Harnessing the Mississippi to rebuild vanishing land p. 334

Scarless wound healing in mouse skin pp. 346 & 362

Tackling racial inequities in medicine p. 348

Science

$15
23 APRIL 2021
sciencemag.org

AAAS

DRY WELLS

Falling groundwater levels threaten wells globally pp. 344 & 418

2022 AWARDS & PRIZES

AAAS DAVID AND BETTY HAMBURG AWARD FOR SCIENCE DIPLOMACY

SUBMIT A NOMINATION TODAY
Deadline June 30, 2021

The **AAAS David and Betty Hamburg Award for Science Diplomacy** recognizes an individual or a limited number of individuals working together in the scientific and engineering or foreign affairs communities who are making an outstanding contribution to furthering science diplomacy.

Over the past 28 years, AAAS has honored an international cadre of science luminaries for their contributions to international scientific cooperation and science diplomacy. **This year, the award has been renamed for David and Betty Hamburg** to recognize their unparalleled commitment to the significant role of science diplomacy to advance science, human rights, peace, and cooperation.

- The award is open to all regardless of nationality or citizenship.
- We accept self-nominations.
- Nominees must be living at the time of their nomination.

To learn more, visit

aaas.org/awards/science-diplomacy/about

AAAS gratefully acknowledges the Carnegie Corporation of New York for their generous support to launch the AAAS David and Betty Hamburg Award for Science Diplomacy and the individuals and foundations whose contributions have begun an endowment that will allow us to sustain it in perpetuity.

PHOTO CREDIT: LETICIA VALVERDES/SILVERBACK

2021 WINNER
Carlos Nobre, Ph.D.

Honored for his career-spanning work to understand and protect the biodiversity and Indigenous Peoples of the Amazon.

Winners will receive:

- Monetary prize of $10,000
- Commemorative plaque
- Worldwide promotion of their achievements through AAAS communication channels, including AAAS publications, member news, website, and social media
- Complimentary registration to the 2022 AAAS Annual Meeting
- Reimbursement for travel and hotel expenses to attend the AAAS Annual Meeting
- The opportunity to publish in *Science & Diplomacy*

AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

CONFERENCES

Alcohol-Induced End Organ Diseases

OCTOBER 24 – 29, 2021 • FOUR POINTS SHERATON / HOLIDAY INN EXPRESS • VENTURA, CA, US

CHAIRS: Gyongyi Szabo and Bernd Schnabl

VICE CHAIRS: Elizabeth J. Kovacs and Gavin E. Arteel

Alcohol-Induced End Organ Diseases GRS

OCTOBER 23 – 24, 2021

CHAIRS: Kyle L. Poulsen and Holly Hulsebus

Biotherapeutics and Vaccines Development

NOVEMBER 14 – 19, 2021 • VENTURA BEACH MARRIOTT • VENTURA, CA, US

CHAIRS: Michael J. Tarlov and Linda O. Narhi

VICE CHAIRS: Peter Tessier and Jessica R. Molek

Biotherapeutics and Vaccines Development GRS

NOVEMBER 13 – 14, 2021

CHAIRS: John S. Schardt and Maria-Monica Castellanos

Cannabinoid Function in the CNS

NOVEMBER 7 – 12, 2021 • FOUR POINTS SHERATON / HOLIDAY INN EXPRESS • VENTURA, CA, US

CHAIRS: Matthew N. Hill and Sachin Patel

VICE CHAIRS: Miriam Melis and Mario van der. Stelt

Cannabinoid Function in the CNS GRS

NOVEMBER 6 – 7, 2021

CHAIRS: Anissa Bara and Simon Chamberland

Carbon Capture, Utilization and Storage

OCTOBER 24 – 29, 2021 • JORDAN HOTEL AT SUNDAY RIVER • NEWRY, ME, US

CHAIR: Roger D. Aines

VICE CHAIR: Andre Bardow

Carbon Capture, Utilization and Storage GRS

OCTOBER 23 – 24, 2021

CHAIRS: Humera Ansari and Tony Feric

Cardiac Arrhythmia Mechanisms

OCTOBER 17 – 22, 2021 • FOUR POINTS SHERATON / HOLIDAY INN EXPRESS • VENTURA, CA, US

CHAIRS: David J. Christini and Ursula Ravens

VICE CHAIRS: Crystal M. Ripplinger and Edward J. Vigmond

Cardiac Arrhythmia Mechanisms GRS

October 16 – 17, 2021

CHAIRS: Enaam Chleilat and Jaclyn A. Brennan

Cell Biology of Metals

OCTOBER 17 – 22, 2021 • MOUNT SNOW • WEST DOVER, VT, US

CHAIRS: Caryn E. Outten and Mitchell Knutson

VICE CHAIRS: Angela Wilks and Kerry Kornfeld

Cell Biology of Metals GRS

OCTOBER 16 – 17, 2021

CHAIRS: Cassandra E. Nelson and Liu Liu

Central Nervous System Injury and Repair

OCTOBER 17 – 22, 2021 • JORDAN HOTEL AT SUNDAY RIVER • NEWRY, ME, US

CHAIRS: Yimin Zou and Simone Di Giovanni

VICE CHAIRS: Elizabeth J. Bradbury and Yishi Jin

Central Nervous System Injury and Repair GRS

OCTOBER 16 – 17, 2021

CHAIRS: Jaclyn T. Eisdorfer and Angela R. Filous

Chronobiology

OCTOBER 10 – 15, 2021 • MOUNT SNOW • WEST DOVER, VT, US

CHAIR: Samer Hattar

VICE CHAIR: Charlotte Helfrich-Forster

Chronobiology GRS

OCTOBER 9 – 10, 2021

CHAIRS: Anna-Marie Finger and Josh Mouland

Computer Aided Drug Design

NOVEMBER 7 – 12, 2021 • MOUNT SNOW • WEST DOVER, VT, US

CHAIR: Georgia B. McGaughey

VICE CHAIR: Matthias Rarey

Computer Aided Drug Design GRS

NOVEMBER 6 – 7, 2021

CHAIRS: Eva Nittinger and Sarah Kochanek

Lung Development, Injury and Repair

NOVEMBER 7– 12, 2021 • WATERVILLE VALLEY • WATERVILLE VALLEY, NH, US

CHAIRS: Anne-Karina Perl and Daniel Tschumperlin

VICE CHAIRS: Rory E. Morty and Xin Sun

Lung Development, Injury and Repair GRS

NOVEMBER 6 – 7, 2021

CHAIRS: Semil P. Choksi and Chimwemwe Mwase

Mammalian DNA Repair

OCTOBER 31 – NOVEMBER 5, 2021 • FOUR POINTS SHERATON / HOLIDAY INN EXPRESS • VENTURA, CA, US

CHAIR: Lee Zou

VICE CHAIR: Patricia L. Opresko

Mammalian DNA Repair GRS

OCTOBER 30 – 31, 2021

CHAIRS: Ragini Bhargava and Ahmed Diab

Medicinal Chemistry

OCTOBER 24 – 29, 2021 • MOUNT SNOW • WEST DOVER, VT, US

CHAIR: Janeta Popovici-Muller

VICE CHAIR: Cynthia A. Parrish

Medicinal Chemistry GRS

OCTOBER 23 – 24, 2021

CHAIR: Thomas R. Simpson

Phagocytes

OCTOBER 24 – 29, 2021 • WATERVILLE VALLEY • WATERVILLE VALLEY, NH, US

CHAIR: Michelle R. Lennartz

VICE CHAIR: Sergio D. Catz

