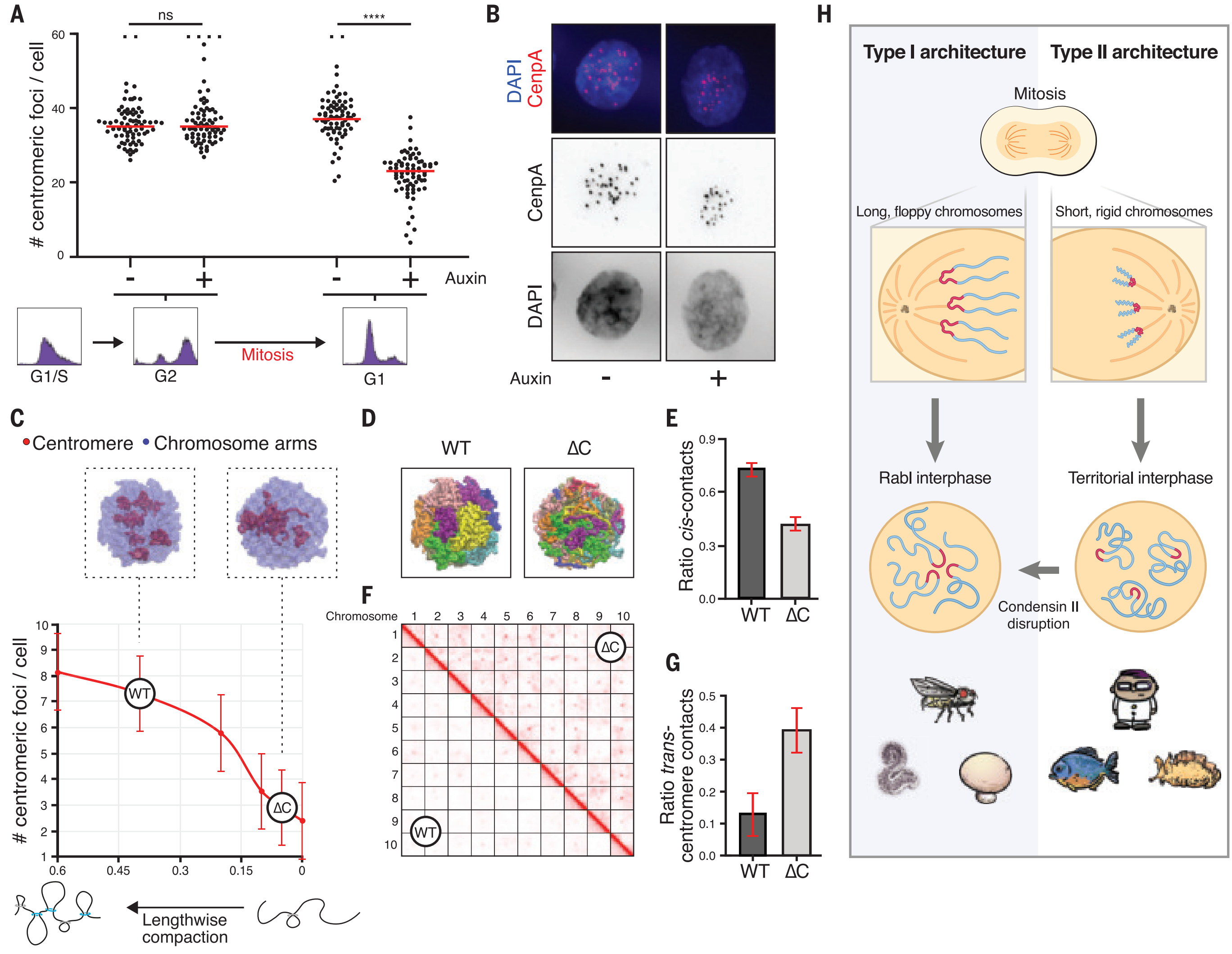


Fig. 4. Centromeric clustering is counteracted by lengthwise compaction and requires mitosis-to-interphase transition. (**A**) Quantification of centromeric foci before or after mitotic progression with or without auxin-mediated condensin II degradation. Fluorescence-activated cell sorting (FACS) plots depict cell cycle stages. Outliers (>60) were truncated and depicted as squares. (**B**) Example images of G_1 cells as quantified in (A). (C to G) Simulation modeling using ten polymer chains as chromosomes. (**C**) Number of centromere clusters upon varying lengthwise compaction (strength of the ideal chromosome term). WT and ΔC correspond to higher and lower lengthwise compaction, recapitulating the experimental data observed in wild type and Δ*CAP-H2* cells. (Top) Representative models for both states. (**D**) Representative simulation snapshots depicting ten chromosomes in different colors. (**E**) Quantification of the ratio of cis contacts. (**F**) Simulated Hi-C matrices depicting contacts between the respective chromosomes. (**G**) Quantification of the proportion of trans-centromeric contacts. (**H**) Model for the establishment of type-I and type-II genome architectures. Having shorter chromosomes during mitosis tends to interfere with adhesion between centromeres, leading to separate centromeres and territorial genome architecture in the subsequent interphase. Reducing lengthwise compaction, for example by condensin II disruption, leads to enhanced centromere clustering, loss of chromosome territories, and a Rabl-like genome architecture. ****$P < 0.0001$; ns, not significant.

chromosome (*19*, *20*). If condensin II–driven lengthwise compaction were the key factor leading to territorialization, rather than chromosome condensation in general, then condensin I depletion would not lead to a shift from territorial to Rabl-like architecture. We found that acute depletion of the condensin I subunit CAP-H did not lead to centromeric clustering (fig. S23, C and D).

Evolution has performed an experiment in which chromosome length varies as a result of chromosome fusions rather than the loss of condensin II. Specifically, the Chinese muntjac has 46 short chromosomes that have merged, in the closely related Indian muntjac, into six chromosomes (in females). By assembling the muntjac genomes, we found that the notable increase in chromosome length in the Indian muntjac coincides, as expected, with the appearance of centromeric clustering (fig. S25).

Taken together, a model emerges in which condensin II establishes interphase 3D genome architecture at the scale of whole chromosomes. We hypothesize that (i) centromeres tend to adhere to one another, a process that is facilitated by proximity during and shortly after mitosis; (ii) the shortening of chromosomes interferes with this adhesion, enabling the centromeres to spread out over the newly formed nuclei; and (iii) chromosome territories emerge as a by-product of the resulting chromosomal separation (Fig. 4H).

The role of condensin II in establishing the overall architecture of the genome appears to be among the most ancient capabilities defining genome folding in the eukaryotic lineage. Changes in condensin II have likely contributed to notable shifts from chromosome territories to Rabl-like features throughout the tree of life. As our exploration of the tree of life continues, one of the many fruits will be a deeper knowledge of our own cellular machinery.

REFERENCES AND NOTES

1. S. S. P. Rao *et al*., *Cell* **159**, 1665–1680 (2014).
2. Materials and methods are available as supplementary materials online.
3. T. Cremer, M. Cremer, *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
4. C. Rabl, *Morphol. Jahrb.* **10**, 214–330 (1885).
5. T. D. King *et al*., *Mol. Biol. Evol.* **36**, 2195–2204 (2019).
6. T. Hirano, *Cell* **164**, 847–857 (2016).
7. K. Kinoshita, T. J. Kobayashi, T. Hirano, *Dev. Cell* **33**, 94–106 (2015).
8. V. Schubert, I. Lermontova, I. Schubert, *Chromosoma* **122**, 517–533 (2013).
9. T. Sakamoto, T. Sugiyama, T. Yamashita, S. Matsunaga, *Nucleus* **10**, 116–125 (2019).
10. K. Nishide, T. Hirano, *PLOS Genet.* **10**, e1004847 (2014).
11. C. R. Bauer, T. A. Hartl, G. Bosco, *PLOS Genet.* **8**, e1002873 (2012).
12. L. F. Rosin, S. C. Nguyen, E. F. Joyce, *PLOS Genet.* **14**, e1007393 (2018).
13. L. Guelen *et al*., *Nature* **453**, 948–951 (2008).
14. J. Padeken *et al*., *Mol. Cell* **50**, 236–249 (2013).
15. L. Brueckner *et al*., *EMBO J.* **39**, e103159 (2020).
16. B. van Steensel, A. S. Belmont, *Cell* **169**, 780–791 (2017).
17. M. Takagi *et al*., *J. Cell Sci.* **131**, jcs212092 (2018).
18. N. Abdennur *et al*., Condensin II inactivation in interphase does not affect chromatin folding or gene expression. bioRxiv 437459 [Preprint]. 7 October 2018. https://doi.org/10.1101/437459.
19. K. Shintomi, T. Hirano, *Genes Dev.* **25**, 1464–1469 (2011).
20. L. C. Green *et al*., *J. Cell Sci.* **125**, 1591–1604 (2012).
21. J. H. Gibcus *et al*., *Science* **359**, eaao6135 (2018).
22. B. Zhang, P. G. Wolynes, *Phys. Rev. Lett.* **116**, 248101 (2016).
23. M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12168–12173 (2016).
24. A. G. Larson *et al*., *Nature* **547**, 236–240 (2017).
25. A. R. Strom *et al*., *Nature* **547**, 241–245 (2017).
26. M. Falk *et al*., *Nature* **570**, 395–399 (2019).
27. C. Hoencamp, Data from: 3D genomics across the tree of life identifies condensin II as a determinant of architecture type, version 1, Harvard Dataverse (2021); https://doi.org/10.7910/DVN/UROKAG.
28. O. Dudchenko, E. Lieberman-Aiden, S. Singh Batra, aidenlab/3d-dna: 3D-DNA Phasing branch 201008, version 201008, Zenodo (2021); http://doi.org/10.5281/zenodo.4619943.
29. T. van den Brand, R. van der Weide, A. Sedeño Cacciatore, M. Schijns, deWitLab/GENOVA v0.94, version v0.94, Zenodo (2021); http://doi.org/10.5281/zenodo.4564568.
30. Á. Sedeño Cacciatore, asedenocacciatore/centromeric_clustering: v0.1, version v0.1, Zenodo (2021); http://doi.org/10.5281/zenodo.4575422.
31. C. Hoencamp *et al*., 3D genomics across the tree of life reveals condensin II as a determinant of architecture type, dataset, Zenodo (2021); http://doi.org/10.5281/zenodo.4582361.
32. O. Dudchenko, dudcha/misc: Analysis of heritability of architectural features, custom script, version Hoencamp, Zenodo (2021); http://doi.org/10.5281/zenodo.4619996.
33. B. van den Broek, bvandenbroek/intensity-distribution-nucleoli: Version 1.0.0, version 1.0.0, Zenodo (2020); http://doi.org/10.5281/zenodo.4350575.

ACKNOWLEDGMENTS

This study is dedicated to the memory of our friend and colleague, José Luis Gómez-Skarmeta. Chinese muntjac cells were kindly provided by W. R. Brinkley. Skin fibroblasts of Indian muntjac were obtained from JCRB (9100). This is a SeaWorld technical manuscript contribution number 2020-12. We thank M. Takagi from the Cellular Dynamics Laboratory at RIKEN for sharing the CAP-H2-AID and the CAP-H-AID cell lines; M. Mertz from the NKI BioImaging Facility for advice on image analyses; the NKI Genomics core facility for sequencing; E. Jaeger and F. Chen (Illumina, Inc.) for fruitful discussions on chromosome-length genome phasing; R. B. Gasser and P. K. Korhonen, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, for their help with the oriental liver fluke sample; J. Alfoldi, J. Johnson, A. Berlin, S. Gnerre, D. Jaffe, I. MacCallum, S. Young, B. J. Walker, J. L. Chang, and E. S. Lander at the Broad Institute Genome Assembly & Analysis Group, Computational R&D Group, and Sequencing Platform for their contribution to the *Aplysia californica* draft genome assembly project; N. Watanabe at Kyoto University for providing the African clawed frog cell line to the Kornberg laboratory; V. Patel at Baylor Genetics for help with sequencing; S. Knemeyer and V. Yeung (SciStories, LLC) for help with Fig. 4H; and A. Fotos from adamfotos.com for help with Fig. 1, Fig. 4H, and figs. S1 and S25. **Funding:** C.H. is supported by the Boehringer Ingelheim Fonds; C.H., Á.S.C., and B.D.R. are supported by an ERC CoG (772471, "CohesinLooping"); A.M.O.E. and B.D.R. are supported by the Dutch Research Council (NWO-Echo); and J.A.R. and R.H.M. are supported by the Dutch Cancer Society (KWF). T.v.S. and B.v.S. are supported by NIH Common Fund "4D Nucleome" Program grant U54DK107965. H.T. and E.d.W. are supported by an ERC StG (637597, "HAP-PHEN"). J.A.R., T.v.S., H.T., R.H.M., B.v.S., and E.d.W. are part of the Oncode Institute, which is partly financed by the Dutch Cancer Society. Work at the Center for Theoretical Biological Physics is sponsored by the NSF (grants PHY-2019745 and CHE-1614101) and by the Welch Foundation (grant C-1792). V.G.C. is funded by FAPESP (São Paulo State Research Foundation and Higher Education Personnel) grants 2016/13998-8 and 2017/09662-7. J.N.O. is a CPRIT Scholar in Cancer Research. E.L.A. was supported by an NSF Physics Frontiers Center Award (PHY-2019745), the Welch Foundation (Q-1866), a USDA Agriculture and Food Research Initiative grant (2017-05741), the Behavioral Plasticity Research Institute (NSF DBI-2021795), and an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375). Hi-C data for the 24 species were created by the DNA Zoo Consortium (www.dnazoo.org). DNA Zoo is supported by Illumina, Inc.; IBM; and the Pawsey Supercomputing Center. P.K. is supported by the University of Western Australia. L.L.M. was supported by NIH (1R01NS114491) and NSF awards (1557923, 1548121, and 1645219) and the Human Frontiers Science Program (RGP0060/2017). The draft *A. californica* project was supported by NHGRI. J.L.G.-S. received funding from the ERC (grant agreement no. 740041), the Spanish Ministerio de Economía y Competitividad (grant no. BFU2016-74961-P), and the institutional grant Unidad de Excelencia María de Maeztu (MDM-2016-0687). R.D.K. is supported by NIH grant R01DK121366. V.H. is supported by NIH grant NIH1P41HD071837. K.M. is supported by a MEXT grant (20H05936). M.C.W. is supported by the NIH grants R01AG045183, R01AT009050, R01AG062257, and DP1DK113644 and by the Welch Foundation. E.F. was supported by NHGRI (grant no. DP2HG010013). M.C.W. is a Howard Hughes Medical Institute Investigator. **Author contributions:** C.H., A.M.O.E., J.A.R., T.v.S., R.G.H.P.v.H., A.N.M., and H.T. performed wet laboratory experiments in human cells, and T.v.S., Á.S.C., B.v.d.B., and E.d.W. performed bioinformatic analysis for these experiments. C.H. performed condensin conservation analyses. Samples for comparative genomics and assembly have been provided by L.L.M., A.K., J.S.L., A.M., K.H., M.C.G., F.L., E.F., N.S., A.H., D.C., A.S.Mut., M.C.W., N.D.Y., E.H., H.H.C., C.J.K., T.L.U.B., K.A.H., A.J.B., P.-J.M., K.M., A.S.Mul., S.P., L.N.-T., and R.R.B., and the corresponding experiments were performed by B.G.S.H., A.D.O., M.P., Z.C., N.M., C.L., W.Y., R.K., P.K., and O.D. Assembly and comparative data analysis were performed by O.D., B.G.S.H., D.W., Z.Y., S.S.P.R., A.J.B., P.-J.M., R.D.K., W.C.W., G.C., J.L.G.-S., V.H., K.L.-T., F.D.P., P.K., and E.L.A. Chromosome-length phasing was done by O.D., D.W., and E.L.A. S.B. and V.G.C. performed simulations. R.H.M., B.v.S., E.d.W., J.N.O., M.D.P., E.L.A., and B.D.R. provided supervision. C.H., O.D., A.M.O.E., S.B., M.D.P., E.L.A., and B.D.R. wrote the original and revised drafts with input from all authors. **Competing interests:** E.L.A., O.D., B.G.S.H., and M.P. are inventors on US provisional patent application 16/308,386, filed 7 December 2018, by the Baylor College of Medicine and the Broad Institute, relating to the assembly methods in this manuscript. E.L.A. and O.D. are inventors on US provisional patent application 62/617,289, filed 14 January 2018, and US provisional patent application 16/247,502, filed 14 January 2019, by the Baylor College of Medicine and the Broad Institute, relating to the assembly methods in this manuscript. E.L.A. is an inventor on US provisional patent application PCT/US2020/064704 filed 11 December 2020, by the Baylor College of Medicine and the Broad Institute, relating to the assembly methods in this manuscript. E.L.A. is Scientific Advisory Board co-chair and consultant for HolyHaid Lab Corporation (Shenzhen, China), whose parent company is Hollyhigh International Capital (Beijing and Shanghai, China). The other authors declare no competing interests. **Data and materials availability:** Sequencing data have been deposited in GEO, accession numbers GSE163641 and GSE169088. Additionally, interactive contact maps for species assembled in this study are available at www.dnazoo.org. Alignments from the conservation analysis have been deposited in Harvard Dataverse (*27*). The Hap1 cell lines are available from B.D.R. under a material transfer agreement with the Netherlands Cancer Institute. The 3D-DNA genome assembly and phasing tools are available on Zenodo (*28*), as well as code for downstream Hi-C data analysis (*29*, *30*). Our molecular simulation package and sample trajectory files can also be found on Zenodo (*31*) along with additional custom scripts for phylogenetic analysis and microscopy image processing (*32*, *33*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6545/984/suppl/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S10
References (*34–114*)
MDAR Reproducibility Checklist

19 August 2020; accepted 16 April 2021
10.1126/science.abe2218

CORONAVIRUS

Resurgence of SARS-CoV-2: Detection by community viral surveillance

Steven Riley[1,2]*, Kylie E. C. Ainslie[1,2], Oliver Eales[1,2], Caroline E. Walters[1,2], Haowei Wang[1,2], Christina Atchison[1], Claudio Fronterre[3,4], Peter J. Diggle[3,4], Deborah Ashby[1], Christl A. Donnelly[1,2,5], Graham Cooke[6,7,8], Wendy Barclay[6], Helen Ward[1,7,8], Ara Darzi[7,8,9], Paul Elliott[1,7,8,10]*

Surveillance of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has mainly relied on case reporting, which is biased by health service performance, test availability, and test-seeking behaviors. We report a community-wide national representative surveillance program in England based on self-administered swab results from ~594,000 individuals tested for SARS-CoV-2, regardless of symptoms, between May and the beginning of September 2020. The epidemic declined between May and July 2020 but then increased gradually from mid-August, accelerating into early September 2020 at the start of the second wave. When compared with cases detected through routine surveillance, we report here a longer period of decline and a younger age distribution. Representative community sampling for SARS-CoV-2 can substantially improve situational awareness and feed into the public health response even at low prevalence.

Ahead of widespread rollout of effective vaccines in most countries (*1*–*3*), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection continues to cause substantial COVID-19 morbidity and mortality globally (*4*). As variants with potentially increased transmissibility emerge (*5*), populations around the world continue to trade off between social interactions and risk of infection (*6*). However, reduced social contact (*7*) has adverse effects on levels of economic activity (*8*), non-COVID-19–related health, and overall well-being (*9*). The ability of both individuals and governments to continue to balance these competing demands requires accurate and timely knowledge of the spread of the virus in the population so that informed choices about interventions can be made.

Data streams based on respiratory symptoms, such as those used for COVID-19 surveillance in most countries, are prone to biases that can obscure underlying trends, such as variations in test availability and test-seeking behavior (*10*). Some countries have augmented these systems with surveys of virus prevalence in the wider population, but these have mostly been one-off activities, for example, as in Wuhan, China (*11*), or were designed explicitly as interventions, for example, as in Slovakia (*12*). Here we show results from the Real-time Assessment of Community Transmission-1 (REACT-1) study, a representative community-wide program that is tracking prevalence of SARS-CoV-2 across England through repeated random population-based sampling (*13*). This program was designed to rapidly detect resurgence of SARS-CoV-2 transmission, including at low prevalence, thus providing early warning of any upturn in infections, which can help with policy response and enable timely implementation of public health interventions.

Over the course of four rounds, from 1 May to 8 September 2020, we invited more than 2.4 million people to join the study, from whom we obtained ~596,000 tested swabs (Table 1) for an overall response rate of ~25% (table S1). Between round 1 (1 May to 1 June 2020) and round 2 (19 June to 7 July) there was a fall in weighted prevalence from 0.16% (95% confidence interval: 0.12%, 0.19%) to 0.088% (0.068%, 0.11%) (Table 1 and Fig. 1). Infections fell further, to their lowest observed value, in round 3 (24 July to 11 August), with 54 positive samples out of 161,560 swabs, giving a weighted prevalence of 0.040% (0.027%, 0.053%). In comparison, a 100-fold higher prevalence of ~5% was seen at the peak of the first UK wave, based on a daily incidence of infection in the UK of >300,000 (*14*) and assuming that individuals would test swab-positive for ~10 days on average (*15*). Prevalence then increased in round 4 (20 August to 8 September), where we found 137 positive samples out of 154,325 swabs, giving a weighted prevalence of 0.13% (0.10%, 0.15%).

Using a model of constant exponential growth and decay (*16*), we quantified this fall and rise in prevalence in terms of halving and doubling times and reproduction number *R* (Fig. 1 and Table 2). Over rounds 2 and 3 (19 June to 11 August), prevalence fell with an estimated halving time of 27 days (95% credible intervals: 20, 42) corresponding to an *R* value of 0.85 (0.79, 0.90). Prevalence then increased over rounds 3 and 4 (24 July to 8 September), with a doubling time of 17 (14, 23) days corresponding to an *R* value of 1.28 (1.20, 1.36). Our estimates of *R* and doubling times were similar in sensitivity analyses among nonsymptomatic people [average 72% (95% confidence interval: 67%, 76%)] or those

Table 1. Unweighted and weighted prevalence (95% confidence interval) of swab positivity across four rounds of REACT-1.