Phagocytes GRS

OCTOBER 23 – 24, 2021

CHAIRS: Catherine M. Buckley and Fernando Montano Rendon

Physical Science of Cancer

NOVEMBER 14 – 19, 2021 • FOUR POINTS SHERATON / HOLIDAY INN EXPRESS • VENTURA, CA, US

CHAIRS: Josef Käs and Sharon Gerecht

VICE CHAIRS: David J. Odde and Jennifer M. Munson

Physical Science of Cancer GRS

NOVEMBER 13 – 14, 2021

CHAIRS: Sara Nizzero and Alexander J. McGhee

Stochastic Physics in Biology

OCTOBER 10 – 15, 2021 • FOUR POINTS SHERATON / HOLIDAY INN EXPRESS • VENTURA, CA, US

CHAIR: Yuhai Tu

VICE CHAIR: Jie Xiao

Stochastic Physics in Biology GRS

OCTOBER 9 – 10, 2021

CHAIRS: Corey Weistuch and Oleksandra Romanyshyn

Stress Proteins in Growth, Development and Disease

NOVEMBER 7– 12, 2021 • JORDAN HOTEL AT SUNDAY RIVER • NEWRY, ME, US

CHAIR: Brian C. Freeman

VICE CHAIR: Harm H. Kampinga

Stress Proteins in Growth, Development and Disease GRS

NOVEMBER 6 – 7, 2021

CHAIRS: Stephanie N. Gates and Jessica Kho

Three Dimensional Electron Microscopy

OCTOBER 31 – NOVEMBER 5, 2021 • JORDAN HOTEL AT SUNDAY RIVER • NEWRY, ME, US

CHAIR: Sharon G. Wolf

VICE CHAIR: Daniela Nicastro

Three Dimensional Electron Microscopy GRS

OCTOBER 30 – 31, 2021

CHAIRS: Cameron Varano and Colin M. Palmer

Join your GRC colleagues for safe, in person conferences featuring cutting edge presentations, stimulating discussion and informal networking opportunities at our remote locations in New England and California.

Advance the Frontiers of Science: October-November

Apply now at: www.grc.org

NOMIS & Science
YOUNG EXPLORER AWARD

Research at the intersection of the social and life sciences

Unconventional. Interdisciplinary. Bold.

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

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

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

Science
AAAS

NOMIS
FOUNDATION
Creating the Spark

CONTENTS

23 APRIL 2021 • VOLUME 372 • ISSUE 6540

NEWS

IN BRIEF

324 News at a glance

IN DEPTH

328 Researchers decry planned road near St. Helens
Project could help avert catastrophic flooding, but threatens research plots
By R. F. Service

329 NIH lifts restrictions on fetal tissue research
Scientists applaud end of intramural research ban and Trump-era review board
By K. Servick

330 Sale of building to Qatar deepens Carnegie rifts
Some scientists see deal and reorganization as "corporatization" of historic institution
By M. Wadman

331 After vaccine failures, France laments biomedical decline
Researchers blame a squeeze in basic research funding and scarce venture capital for biotech startups *By T. Rabesandratana*

332 Ethical controversy swirls around shark fossil from Mexico
Paleontologists raise questions of access to exquisite specimen *By R. Pérez Ortega*

FEATURES

324 Unleashing Big Muddy
By sending Mississippi waters on new course, engineers hope to build new land—and test ways to save a retreating coast *By W. Cornwall*
PODCAST

INSIGHTS

POLICY FORUM

338 Reframing rankings in educational assessments
Ability and test-taking behavior should be disentangled and jointly reported to improve interpretation and fairness
By S. Pohl et al.

PERSPECTIVES

341 Distributed manufacturing for and by the masses
Networked systems, artificial intelligence, and an engaged public can enable mass customization of local products
By C. E. Okwudire and H. V. Madhyastha

342 Expanding cell-to-cell interactions
A new single-cell sequencing approach reveals astrocyte and microglia cross-talk during inflammation
By A. Silvin and F. Ginhoux
RESEARCH ARTICLE p. 360

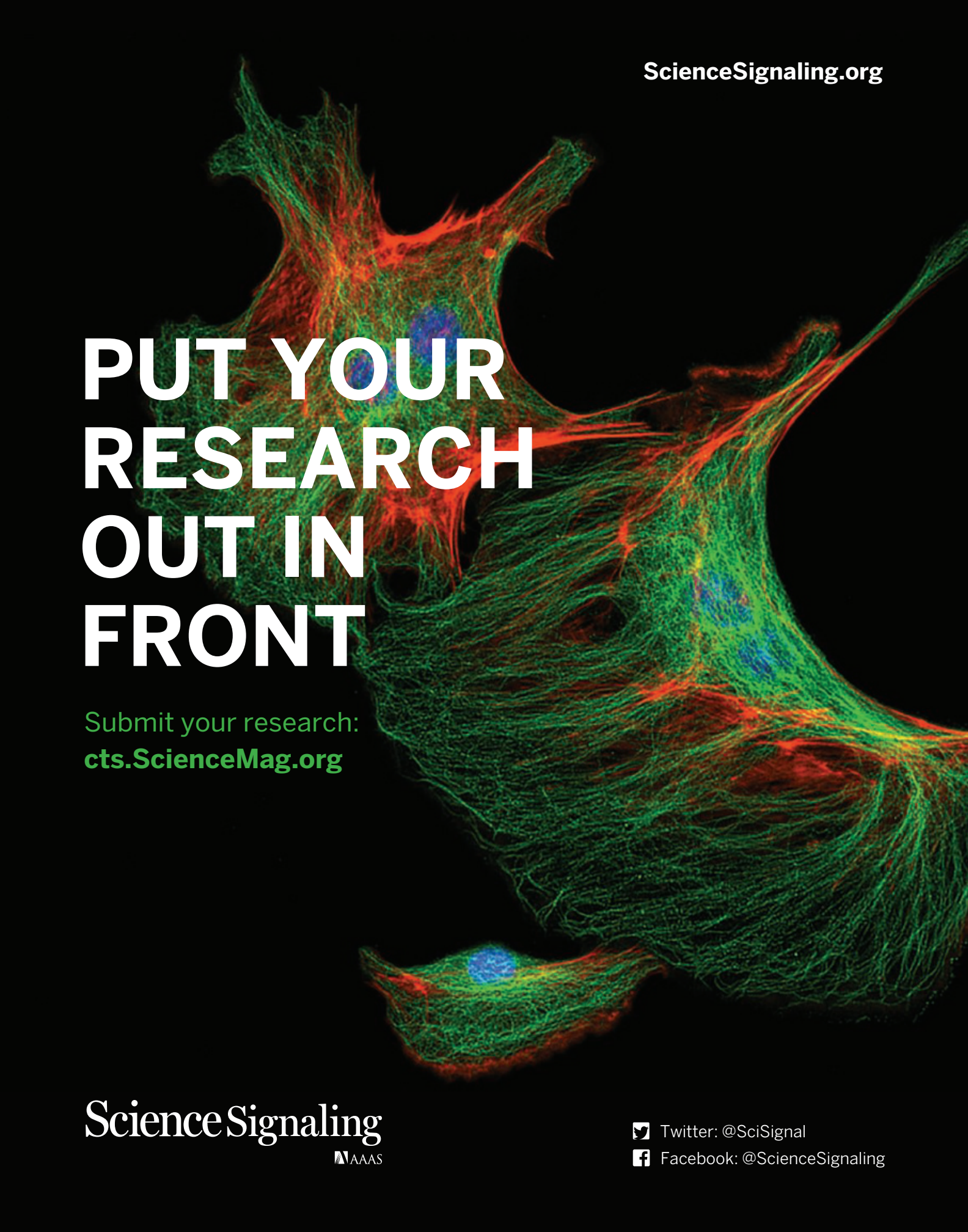
344 The hidden crisis beneath our feet
Disappearing groundwater requires action to prevent widespread water scarcity
By J. S. Famiglietti and G. Ferguson
REPORT p. 418

345 A spotlight on the elusive striatal cholinergic interneuron
A new role for the ubiquitous integrated stress response pathway in striatal-dependent learning and memory
By A. E. Ingebretson and J. C. Lemos
RESEARCH ARTICLE p. 361

346 Healing without scarring
A gene regulatory circuit directs tissue regeneration and limits fibrosis
By P. Konieczny and S. Naik
RESEARCH ARTICLE p. 362

PHOTOS (FROM TOP): GERRIT VYN/NPL/MINDEN PICTURES; GARETH FULLER/PA WIRE/AP IMAGES

ScienceSignaling.org



PUT YOUR RESEARCH OUT IN FRONT

Submit your research:
cts.ScienceMag.org

Science Signaling
AAAS

Twitter: @SciSignal
Facebook: @ScienceSignaling

348 Addressing racial inequities in medicine
The COVID-19 pandemic highlights the need to dismantle structural racism that perpetuates health inequities
By M. A. Fair and S. B. Johnson

350 José Manuel Baselga (1959–2021)
Transformative oncologist who laid the groundwork for precision medicine *By M. Scaltriti*

BOOKS *ET AL.*

351 Crafting a culture of secrecy
Clandestine research defined America's atomic age
By K. Brown

352 Man's (next) best friend
Our relationships with animals offer insight into the future of human–robot relations *By J. C. Gellers*