Parameter	Round 1	Round 2	Round 3	Round 4
First sample	1 May 2020	19 June 2020	24 July 2020	20 August 2020
Last sample	1 June 2020	7 July 2020	11 August 2020	8 September 2020
Recruitment letters sent	395,000	600,000	710,000	710,000
Swabs sent	161,497	219,633	225,615	211,291
Tested swabs	120,620	159,199	161,560	154,325
Swab response rate	75%	72%	72%	73%
Overall response rate	31%	27%	23%	22%
Positive swabs	159	123	54	137
Unweighted prevalence (95% CI)	0.132% (0.113%, 0.154%)	0.077% (0.065%, 0.092%)	0.033% (0.025%, 0.043%)	0.089% (0.075%, 0.105%)
Weighted prevalence (95% CI)	0.156% (0.124%, 0.188%)	0.088% (0.068%, 0.109%)	0.040% (0.027%, 0.053%)	0.125% (0.096%, 0.154%)

---

[1]School of Public Health, Imperial College London, London, UK. [2]MRC Centre for Global Infectious Disease Analysis and Abdul Latif Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, UK. [3]Centre for Health Informatics, Computing, and Statistics (CHICAS), Lancaster Medical School, Lancaster University, Lancaster, UK. [4]Health Data Research UK, London, UK. [5]Department of Statistics, University of Oxford, Oxford, UK. [6]Department of Infectious Disease, Imperial College London, London, UK. [7]Imperial College Healthcare NHS Trust, London, UK. [8]National Institute for Health Research Imperial Biomedical Research Centre, London, UK. [9]Institute of Global Health Innovation, Imperial College London, London, UK. [10]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK.
***Corresponding author. Email: s.riley@imperial.ac.uk (S.R.); p.elliott@imperial.ac.uk (P.E.)**

positive for both the envelope protein (E) gene and nucleocapsid protein (N) gene (table S2).

We compared epidemic trends estimated from the REACT-1 data above with those based on routine surveillance data (Fig. 1, figs. S1 and S2, and Table 2) over the same period. Numbers of routine surveillance cases were growing from the start of round 2 to the end of round 3 (19 June to 11 August), with a corresponding *R* of 1.05 (1.02, 1.07) (Table 2), when swab positivity was declining in REACT-1. *R* estimates from routine surveillance data were likely upwardly biased, because there was a near-doubling of test capacity during this period (*17*) (fig. S1). These findings are consistent with experience in the UK during the 2009 influenza pandemic, when there were substantial temporal variations in the sensitivity of case-based polymerase chain reaction surveillance (*18*).

We also observed an apparent shift from decline to growth using within-round data (fig. S3 and Table 2). During round 3 (24 July to 11 August), with 94% probability, the epidemic had started to grow with a doubling time of 14 days (95% credible interval: from halving every 59 days to doubling every 6.4 days), corresponding to an *R* of 1.34 (0.93, 1.83) (Fig. 1 and Table 1). During round 4 (20 August to 8 September), the doubling time decreased to 8.0 (5.7, 14) days, with an *R* of 1.64 (1.35, 1.95). In response to the rapidly increasing epidemic, the UK government announced a more stringent social distancing measure called the “rule of six,” prohibiting gatherings of more than six people (*19*).

We relaxed our assumption of constant growth or decay using a flexible p-spline (*16*) (fig. S1) and inferred a plateau or slight increase in prevalence in July 2020 in the gap between rounds 2 and 3. As a result, the prevalence for round 3 started higher than

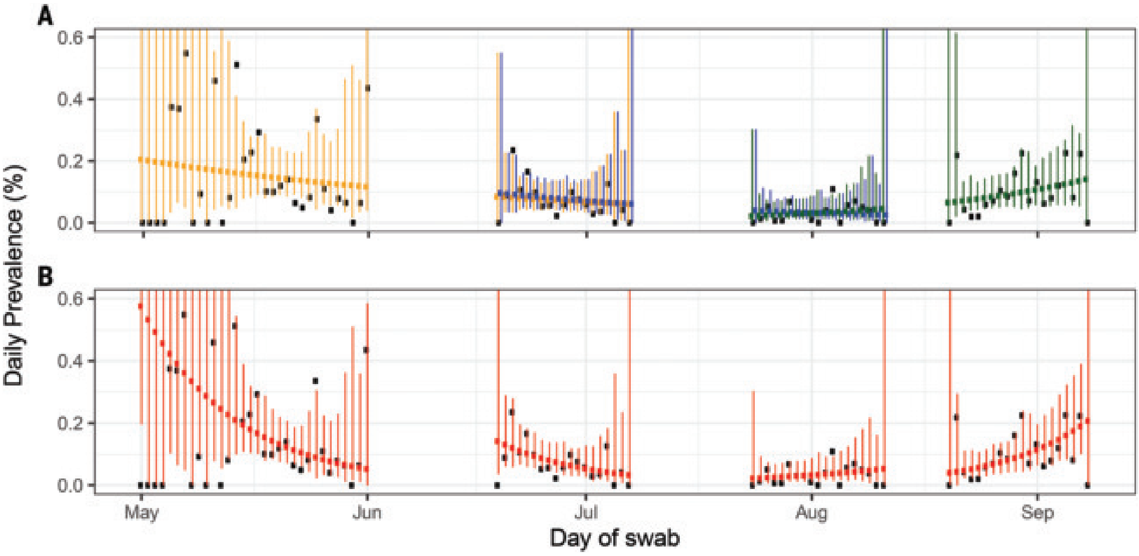


Fig. 1. Constant growth rate models fit to REACT-1 data for sequential and individual rounds. (**A**) Model fits to REACT-1 data for sequential rounds 1 and 2 (yellow), 2 and 3 (blue), and 3 and 4 (green). Vertical lines show 95% prediction intervals for models. Black points show observations. See Table 1 for *R* estimates. (**B**) Models fit to individual rounds only (red). Note that only 585,004 of 596,965 tests had dates available and were included in the analysis (465 out of 473 positives were included).

Table 2. Fitted growth rates, reproduction numbers, and doubling times (95% credible intervals) for SARS-CoV-2 swab positivity in England.

Data	Round(s)*	Number of participants or cases	Growth rate *r* (per day)	*P* (*r* > 0)	Reproduction number	Doubling (+) or halving (–) time (days)
REACT All	1	110,944	–0.077 (–0.107, –0.046)	<0.01	0.57 (0.44, 0.73)	–9.0 (–6.5, –14.9)
	2	157,428	–0.089 (–0.130, –0.032)	<0.01	0.52 (0.36, 0.81)	–7.8 (–5.3, –21.4)
	3	162,619	0.049 (–0.012, 0.109)	0.94	1.34 (0.93, 1.83)	14.2 (–58.6, 6.4)
	4	153,964	0.086 (0.050, 0.122)	>0.99	1.64 (1.35, 1.95)	8.0 (13.8, 5.7)
	1 and 2	268,422	–0.018 (–0.025, –0.012)	<0.01	0.89 (0.85, 0.92)	–37.9 (–28.0, –57.5)
	2 and 3	320,047	–0.025 (–0.034, –0.017)	<0.01	0.85 (0.79, 0.90)	–27.3 (–20.1, –41.6)
	3 and 4	316,583	0.041 (0.030, 0.051)	>0.99	1.28 (1.20, 1.36)	17.0 (22.8, 13.5)
Routine surveillance data	1	69,299	–0.034 (–0.042, –0.027)	<0.01	0.80 (0.75, 0.84)	–20.3 (–16.7, –26.0)
	2	11,523	–0.018 (–0.032, –0.004)	0.01	0.89 (0.81, 0.97)	–38.5 (–21.9, –164.1)
	3	15,172	0.026 (0.004, 0.049)	0.99	1.18 (1.03, 1.34)	26.2 (161.4, 14.2)
	4	31,209	0.085 (0.067, 0.104)	>0.99	1.63 (1.48, 1.79)	8.1 (10.3, 6.7)
	1 and 2	97,255	–0.029 (–0.031, –0.027)	<0.01	0.82 (0.81, 0.84)	–23.7 (–22.0, –25.7)
	2 and 3	36,393	0.007 (0.003, 0.012)	>0.99	1.05 (1.02, 1.07)	93.6 (207.6, 60.6)
	3 and 4	56,064	0.029 (0.023, 0.035)	>0.99	1.19 (1.15, 1.24)	24.1 (30.7, 20.1)

*See Table 1 for start and end dates of rounds

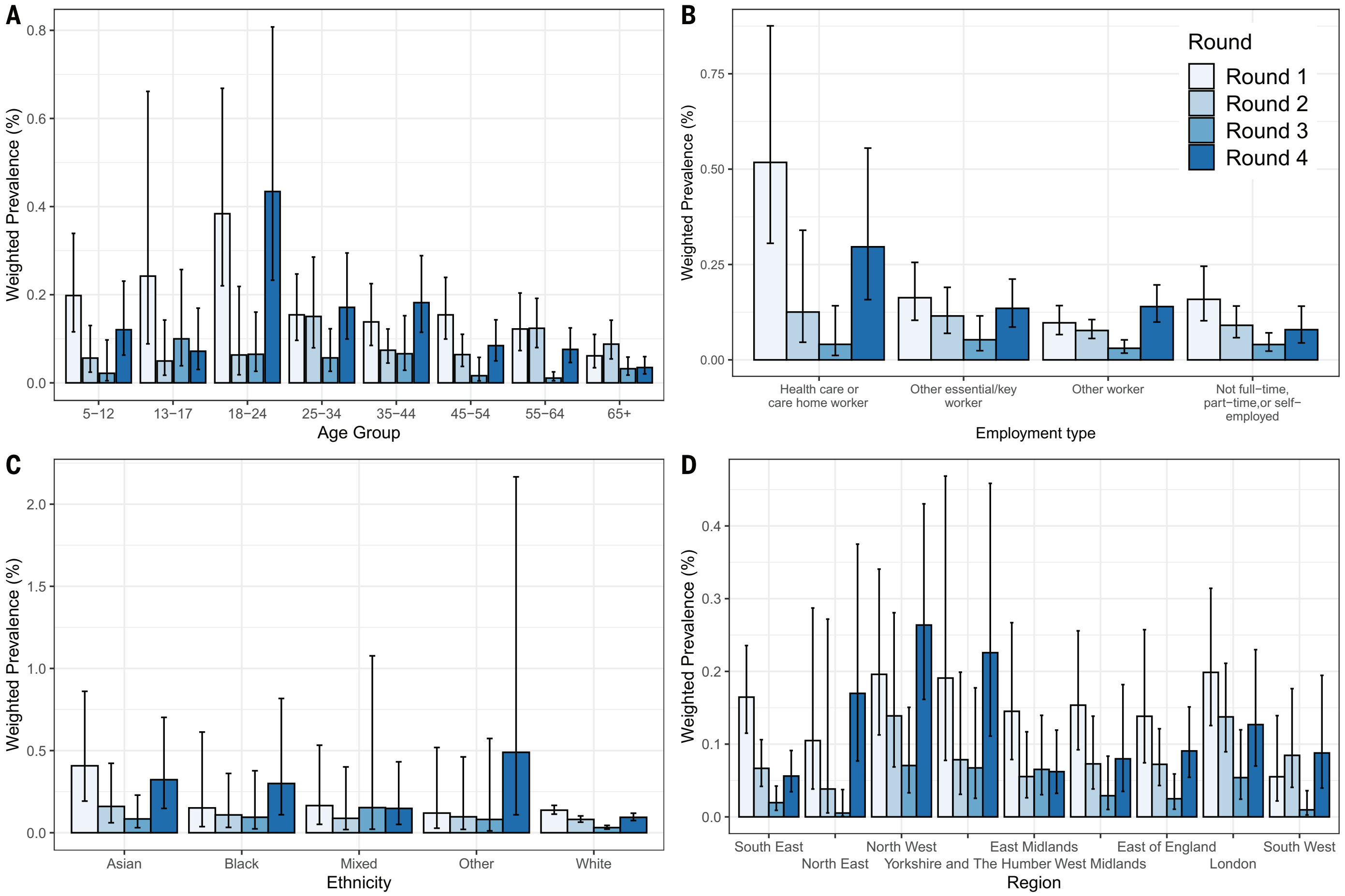


Fig. 2. Prevalence of unweighted swab positivity. Covering four rounds of the REACT-1 study by (**A**) age, (**B**) employment type, (**C**) ethnicity, and (**D**) region. Vertical bars show 95% confidence intervals. Rounds are differentiated by color.