LETTERS

354 Partial immunity and SARS-CoV-2 mutations
By William P. Hanage and Colin A. Russell

354 Response
By C. M. Saad-Roy et al.
RESEARCH ARTICLE p. 363

355 A new frontier for mi familia
By R. D. Otero-Symphony

RESEARCH

IN BRIEF

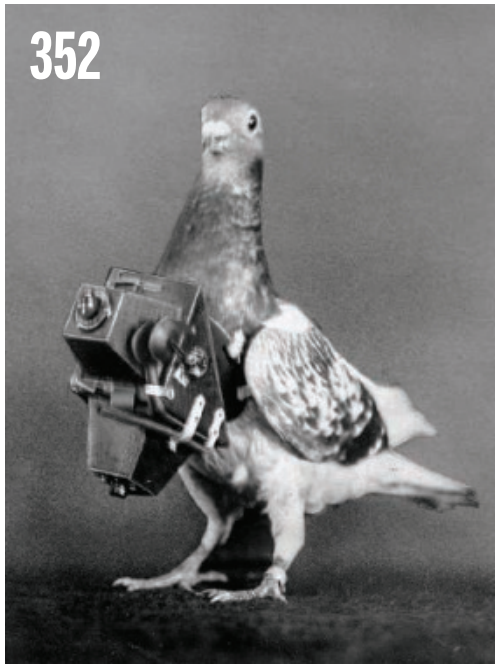
356 From *Science* and other journals

REVIEW

359 Plant science
Fluctuations shape plants through proprioception *B. Moulia* et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABC6868

RESEARCH ARTICLES

360 Neuroscience
Barcoded viral tracing of single-cell interactions in central nervous system inflammation
I. C. Clark et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF1230
PERSPECTIVE p. 342



361 Neuroscience
Cholinergic neurons constitutively engage the ISR for dopamine modulation and skill learning in mice *A. R. Helseth* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABE1931
PERSPECTIVE p. 345

362 Regeneration
Preventing *Engrailed-1* activation in fibroblasts yields wound regeneration without scarring *S. Mascharak* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABA2374
PERSPECTIVE p. 346

363 Coronavirus
Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes *C. M. Saad-Roy* et al.
LETTERS p. 354

371 Epigenome
Replication timing maintains the global epigenetic state in human cells
K. N. Klein et al.

378 Climate economics
Assessing China's efforts to pursue the 1.5°C warming limit *H. Duan* et al.

385 Neurodevelopment
Conserved genetic signatures parcellate cardinal spinal neuron classes into local and projection subsets *P. J. Osseward II* et al.

REPORTS

393 Heat transport
Broadband directional control of thermal emission *J. Xu* et al.

398 Organic chemistry
C(sp^3)–H methylation enabled by peroxide photosensitization and Ni-mediated radical coupling
A. Vasilopoulos et al.

403 Lasers
Higher-dimensional supersymmetric microlaser arrays *X. Qiao* et al.

409 Quantum devices
Josephson junction infrared single-photon detector *E. D. Walsh* et al.

412 Coronavirus
Timing the SARS-CoV-2 index case in Hubei province *J. Pekar* et al.

418 Groundwater
Global groundwater wells at risk of running dry *S. Jasechko and D. Perrone*
PERSPECTIVE p. 344

DEPARTMENTS

323 Editorial
Science journalism grows up
By Deborah Blum

430 Working Life
When do I tell my supervisor?
By Nancy Saana Banono

ON THE COVER

A farmer overlooks a dry well in Madhya Pradesh, India. Wells provide billions of people with drinking water and supply 40% of the water used for global irrigation. Scientists have shown that millions of wells may run dry if groundwater levels decline by only a few meters. This research can inform policies designed to support sustainable groundwater use and protect wells from running dry. See pages 344 and 418. *Photo: Prashanth Vishwanathan/ Bloomberg via Getty Images*

Science Staff 322
New Products 422
Science Careers 426

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.

PHOTO: BOYER/ROGER VIOLLET/GETTY IMAGES

Cottrell Scholars 2021: the Future *of* Science

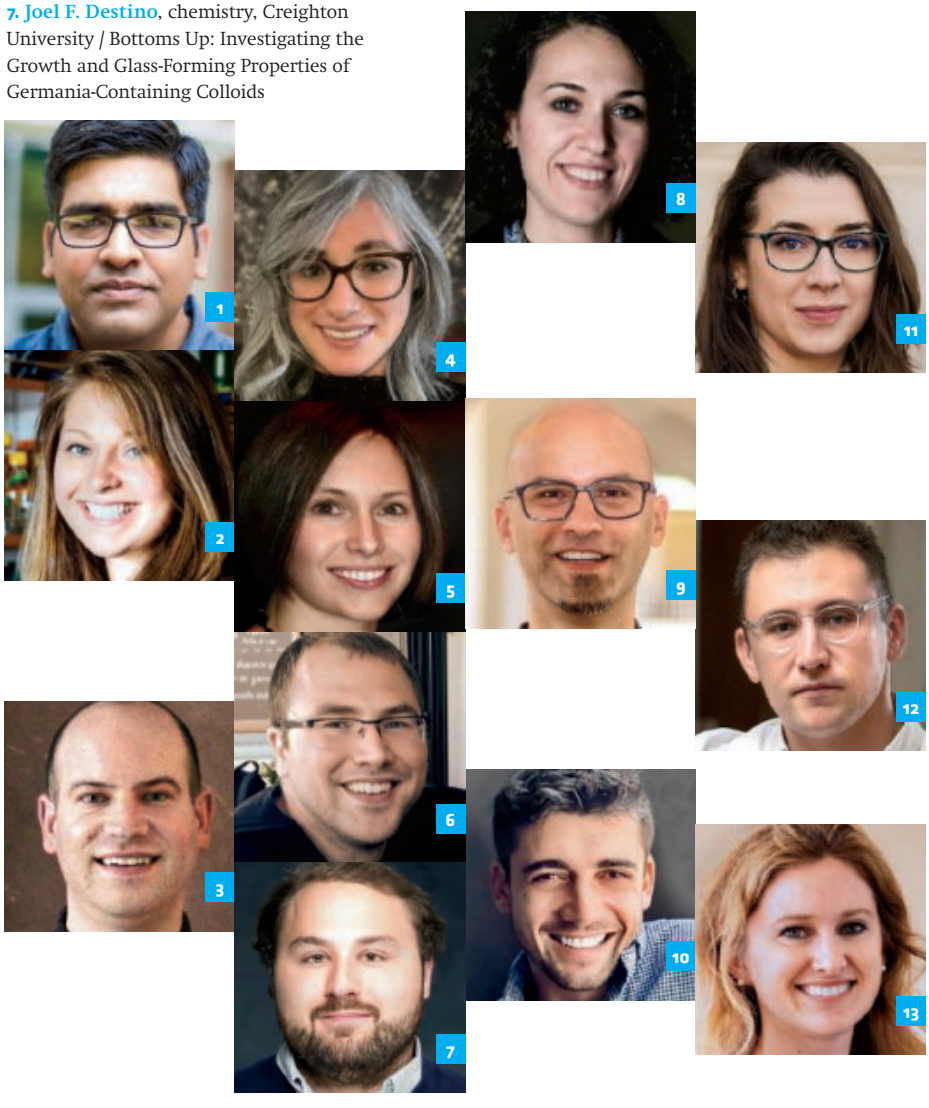
Research Corporation for Science Advancement, America's first foundation dedicated wholly to science, has named 25 early-career scholars in chemistry, physics, and astronomy as its 2021 Cottrell Scholars. Each awardee receives $100,000. Recipients are recognized for high-quality research, innovative educational initiatives, and academic leadership.

Three accomplished Cottrell Scholars (far right) further along in their careers have also received recognition for their achievements. The IMPACT Award is given for national impact in science through leadership and service. STAR Awards reward excellence in science teaching and research.

For additional information visit rescorp.org or call 520.571.1111.