expected from the data observed at the end of round 2, a pattern similar to that seen in data from the Office for National Statistics Coronavirus (COVID-19) Infection Survey pilot (*20*). Using the p-spline, we estimated that lowest prevalence occurred on 20 July (13 July, 15 August) (fig. S3), compared with 5 July (30 June, 16 July), as estimated from the routine surveillance data, likely reflecting the rapid increase in testing capacity (fig. S3).

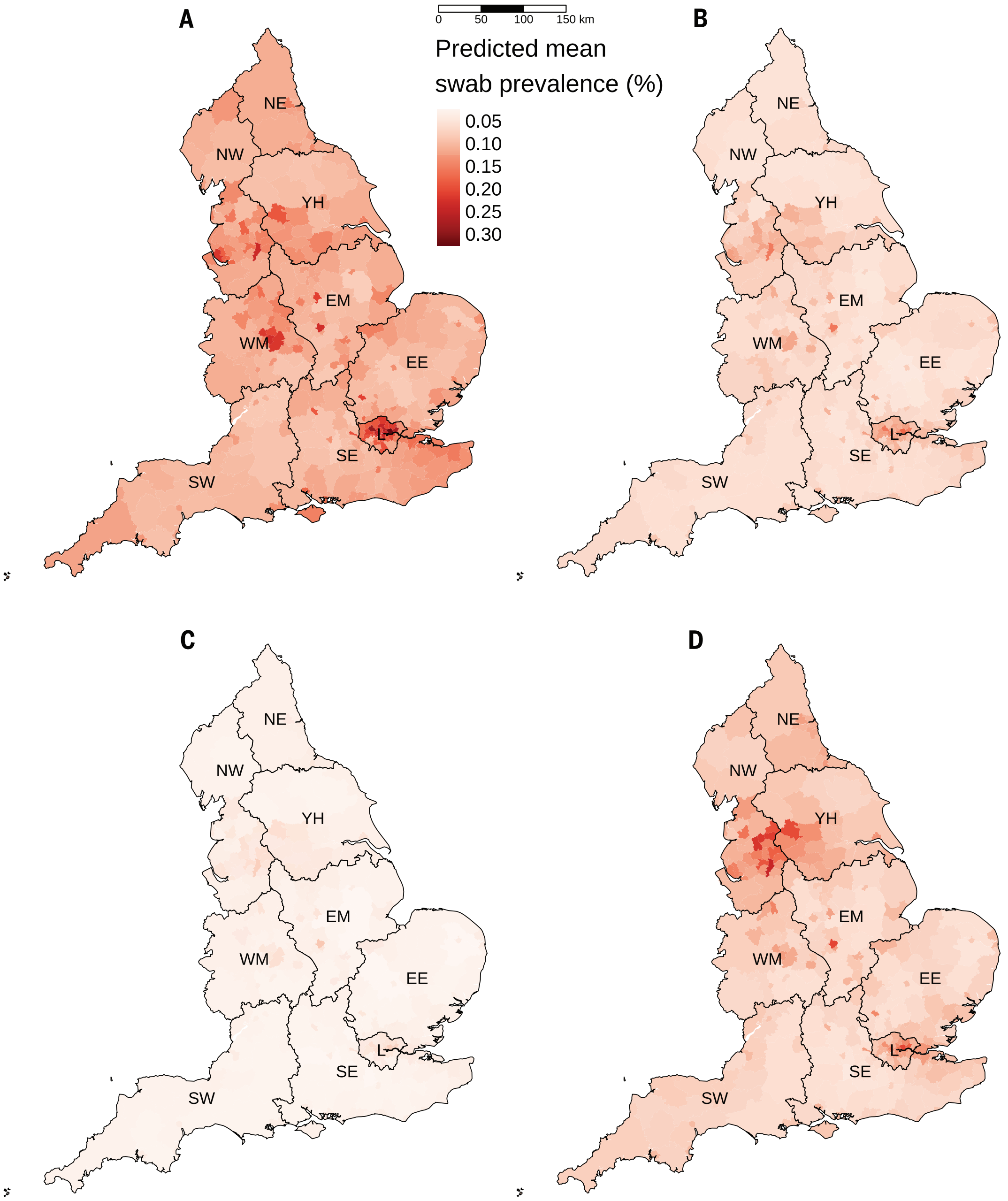
During March and April, the highest prevalence regionally was recorded in London, which experienced the highest incidence of cases during the first wave (*21*, *22*). Prevalence fell in all regions between round 1 (1 May to 1 June) and round 3 (24 July to 11 August). There was then positive growth (>95% probability) between round 3 and round 4 (20 August to 8 September) in all regions except East and West Midlands (table S3 and figs. S4 and S5), with the highest growth in the North East region [R = 1.67 (1.20, 2.48)]. During round 4 (20 August to 8 September), we observed about a threefold difference between the highest prevalence in both the North West region at 0.17% (0.12%, 0.24%) and Yorkshire and the Humber at 0.17% (0.11%, 0.27%) and the lowest at 0.06% (0.04%, 0.09%) in the South East region (Fig. 2, table S4, and fig. S4).

We found spatial heterogeneity in prevalence at a subregional level using a geospatial model (*16*) with a range parameter estimate of 22.6 km (95% confidence interval: 16.1, 31.7) (Fig. 3 and table S5). We observed areas of higher prevalence in parts of the North West region, Yorkshire and the Humber, Midlands, and the London conurbation in round 1 (1 May to 1 June). These patterns persisted at lower prevalence in round 2 (19 June to 7 July) before reaching lowest prevalence in round 3 (24 July to 11 August). The epidemic then resurged in round 4 (20 August to 8 September), with geographical patterns similar to those seen in rounds 1 and 2 and an indication that prevalence in each local area had increased between rounds 3 and 4 (fig. S5).

Our findings show substantial variations in age patterns over time. In round 4 (20 August to 8 September), the highest prevalence at 0.25% (0.15%, 0.41%) was found in participants aged 18 to 24 years, increasing more than threefold from 0.08% (0.04%, 0.18%) in round 3 (24 July to 11 August) (Fig. 2 and table S4). The lowest prevalence at 0.04% (0.02%, 0.06%) was in those aged 65 years and older, similar to round 3. These patterns suggest that the second wave started in young adults–likely driven by higher numbers of social contacts (*23*)–before spreading into older (*22*, *24*) and more at-risk populations (*25*).

We compared age patterns from REACT-1 with those in the routine surveillance case incidence data (*17*); in each dataset, we estimated odds ratios for each age group (35 to 44 years as comparator) (fig. S6). We found that the symptomatic case data in round 1 (1 May to 1 June) overestimated odds at older ages and underestimated odds at younger ages relative to REACT-1, reflecting the limited availability of symptomatic testing at that time, when testing was carried out mainly among hospitalized patients (*17*). In subsequent rounds, the case data consistently underestimated

Fig. 3. Geospatial patterns. Estimated prevalence from geospatial model for (**A**) round 1, (**B**) round 2, (**C**) round 3, and (**D**) round 4. Regions: NE, North East; NW, North West; YH, Yorkshire and the Humber; EM, East Midlands; WM, West Midlands; EE, East of England; L, London; SE, South East; SW, South West.



odds at ages 5 to 14 years, while odds at older ages continued to be overestimated relative to REACT-1. Similar biases in case data may have contributed to reports of reduced susceptibility to infection in younger children (*26*).

We found differences over time in the odds of infection for health care and care home workers, with odds of 5.5 (3.1, 9.7) relative to other workers during round 1 (1 May to 1 June) but much-reduced odds in subsequent rounds (table S6). These findings indicate that there was a shift away from rapid transmission in hospitals (*27*) and care homes (*28*) during the first wave to predominantly community transmission at the start of the second wave.

We found about a twofold greater unweighted prevalence of swab positivity in participants of Asian ethnicity (mainly South Asian) at 0.14% (0.10%, 0.20%), compared with 0.07% (0.07%, 0.08%) in white participants across all four rounds combined (table S4); odds were 2.2 (1.2, 4.0) relative to white participants in round 4 (20 August to 8 September), with multiple adjustment (table S6). There was also a higher unadjusted prevalence of infection in Black people compared with white people across all four rounds combined at 0.15% (0.09%, 0.27%) (table S4). These higher rates of swab positivity are consistent with higher SARS-CoV-2 seroprevalence among Asian and Black people and people of other nonwhite ethnicity in England (*22*). This supports the view that higher rates of hospitalization and mortality from COVID-19 reported among minority ethnic groups in England (*29*) reflect their higher rates of infection rather than a poorer prognosis once infected.

Although we aimed to be representative of the population of England by inviting a random sample of people on the National Health Service patient register (*16*), we found differential response rates by age, area, and round. For example, response rates ranged from 21.8% in round 4 (20 August to 8 September) to 30.8%

in round 1 (1 May to 1 June) and differed across age groups, from 10.7% for ages 18 to 24 years to 31.1% for ages 55 to 64 years (round 4) (table S1). However, unlike the symptomatic testing, we were able to correct for variations in response given that we have a known denominator. We were thus able to estimate prevalence weighted to the population of England as a whole, taking into account sample design and nonresponse, although we did not reweight prevalence estimates for subgroups, because of lower numbers of positives.

We converted growth rates into reproduction numbers using serial interval parameters from (*30*). However, we also tested the sensitivity of our results to a wide range of other published estimates (table S7). We found that by using (*30*) our estimates of *R* above 1 were conservative and that using other published parameters lowered our *R* estimates. The converse was true for *R* values less than 1; estimates using (*30*) were lower than those using results from other studies. Essentially, uncertainty in our estimates of *R* reflect uncertainty in our estimate of the growth rate and do not propagate uncertainty about the serial interval present in the literature.

We relied on self-swabbing to obtain estimates of swab positivity. A throat and nose swab is estimated to have between ~70 and ~80% sensitivity (*31*), so we are likely to have underestimated true prevalence, although this would be unlikely to have affected trend analyses or estimation of *R*. During the period of our study, there was changing availability of symptom-driven test capacity, which likely explains the earlier increase in swab positivity in the symptomatic data compared with our own data (*17*). The trends in our data were supported by results of analyses among the subset of nonsymptomatic individuals, who would not have presented to the national case-testing program (table S2).

Our study provides timely community-based prevalence data to increase situational awareness and inform the public health response during the current SARS-CoV-2 pandemic. The scenario of declining prevalence to low levels followed by resurgence reported here may reoccur in the future in the absence of protective population immunity; this depends on levels of vaccine coverage of the population (*32*), degree of waning of natural immunity and vaccine efficacy (*33*), and potential for antigenic escape (*34*). Also, as of early 2021, some populations have successfully avoided large waves of infection but may not be able to do so in the future because of intervention fatigue or increased transmissibility of the virus (*35*).

Accurate estimates of prevalence with robust descriptions of trends by time, person, and place would support sustainable policies designed to maintain low levels of prevalence. Unlike China, New Zealand, and Australia, the UK did not attempt functional elimination (so-called COVID-zero) during periods of low prevalence in February or August 2020, in common with all other European nations. However, with the rollout of effective vaccines from December 2020 (*36*) and with accumulating evidence of antigenic change (*37*), the cost-benefit assessment of policies designed to achieve sustained low levels of prevalence may be different in the future. For example, during the declining phase, prevalence may be high in some areas because of low vaccine uptake, variant emergence, or increased social mixing. Data from REACT-1 or similar studies could be used to target local public health or vaccination campaigns more effectively than would be possible with routine surveillance data alone, similar to how REACT-1 results fed into the government policy of the rule of six in early September 2020 (*19*).

Additionally, knowledge from community-based surveillance can be used to calibrate other data streams, not only symptomatic testing (*38*) but also the use of mobility data (*39*) and sewage-based sampling of viral RNA (*40*). Given the different spatial and temporal resolutions of alternate data sources, ground-truth data such as those from REACT-1 can substantially improve evidence synthesis for infectious disease (*41*).

We demonstrate the capability of a large national community surveillance program to detect a resurgence of SARS-CoV-2 infection at low prevalence. Our findings have implications for policies to contain the COVID-19 pandemic. While we wait for the vaccination of all risk groups in England and across the world, control of the SARS-CoV-2 virus must continue to rely on established public health measures (*42*), including social distancing, frequent handwashing, face coverings, and an effective test, trace, and isolate system. Although we show high levels of effectiveness of stringent social distancing during the first lockdown in England, prevalence subsequently increased. This perhaps reflects holiday travel, return to work, or a more general increase in the number and transmission potential of social interactions, with a rapid rise evident in early September 2020 at the start of the second wave. A combination of vaccination, social distancing, and other public health measures should again result in substantial reductions in prevalence. Studies similar to REACT-1 could then detect any upturn in prevalence and help trigger an effective public health response.

REFERENCES AND NOTES

1. E. J. Anderson *et al.*, *N. Engl. J. Med.* **383**, 2427–2438 (2020).
2. M. Voysey *et al.*, *Lancet* **397**, 99–111 (2021).
3. F. P. Polack *et al.*, *N. Engl. J. Med.* **383**, 2603–2615 (2020).
4. Johns Hopkins University, Coronavirus Resource Center; https://coronavirus.jhu.edu/.
5. Public Health England, New SARS-CoV-2 variant: Information on the new variants of the SARS-CoV-2 virus, www.gov.uk; www.gov.uk/government/collections/new-sars-cov-2-variant.
6. J. Zhang *et al.*, *Science* **368**, 1481–1486 (2020).
7. C. I. Jarvis *et al.*, *BMC Med.* **18**, 124 (2020).
8. S. Luo, K. P. Tsang, *Contemp. Econ. Policy* **38**, 583–592 (2020).
9. R. Layard *et al*, When to release the lockdown? A wellbeing framework for analysing costs and benefits, IZA Discussion Paper No. 13186, SSRN (2020); https://ssrn.com/abstract=3590884.
10. R. Moss, A. E. Zarebski, S. J. Carlson, J. M. McCaw, *Trop Med Infect Dis* **4**, 12 (2019).
11. S. Cao *et al.*, *Nat. Commun.* **11**, 5917 (2020).
12. M. Pavelka *et al.*, *Science* eabf9648 (2021).
13. S. Riley *et al.*, *Wellcome Open Res.* **5**, 200 (2020).
14. S. Flaxman *et al.*, *Nature* **584**, 257–261 (2020).
15. L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, J. Lessler, *Ann. Intern. Med.* **173**, 262–267 (2020).
16. Materials and methods are available as supplementary materials.
17. UK Government, Coronavirus (COVID-19) in the UK dashboard, www.gov.uk; https://coronavirus.data.gov.uk/.
18. I. Dorigatti, S. Cauchemez, N. M. Ferguson, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13422–13427 (2013).
19. Home Office, Rt Hon Priti Patel MP, "Rule of six comes into effect to tackle coronavirus," www.gov.uk, 14 September 2020; www.gov.uk/government/news/rule-of-six-comes-into-effect-to-tackle-coronavirus.
20. Office for National Statistics, "Coronavirus (COVID-19) Infection Survey pilot: England, 9 July 2020," 9 July 2020; www.ons.gov.uk/releases/coronaviruscovid19infectionsurveypilotengland9july2020.
21. Office for National Statistics, "Deaths registered weekly in England and Wales, provisional: week ending 4 September 2020," 15 September 2020; www.ons.gov.uk/releases/deathsregisteredinenglandandwalesprovisionalweekending4september2020.
22. H. Ward *et al.*, *Nat. Commun.* **12**, 905 (2021).
23. J. Mossong *et al.*, *PLOS Med.* **5**, e74 (2008).
24. Office for National Statistics, "Coronavirus (COVID-19) Infection Survey, UK: 8 January 2021," 8 January 2021; www.ons.gov.uk/releases/coronaviruscovid19infectionsurveyuk.
25. R. Verity *et al.*, *Lancet Infect. Dis.* **20**, 669–677 (2020).
26. N. G. Davies *et al.*, *Nat. Med.* **26**, 1205–1211 (2020).
27. L. Rivett *et al.*, *eLife* **9**, e58728 (2020).
28. N. S. N. Graham *et al.*, *J. Infect.* **81**, 411–419 (2020).
29. Office for National Statistics, "Coronavirus (COVID-19) related deaths by ethnic group, England and Wales: 2 March 2020 to 10 April 2020," 7 May 2020; www.ons.gov.uk/releases/coronavirusrelateddeathsbyethnicgroupenglandandwales2march2020to10april2020.
30. Q. Bi *et al.*, *Lancet Infect. Dis.* **20**, 911–919 (2020).
31. B. Böger *et al.*, *Am. J. Infect. Control* **49**, 21–29 (2021).
32. P. Peretti-Watel *et al.*, *Lancet Infect. Dis.* **20**, 769–770 (2020).
33. M. Jeyanathan *et al.*, *Nat. Rev. Immunol.* **20**, 615–632 (2020).
34. C. K. Wibmer *et al.*, *Nat. Med.* **27**, 622–625 (2021).
35. J. Wise, *BMJ* **371**, m4857 (2020).
36. E. Mahase, *BMJ* **371**, m4714 (2020).
37. A. J. Greaney *et al.*, *Cell Host Microbe* **29**, 463–476.e6 (2021).
38. D. A. Drew *et al.*, *Science* **368**, 1362–1367 (2020).
39. B. Jeffrey *et al.*, *Wellcome Open Res.* **5**, 170 (2020).
40. I. Bar-Or *et al.*, medRxiv 2020.04.26.20073569 [Preprint]. 1 May 2020. https://doi.org/10.1101/2020.04.26.20073569.
41. E. van Leeuwen, P. Klepac, D. Thorrington, R. Pebody, M. Baguelin, *PLOS Comput. Biol.* **13**, e1005838 (2017).
42. Cabinet Office, "Coronavirus (COVID-19): What has changed – 9 September," www.gov.uk, 9 September 2020; www.gov.uk/government/news/coronavirus-covid-19-what-has-changed-9-september.
43. S. Riley *et al.*, reactidd R package with data, Version 0.91, Zenodo (2021); http://doi.org/10.5281/zenodo.4681283.

ACKNOWLEDGMENTS

We thank key collaborators on this work: K. Beaver, S. Clemens, G. Welch, A. Cleary, K. Ward, and K. Pickering at Ipsos MORI; G. Fontana, S. Satkunarajah, and L. Naar at the Institute of Global Health Innovation at Imperial College; D. Fecht at the MRC Centre for Environment and Health, Imperial College London; and G. Taylor at the Molecular Diagnostic Unit, Imperial College London. We acknowledge North West London Pathology and Public Health

England for help with calibration of the laboratory analyses; NHS Digital for access to the NHS register; and the Department of Health and Social Care for logistic support. S.R. acknowledges helpful discussion with members of the UK Government Office for Science (GO-Science) Scientific Pandemic Influenza Group on Modelling (SPI-M) committee. **Funding:** This research was funded by the Department of Health and Social Care in England. S.R. and C.A.D. acknowledge support from the MRC Centre for Global Infectious Disease Analysis, the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU), the Wellcome Trust (200861/Z/16/Z, 200187/Z/15/Z), and the US Centers for Disease Control and Prevention (U01CK0005-01-02). G.C. is supported by an NIHR Professorship. P.E. is director of the MRC Centre for Environment and Health (MR/L01341X/1, MR/S019669/1) and receives support from the NIHR Imperial Biomedical Research Centre and the NIHR HPRUs in Chemical and Radiation Threats and Hazards and Environmental Exposures and Health, the British Heart Foundation Centre for Research Excellence at Imperial College London (RE/18/4/34215), the UK Dementia Research Institute at Imperial (MC_PC_17114), and Health Data Research UK (HDR UK). We thank the Huo Family Foundation for their support of our work on COVID-19. **Ethics:** We obtained research ethics approval from the South Central–Berkshire B Research Ethics Committee (Integrated Research Application System ID 283787). **Author contributions:** S.R. and P.E. conceptualized and designed the study and drafted the manuscript. S.R., K.E.C.A., O.E., H.Wan., C.F., and C.E.W. undertook the data analysis. P.J.D., D.A., and C.A.D. provided statistical advice. G.C., W.B., H.War., C.A., and A.D. provided study oversight. A.D. and P.E. obtained funding. S.R., K.E.C.A., O.E., H.Wan., C.E.W., C.A., P.J.D., D.A., C.A.D., G.C., W.B., H.War., G.T., A.D., and P.E. critically reviewed the manuscript. All authors read and approved the final version of the manuscript. P.E. is the guarantor for this paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted, had full access to all the data in the study, and had final responsibility for the decision to submit for publication. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Code and additional data to support the figures are freely available in Zenodo (*43*).