Cottrell Scholar Awardees

1. **Vinayak Agarwal**, chemistry, Georgia Institute of Technology / Unlocking Marine Eukaryotic Natural Product Biosynthetic Schemes in Research and Education
2. **Jeanine Amacher**, chemistry, Western Washington University / Investigating Sortase Enzyme Activity and Specificity Using Natural Sequence Variation and Ancestral Sequence Reconstruction
3. **Jeffrey Bandar**, chemistry, Colorado State University / Salt-Promoted Electron Transfer Processes for Reductive Cross-Coupling Reactions
4. **Rachel S. Bezanson**, astronomy, University of Pittsburgh / Building Bridges in the Steel City: Leveraging the Nearby to Follow Galaxies Across Cosmic Time
5. **Laura Blecha**, physics, University of Florida / The Making of a Gravitational Wave Source: Probing the Role of Galaxy Assembly in Black Hole Binary and Triple Formation
6. **Justin R. Caram**, chemistry, University of California, Los Angeles / Chemical Physics Informed Design of SWIR Emissive Molecules
7. **Joel F. Destino**, chemistry, Creighton University / Bottoms Up: Investigating the Growth and Glass-Forming Properties of Germania-Containing Colloids
8. **Daniela Fera**, chemistry, Swarthmore College / Dissecting the Interactions and Conformations of Protein Kinases to Understand Biochemical Signaling
9. **Alex Frañó**, physics, University of California, San Diego / Creating Artificial Angstrom-scale Periodic Potentials Using X-ray Standing Waves and Enhancing Creativity in Physics Education
10. **Thomas L. Gianetti**, chemistry, University of Arizona / Developing a Photo-Rechargeable and Symmetrical Organic Redox Flow Battery
11. **Natalie M. Gosnell**, astronomy, Colorado College / Constraining the Complexities of Stellar Activity with Sub-Subgiant Stars
12. **Christopher H. Hendon**, chemistry, University of Oregon / Inorganic Defects in Metal-Organic Frameworks
13. **Alexis C. Komor**, chemistry, University of California, San Diego / Harnessing Precision Genome Editing Tools to Study DNA Repair Proteins in Live Cells

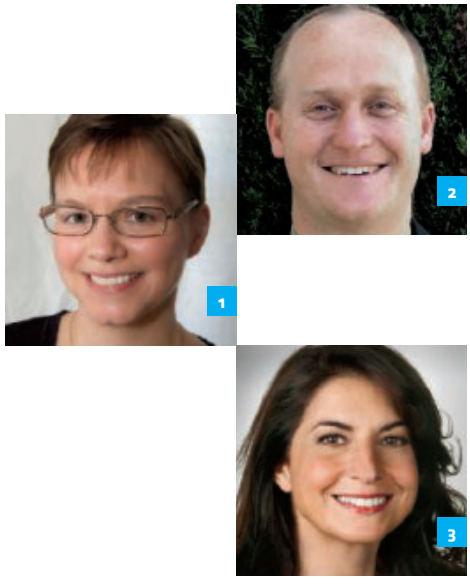


IMPACT and STAR Awardees

1. IMPACT awardee Penny Beuning, chemistry, Northeastern University / Cottrell Scholar 2009
2. STAR awardee Tim Clark, chemistry, University of San Diego / Cottrell Scholar 2007
3. STAR awardee Hanadi Sleiman, chemistry, McGill University / Cottrell Scholar 2002

14. Eli M. Levenson-Falk, physics, University of Southern California / Creating Custom Quantum Environments with Superconducting Circuits (For Beginners)
15. Rosario Porras-Aguilar, physics, University of North Carolina at Charlotte / Label-Free Reconfigurable Microscopy with High Specificity **16. Davit Potoyan**, chemistry, Iowa State University / Uncovering Principles of Bio-molecular Condensation: from Single Molecules to Cellular Organelles
17. Tyler D. Robinson, astronomy, Northern Arizona University / Understanding the True Utility of Prior Mass Constraints in Characterizing Exoplanet Atmospheres and Diversifying STEM at NAU
18. Brian J. Shuve, physics, Harvey Mudd College / Matter-Antimatter Asymmetry from Dark Matter Freeze-In **19. Marcelle Soares-Santos**, physics, University of Michigan / Cosmology with Merging Black Holes and Neutron Stars **20. Ruby May A. Sullan**, chemistry, University of Toronto – Scarborough / Multifunctional and Stimuli-Responsive Nanotherapeutic Platform for Targeted Disruption of Bacterial Biofilms

21. Alexandra Velian, chemistry, University of Washington / Synthesis of Functional Metal Chalcogenide Lattices Using Symmetry-Encoded, Atomically Precise Clusters **22. Rongsheng (Ross) Wang**, chemistry, Temple University / A Fluorine Displacement Based Bioorthogonal Labeling Approach to Interrogate Non-Histone Substrates of "Histone" Deacetylases
23. Leah S. Witus, chemistry, Macalester College / Investigation of Beta-Hairpin Hydrolytic Peptides and Development of an Advanced Undergraduate Scientific Communication Course
24. Joseph M. Zadrozny, chemistry, Colorado State University / Harnessing Ligand-Shell Nuclear Spins to Control Molecular Spin Coherence
25. Gail Zasowski, astronomy, University of Utah / Understanding the Chemical Enrichment of Our Universe: Unifying Evidence from the Milky Way and Other Galaxies



RESEARCH CORPORATION
for SCIENCE ADVANCEMENT
A foundation dedicated to science since 1912.