SUPPLEMENTARY MATERIALS

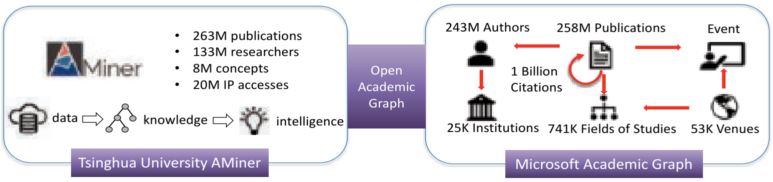
science.sciencemag.org/content/372/6545/990/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S7
References (*44*–*52*)
Data S1
MDAR Reproducibility Checklist

5 October 2020; accepted 20 April 2021
Published online 23 April 2021
10.1126/science.abf0874

SIGKDD TEST-OF-TIME AWARD

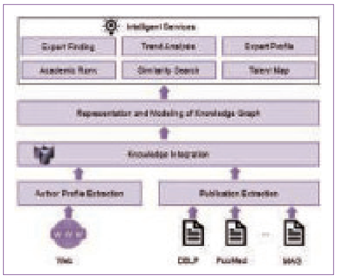
AMINER: AI-POWERED ACADEMIC NETWORK MINING

As one of China's most prestigious and influential universities, Tsinghua is committed to cultivating global citizens who will thrive in today's world and become tomorrow's leaders. Through the pursuit of education and research at the highest level of excellence, Tsinghua is developing innovative solutions that will help solve pressing problems in China and the world.



AMiner, an academic search and mining system which has been developed by professor Jie Tang and his team from Tsinghua fourteen years ago, has now changed the way of academic research.

Comparing with other popular academic systems, such as Google Scholar, Microsoft Academic Search, CiteSeerX, CiteULike, ResearchGate, Semantic Scholar, DBLP and so on, AMiner can quickly scan and understand scientific text such as news articles, research papers and researchers' profiles. By automatically highlighting the most important features of articles and profiles from different sources, it can generate semantic-based profiles of researchers, find the connections among a cluster of papers, explore emerging technologies, and even predict the future directions of science and technology.



Last year, at the early stage of the COVID-19 outbreak, AMiner timely launched the COVID-19 knowledge graph, which promoted the progress of COVID-19 related research. The knowledge graph included the comprehensive profile information of scholars, such as their published papers, filed patents, research projects, academic services, and related news articles.

Technique-wise, however, the above semantic information is often scattered across isolated and heterogeneous data sources. Tang and his team devoted themselves to developing novel algorithms to collect, extract, and infer different categories of information for scholars, based on which they construct a dynamically increasing scholar-centered academic knowledge graph. To build the knowledge graph, explicit information across isolated data sources, such as papers, patents and projects, need to be integrated, where normalization and disambiguation are key challenges to be addressed. They organized isolated data sources into graphs, and then leveraged the graph structures to reduce the impact of the unnormalized and ambiguous text information. Moreover, implicit information such as the advisor-advisee relationships, scholar's research interests and the semantic relationships between research concepts are inferred using deep learning-based algorithms. Without the need of annotating massive implicit relationships, the embeddings of entities in knowledge graphs are learned by self-supervised models to facilitate the inference of the implicit relationships.

With the underlying knowledge graphs, many useful tools and applications have been developed by the team. For instance, Topic Trend module can demonstrate the macro evolution of general research topics such as "deep learning" or more specific techniques such as "auto encoder". On the other hand, Master Reading Tree module is a tool to help scholars learn the micro evolution of a popular paper. Another example is Must Reading Papers module, which targets at sorting out the state-of-the-art papers and the influential scholars about a topic. These tools are helpful for scholars to trace the up-to-date research progresses. Star Talents module is an application to predict the academic potential of a scholar in the future, which can shed light on choosing supervisors for students or seeking collaborator for companies.

Every year, AMiner attracts more than 20 million independent IP accesses from over 220 countries/regions (per Google Analytics). In addition, together with Microsoft, AMiner has released the most massive open academic graph (OAG) with over 300 million publication papers and 100 million authors. It is definitely the largest released academic data. Because of these achievements, Tang and his team were awarded the Microsoft Research Asia Collaborative Research Award. The original paper about the AMiner system published on ACM SIGKDD 2008 received the inaugural Test of Time Award for Applied Science in recognition of the team's study of mining academic social networks.

Currently, the technology is still facing bottlenecks, despite tremendous developments in the past years. According to Tang and his team, the current status is still far away from real AI. Intelligent functionalities are required to improve the efficiency of academic research. For example, there are still mistakes in the built knowledge graph, such as wrong authors matched to some papers,or advisor-advisee relationship errors. Deploying a monitor robot is a way to dynamically update the knowledge graph, and to correct potential mistakes using reasoning. In addition, an intelligent service is needed to deeply analyze the content and summarize the complex knowledge including research problems, challenges and approaches presented in papers, and even invent new ideas based on existing knowledge. Finally, a chatbot is also necessary to answer any natural language questions from end users. Tang and his team are making their efforts on the direction. In 2019, Zhipu AI was founded by part of the research team. The Tsinghua's spinoff also used large pre-trained models and other technologies to understand and analyze information of cutting-edge sciences and technologies.

We seek top-level applicants with an outstanding research record in the broad area of data mining for intelligent systems (theory, computation, and practice). Research areas of interest include, but are not limited to: (i) data mining, (ii) distributed computing, (iii) theory and algorithm, (iv) artificial intelligence, (v) machine learning. Please send the application package to the following address by email to: **jietang@tsinghua.edu.cn**.

ADVERTISEMENT

清華大学
电机工程与应用电子技术系
Department of Electrical Engineering
Tsinghua University

Department of Electrical Engineering (DEE) of Tsinghua University, founded in 1932, is one of the three earliest engineering departments since the foundation of the university. Bearing the historical mission of cultivating top talents for the nation and advancing the electric power industry in China, DEE has continuously made breakthroughs and developments in a broad range of areas for nearly 90 years. DEE not only was ranked the first or A+ class in the past national evaluations of disciplines, but also established an outstanding global reputation. In recent years, the QS ranking of Tsinghua University in Electrical Engineering has been around top 10 in the world.

In the 21st century, the world is facing a major historical opportunity for energy evolution. China has announced the goal of "peak carbon dioxide emissions before 2030 and achieve carbon neutrality before 2060". To achieve this goal, DEE commits to lead the research and education in power engineering in China, aiming at cutting edge contributions in discipline development and scientific research, serving major national needs and promoting the in-depth development of basic disciplines including power systems, power electronics and electric machine systems, high-voltage technology, flexible transmission and distribution, and new electrical technologies. At the same time, DEE continues to expand the extension of disciplines, vigorously promote the interdisciplinary integration, and foster the technology-to-market progress and industrial applications, for which, DEE has gradually formed a strategic discipline structure of "one-axis-two-wings" (Fig.1). Energy Internet Research Institute of Tsinghua University and Sichuan Energy Internet Research Institute of Tsinghua University have been established to promote the concept and technology development of energy internet and therefore support "one-axis-two-wings".

DEE has been focusing on "organized scientific research" and has formed 11 research teams across multidisciplinary academic fields. A series of achievements have been made in the field of stability control and protection of ac/dc complex power grid, including autonomous-synergic voltage control, traveling waves-based protective relaying of power lines, on-line analysis of damping characteristics and its adaptive control, active defense again cascading failures, IGCT device and high voltage DC breaker, smart and advanced dielectric materials, UHVAC and UHVDC technology, power electronics transient analysis methodology. These technologies have been successfully applied in major power engineering infrastructures operated by State Grid Corporation of China and China Southern Power Grid.

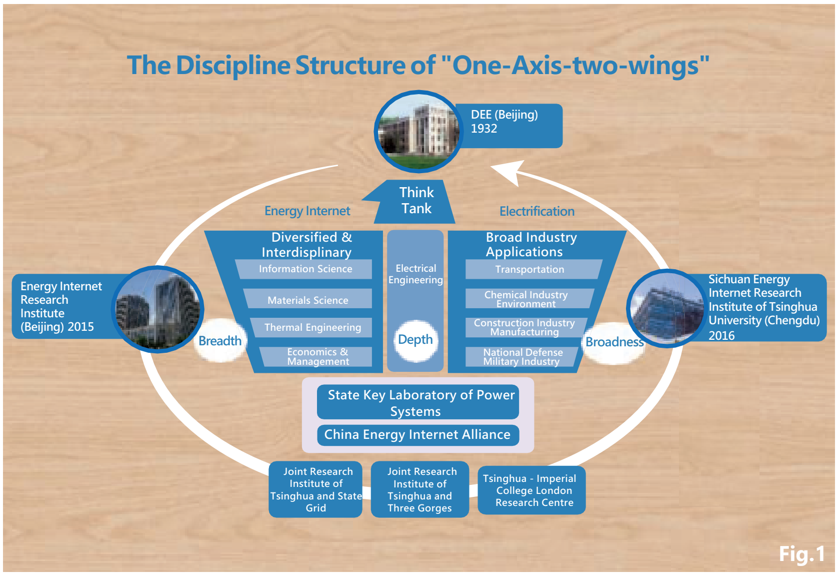


Fig.1

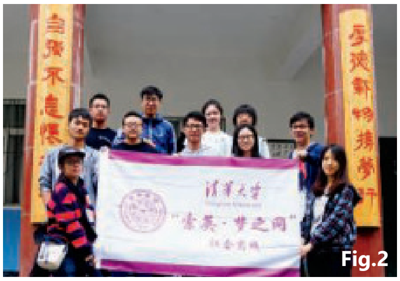


Fig.2

In terms of teaching and education, DEE has developed a new International Energy Internet Program, dedicated to cultivating future top-profile innovation leaders with international vision in the energy field. Besides offering the first MOOC (massive open online course) for the course of Principles of Electric Circuits in China, DEE actively explores "student-centered" teaching mode, using modern teaching tools such as Rain Classroom to carry out flipped classroom and blended learning reforms, and improves students' learning qualities. Furthermore, DEE encourages students to undertake social services and launches the Dream Grid Program to build photovoltaic microgrid for primary schools in remote areas in western China (Fig.2).

Department of Electrical Engineering has been actively bridging China and the world. During the past 5 years, about 30 students won best paper awards in international conferences. DEE founded or co-founded influential internal conferences such as IEEE EI^2 and iSpec, and established the Energy Internet Coordination Committee at the IEEE Power and Energy Society. Hitherto, Department of Electrical Engineering has 9 IEEE Fellows and 26 IET Fellows, and 85 faculty members are editor-in-chief or associate-editor in internationally renowned journals.

Facing future challenges, Department of Electrical Engineering of Tsinghua University will continue to carry forward the tradition of "Dedication to the Country and Pursuit of Excellence", uphold the department motto of "Rigorous in Learning and Righteous in Conduct", actively promote the development of disciplines, improve the quality of talent training, and make significant contributions in serving the national energy revolution strategy. In April 2022, DEE will celebrate its 90th anniversary and would like to welcome global alumni, collaboration partners and friends to join the celebration events.

Tel: +86-10-62782014 **Email:** ee@tsinghua.edu.cn

浙江大學
ZHEJIANG UNIVERSITY
1897
ZJU
求是大讲堂

Zhejiang University (ZJU) is one of China's top higher education institutions, as well as one of its oldest; its roots can be traced back to 1897 and the founding of the Qiushi Academy.

Located in Hangzhou – one of China's most picturesque cities – the University is organized across 7 faculties and 37 schools. It is home to 3,741 full-time faculty members, including 46 members of the Chinese Academy of Sciences and the Chinese Academy of Engineering. ZJU has 54,641 students, over 53% of whom are postgraduate students.

Laying claim to several areas of research strength, ZJU currently ranks among the top three on Chinese mainland and within the top 100 in the Times Higher Education World Reputation Rankings and QS World University Rankings. Eighteen disciplines of ZJU have been selected for China's "Double First-class" Initiative (3rd in China) and 39 disciplines graded A in the recent national assessment (1st in China).

The University prides itself on a culture of innovation and entrepreneurship. ZJU researchers are making an impact across many priority areas that address global challenges, including artificial intelligence, assembly technology for large aircraft, clean energy, ocean technology, industrial control technology, and global public health initiatives related to the prevention and treatment of infectious diseases. ZJU is also renowned for the number of business start-ups it spins off. Over 100 of its alumni sit at the helm of domestic or overseas listed companies, making the University synonymous with excellence and leadership.

ZJU is committed to transforming China and the world through active engagement. Major innovative developments include the creation of a high-level platform for intellectual property exchange, as well as the formation of a number of think-tanks, including the China Academy of Western Region Development, the National Research Center for Agricultural and Rural Development, and the Institute for Public Policy, which exist to extend the scope of ZJU's research in social sciences.

ZJU has partnerships in place with more than 190 institutions from more than 30 countries worldwide. Included among them are such leading institutions as Imperial College London, the University of Sydney and the University of Illinois at Urbana-Champaign.

With a cohort of 7,074 international students, and around 8,000 faculty and students who participate annually in various overseas mobility programs, ZJU fully harnesses its extensive network to nurture talent with a global outlook. In collaboration with the Universities of Edinburgh and Illinois it has also established the ZJU-UoE and ZJU-UIUC Institutes on Haining International Campus.

Chief among ZJU's aims is the aspiration to become a world-class university with a distinctively Chinese character, where tradition and modernity are successfully combined.

Welcome to join us in ZJU.

Tel: 0571-88981390 88981345　　Fax: 0571-88982640　　Email : tr@zju.edu.cn

Join Hands with Health to Compose a Glorious Chapter!

Shanghai University of Medicine & Health Sciences sincerely welcomes outstanding talents at home and abroad

Shanghai University of Medicine & Health Sciences is an applied technique-oriented undergraduate medical university affiliated to Shanghai, the university is oriented towards application, characteristics and internationalism, adheres to the development direction of "medical and industrial integration, medical and nursing integration, and medical insurance integration", and aims to cultivate human health promoters with sound personality and psychology, who can solve practical problems and have the potential to lead the industry. There are more than 11,000 students in 17 majors such as Clinical Medicine, Nursing, Rehabilitation, Medical Devices, Medical Examination, Medical Imaging, Health Management and Service, Pharmacy, Health Inspection and Quarantine, Clinical Engineering, Medical Product Supervision, Big Data Technology, Emergency Management, Dental Technology, and Intelligent Imaging Engineering, and more than 800 colleagues here are welcome you. The university has built affiliated hospitals such as Jiading District Central Hospital, Zhoupu Hospital, Pudong New Area People's Hospital, Jinshan District Central Hospital, Chongming District Central Hospital, Landseed Hospital, Shanghai, affiliated teaching hospital system composed of tertiary, secondary and first-class hospitals and medical institutions, and the teaching base composed of hundreds of high-level medical and health service institutions, medical device enterprises and testing units. The university has Shanghai molecular imaging key laboratory, Shanghai universities key laboratory and engineering center. The university has carried out extensive and in-depth international cooperation and exchanges with dozens of well-known universities in the United States, Japan, France, Britain, Australia, Germany and other countries and regions.

II. Job Recruitment

[Discipline Leading Talents] The candidates should be professionals and scholars with noble ethics and rigorous academic style, who are active in the frontline of academic research, well known at home and abroad, able to quickly raise their discipline to the advanced level at home and abroad, and have achieved recognized academic achievements in their field of study.

[Excellent Talents] In general, excellent talents with a high-level doctorate degree and more than 2 years of frontline teaching and research practice experience. Qualifications for cultural, artistic and physical education talents may be appropriately relaxed. Their academic level should generally be above average in their own filed.

III. Discipline Direction

Outstanding talents in Clinical Medicine, Nursing, Medical Technology, Biomedical Engineering, Medical Imaging Technolgy, Basic Medicine, Stomatology, Public Health and Preventive Medicine, Artificial Intelligence, Biomedicine and other disciplines.

IV. Relevant Benefits

We will provide a series of policy support based on school talent introduction package and other related measures, including "One person, One policy" consultation in certain circumstances.

Contact person: Mr. / Ms. Liu **Tel:** 86-21-65882192 **E-mail:** jkyxyhr@163.com **Web:** www.sumhs.edu.cn

Address: Human Resources Office, Shanghai University of Medicine & Health Sciences, 279 Zhouzhu Road, Pudong New Area, Shanghai (Postal code: 201318)

哈爾濱工程大學

Rise in the Marine Field with the Mind of a Big Power, and Build a Maritime Power through Talents First

This is Harbin Engineering University, originated from Harbin Institute of Military Engineering. Subordinated to the Ministry of Industry and Information Technology as one of the "seven major universities of national defense", I am a national key university covering multiple fields of disciplines including science, engineering, liberal arts, law, economics, management, etc. As one of the first batch of key construction universities of the Project 211 and a university for the construction of innovation platforms of superior disciplines of the Project 985, I have now been selected for the national "Double First-class" construction.

"Three-marine and one nuclear field" are my distinctive school-running features. The ocean and the shipbuilding industry are where my voyage leads to, and the nuclear industry is my arena.

I am the largest talent training base for shipbuilding and nuclear industries in China, as well as one of the most important scientific research bases for shipbuilding and nuclear industries in China.

In addition to the main campus, I am focusing on building Qingdao Innovation Development Base/Qingdao Research (Postgraduate) Institute, Yantai Research (Postgraduate) Institute, and Nanhai Innovation and Development Base, based on China's major development strategies and taking into account my school-running features of "three-marine and one nuclear field".

Leading talents

A leading talent should have strong academic influence and have made domestically leading or globally advanced academic achievements, with academic influence ranking forefront among that of domestic and foreign young and middle-aged scholars in the filed. A young leading talent should aim to be selected as a national-level talent in the future.

Excellent teachers

An excellent teacher should have strong academic potential, and hold a doctorate degree from a high-level domestic or foreign university. He/she should own good academic foundation and present the potential to grow into a young leading talent.

Postdoctoral innovative talents

A postdoctoral innovative talent should have strong academic innovation abilities, with a doctorate degree which was obtained less than 3 years ago. The talent should be able to meet the requirements of a full-time scientific research position, and present the potential to quickly grow into an excellent teacher.

Contact person: Li Guangdong, Sun Mengnan (For postdoctoral fellows)
E-mail: 66710432@qq.com, sunmengnan@hrbeu.edu.cn (For postdoctoral fellows)
Correspondence address: Human Resources Office, Room 809, Main Building No.1, Harbin Engineering University, 145 Nantong Street, Nangang District, Harbin

---

CHINESE UNIVERSITIES NOW RECRUITING THE VERY BEST SCIENTISTS COME AND GET REGISTERED ONLINE!

Welcome back to hometown.
Thousands of academic job vacancies are in fast-developing China.

Participation in our
Online Job Fair

800 Chinese Universities
Complete the information to apply | Make an appointment in advance

We also provide
Authentic job opportunities and professional consultation

2021

From May to December, this is full schedule,
looking forward to your join in us

May 28
Jun 18
Sep 24
Oct 18
Oct 29
Nov 19
Dec 03

Please visit **www.acabridge.edu.cn**

Scan to apply

Welcome to our
Young Scholars' Forum

Meet the HR and build your career from over 29 cities

Shanghai
Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Translational Medicine Research Institute, Shanghai University

Ningbo / Wenzhou, Zhejiang
Ningbo University, Wenzhou University of Technology

Qingdao / Weifang, Shandong
Shandong University of Science and Technology, Rehabilitation University (in preparation), Weifang College

Hefei / Wuhu, Anhui
Anhui Medical University, Anhui University of Technology, Wannan Medical College

Changsha, Hunan
Hunan University

Shijiazhuang / Baoding, Hebei
Shijiazhuang Tiedao University, Hebei Medical University, The National Police University for Criminal Justice

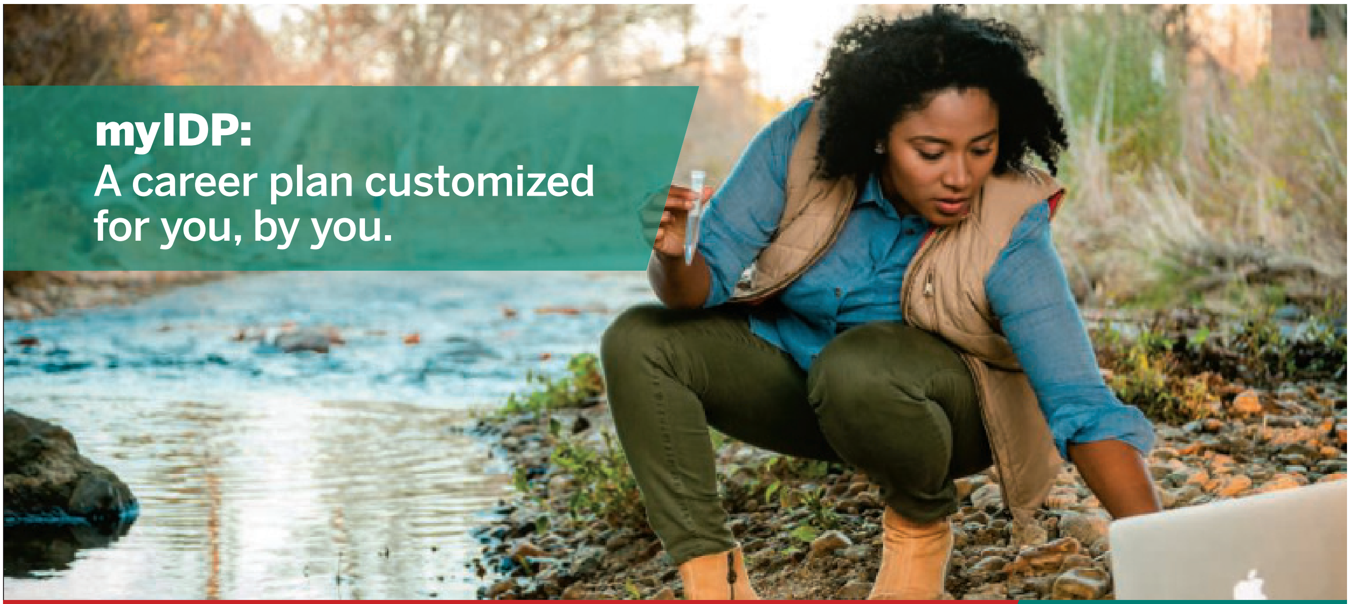
Chengdu, Sichuang
University of Electronic Science and Technology of China

Harbin, Heilongjiang
Harbin Engineering University

黑龙江
河北
山东
安徽
上海
浙江
四川
湖南
南海诸岛

Scan to apply

There are still more …


myIDP:
A career plan customized for you, by you.

online @sciencecareers.org

ScienceCareers

OPPORTUNITIES IN CHINA

National University of Defense Technology (NUDT) is a comprehensive research university and a key university supported by the national “985 Project”, “211 Project” and “Double First Class” project. The University consists of 13 colleges and research institutes, including College of Liberal Arts and Sciences, College of Computer Science and Technology, College of Electronic Science and Technology, and College of Advanced Interdisciplinary Studies, etc., which are located in Changsha (Hunan Province), Nanjing (Jiangsu Province), Hefei (Anhui Province), Wuhan (Hubei Province), Xi’an (Shaanxi Province).

Among its 6 discipline categories ranging from science and engineering to management and law, etc., NUDT is authorized to confer doctoral degrees in 23 first-level disciplines and master’s degrees in 26 first-level disciplines.

In the latest national assessment of first-class disciplines, eight disciplines of the University were rated ‘A level’, including 4 A+ disciplines, 3 A disciplines, and 1 A- discipline. NUDT ranks 11th nationally in terms of the number of A+ discipline. Five disciplines of the University, namely information and communication engineering, computer science and technology, aerospace science and technology, software engineering, management science and engineering, have been selected into the list for the national initiative to construct first-class disciplines.

NUDT boasts a high-level faculty composed of “leading talents and innovative teams”. Among its faulty, there are 16 academicians of the Chinese Academy of Sciences and the Chinese Academy of Engineering, and over 500 recipients of major national talent programs support or awards. There are also 10 national-level innovation teams, 8 national-level teaching teams, 2 NSFC Innovation Research Team, and 1 team winning the National Award for Excellence in Innovation.

NUDT is home to an array of advanced teaching and research facilities, including 10 national key laboratories, 6 national-level experimental teaching demonstration centers, 3 national and local joint engineering research centers.

Learn more and join us >>>
Email: rcfw@nudt.edu.cn

学术桥
AcaBridge.edu.cn

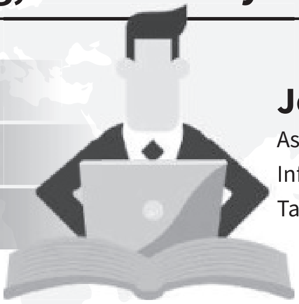
One-on-one Career Consultation

Professional in Job seeking, Customize your career plan

RESUME EVALUATION

CAREER CONSULTATION

JOB RECOMMENDATION



Join us and get
Assistance in your interview
Informed of latest activities
Talent policy interpretation

OUR FEATURE »

Information Management
Search for thousands of job opportunities
Direct delivery of applications

Workstation in 28 Cities
Close cooperation with School HR Dept., Colleges and Research Institutes

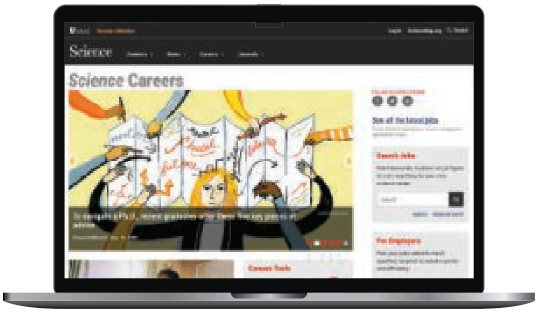
Thousand Jobs in
All disciplines
Faculty & postdoc

Recruitment Activities
Online job fair
“Overseas scholars visit to China“ forums

If you are seeking job opportunities in Chinese universities or research institutes, please contact the talent service agency of AcaBridge (**consultant@acabridge.edu.cn**), which provides one-on-one consultations. For more information, visit us at **www.acabridge.edu.cn**.

Scan to apply

CAREER
PLANNING


Science
Science Careers

WORKING LIFE

By **Robert Blasiak**

Family first

"Don't worry," my boss said. "No one will even notice you were gone." I knew he was trying to calm my worries that I would inconvenience others by taking parental leave after the birth of my second child. But in some ways, deciding to take time off turned out to be more complicated the second time around. I had just started to take on leadership roles at work, and I did not want to let down my colleagues and collaborators. I feared that if I took a leave, everything would fall apart and they'd realize I never should have been trusted in the driver's seat in the first place. Now, I had the opposite worry: "If they don't even notice I'm gone, will they realize everything works fine without me?"

When our first child was born 4 years earlier, my wife and I lived in Japan, where paternity leave is a somewhat unusual ask. We had a long conversation about what would work best for us as professionals and as a family. My wife's contract as a university program manager allowed 2 months of maternity leave; my contract didn't offer any parental leave. But I knew I would regret not taking time off to be with our son. Plus, none of my work colleagues would be affected by any delays to finishing my thesis, preparing for my defense, or submitting my grant application. The decision was clear: My wife and I would both take leave before returning to work and starting our son in child care.

"My contract didn't offer any parental leave. But I knew I would regret not taking time off."

I was lucky to have a supportive boss who helped me adjust my contract; for 9 months, I worked 2 days a week before returning to full time. At home, life revolved around the baby, full of joy and exhaustion. Work provided a space for adult interactions and surely helped my mental health. I also quickly become more realistic and pragmatic with my time management at work—a skill that seems more important with every passing year.

Over the next few years, I grew as a parent and an academic, navigating diapers and day care and moving to a postdoc position in Sweden. My new supervisors nudged me out of my comfort zone and encouraged me to step up into leadership roles, including co-leading a project with 11 colleagues scattered across six time zones and another with eight colleagues in eight different countries.

So, when my wife and I were expecting our second child in 2019, our considerations about taking leave were quite different. Sweden offers generous support for new parents, which in some ways made things easier—but at the same time, I was now responsible for collaborations that involved people I didn't want to let down. Somewhere not-so-deep inside, I had a feeling I was about to blow my big chance. I had entered academia in my 30s and was a decade behind my peers in terms of experience. Could I afford to take more time away from work?

After much discussion with my wife and colleagues, I realized that—as much as I love science—my family comes first. My wife decided she was comfortable taking a year off from her job, and I ended up taking 8 months off over the course of a year and a half—the first 3 months after our son was born, some time over the summer when day care was closed and both kids were at home, and then 4 months after my wife went back to work.

I found that my fears about taking leave were mostly unfounded. Collaborators immediately stepped up to keep projects on track or bring them over the finish line. No one ever blamed me for taking time off. I missed a couple of funding calls and was left off a few papers I think maybe I should have been on, but I'm already forgetting the specifics. I don't have any regrets—and if I'm another step behind my peers as a result, so be it.

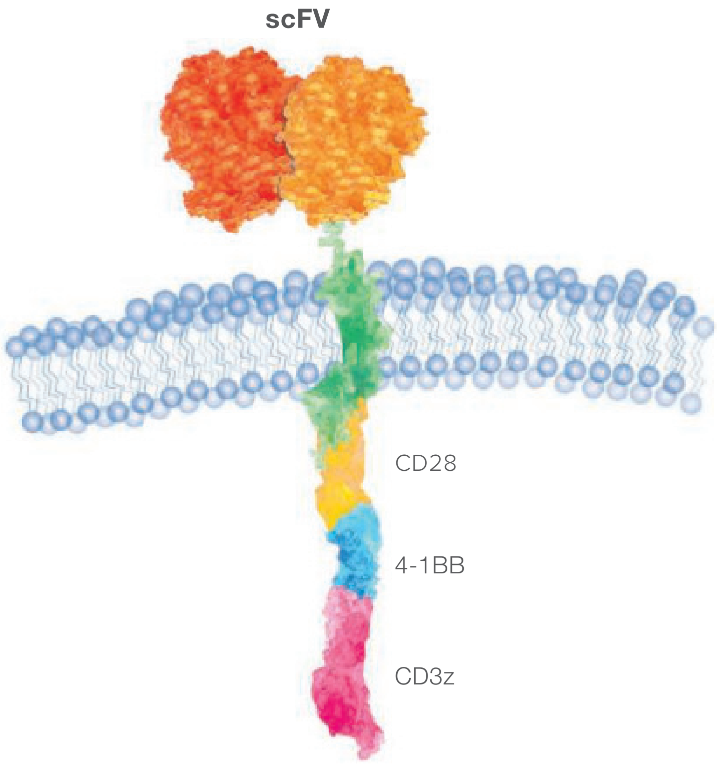
For me, accepting that my career is secondary to my family made every other decision easier. And when I articulated that to my colleagues, they all understood and accepted my priorities. Sometimes academia is described as a hypercompetitive space in which people are pushed to the extremes to succeed. But the support and generosity of my colleagues and collaborators has taught me it can also be a kind space, and this is something to foster and treasure. ■

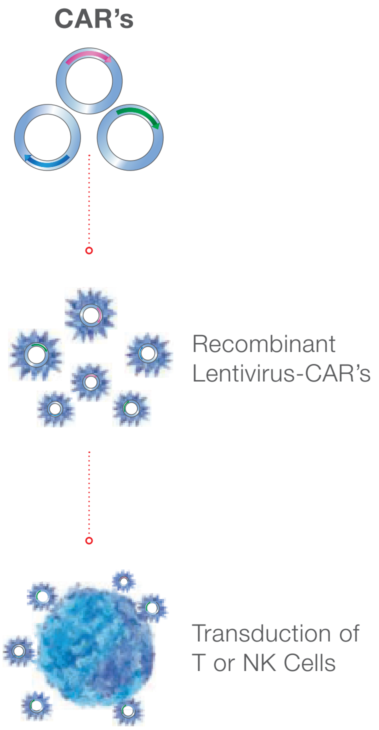
Robert Blasiak is a researcher at Stockholm University. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Custom Lentivirus Generation

High Quality Lentiviral Vectors





Our standardized and highly optimized platform enables ProMab to produce recombinant lentivirus with many benefits:

- Custom vector engineering from gene synthesis to plasmid production
- High titer: 10^9 to 10^11 particles/ml
- Fast turnaround time: 2 to 3 weeks
- Variety of analytical options
- Ready to use CAR-T lentivirus
- Downstream to CAR-T or stable cell line development services

ProMab Biotechnologies offers custom lentivirus generation services to deliver specific genes of interest to advance cell and gene therapy applications. Lentivirus offers high efficiency transduction and stably integrated expression to create innovative cell lines and engineered primary cells. From small-scale proof-of-concept to large scale production, utilize ProMab's platform to accelerate your research.

All products are for research only

Discover more | www.promab.com

ProMab
Biotechnologies, Inc.
2600 Hilltop Dr, Building B, Suite C320, Richmond, CA 94806
1.866.339.0871 | info@promab.com
SCAN ME