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, 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, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, 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 **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 **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

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

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

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

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

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

MEDIA CONTACTS
scipak@aaas.org

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

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

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

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

AAAS BOARD OF DIRECTORS
CHAIR Claire M. Fraser
PRESIDENT Susan G. Amara
PRESIDENT-ELECT Gilda A. Barabino
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Rosina M. Bierbaum, Ann Bostrom, Jane Austin Clayton, Laura H. Greene, Kaye Husbands Fealing, Maria M. Klawe, Robert B. Millard, William D. Provine

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **(S)**)

Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena, *Institut Curie*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *UT Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Kiros T. Berhane, *Columbia U.*
Bradley Bernstein, *Mass. General Hospital*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki, *NYU School of Med.*
Mariana Byndloss, *Vanderbilt U. Med. Ctr*
Annmarie Carlton, *UC Irvine*
Nick Chater, *U. of Warwick*
Ling-Ling Chen, *SIBCB, CAS*
M. Keith Chen, *UCLA*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Christian Davenport, *U. of Michigan*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Ulrike Diebold, *TU Wien*
Stefanie Dimmeler, *Goethe-U. Frankfurt*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn*
Jennifer A. Doudna, *UC Berkeley*
Ruth Drdla-Schutting, *Med. U. Vienna*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Jennifer Elisseeff, *Johns Hopkins U.*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *U. of Penn.*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *GWU*
Toren Finkel, *U. of Pitt. Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *UPenn*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Taekjip Ha, *Johns Hopkins U.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. London*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Wash.*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *UC Berkeley*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med.*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Gabriel Lander, *Scripps Res.* **(S)**
Mitchell A. Lazar, *UPenn*
Wendell Lim, *UCSF*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *Wash. U. in St. Louis*
Ke Lu, *Inst. of Metal Res., CAS*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
David Lyons, *U. of Edinburgh*
Fabienne Mackay, *QIMR Berghofer*
Anne Magurran, *U. of St. Andrews*
Asifa Majid, *U. of York*
Oscar Marín, *King's Coll. London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
Jason Matheny, *Georgetown U.*
Heidi McBride, *McGill U.*
C. Robertson McClung, *Dartmouth*
Rodrigo Medellín, *U. Nacional Autónoma de México*
Jane Memmott, *U. of Bristol*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci., Res. & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Daniel Pauly, *U. of British Columbia*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Institut Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UCSF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Lei Stanley Qi, *Stanford U.*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial Coll. London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Joachim Saur, *U. zu Köln*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Wash.*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *UCSF*
Robert Siliciano, *JHU School of Med.*
Lucia Sivilotti, *U. Coll. London*
Alison Smith, *John Innes Centre*
Richard Smith, *UNC* **(S)**
Mark Smyth, *QIMR Berghofer*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Dartmouth*
Ira Tabas, *Columbia U.*
Eriko Takano, *U. of Manchester*
Patrick Tan, *Duke-NUS Med. School*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Kimani Toussaint, *Brown U.*
Wim van der Putten, *Netherlands Inst. of Ecology*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
David Wallach, *Weizmann Inst.*
Jane-Ling Wang, *UC Davis* **(S)**
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* **(S)**
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *UPenn School of Med.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

Science journalism grows up

By the early 1920s, an unlikely pair—a powerful national newspaper publisher and a California-based zoologist—decided that they'd had enough. Enough of half-baked reporting on research results, enough of stories that left readers confused about even the basic principles of science. They wanted something better. They wanted reporting that encouraged a "scientific habit of mind," a citizenry aware of the role of research in everyday life.

However unlikely, the alliance between Edward Willis Scripps, founder of one of America's largest newspaper chains, and Harvard-trained zoologist William Emerson Ritter, ran deep. The two men shared a belief in science as the new century's most powerful transformative agent—and also a belief that scientists were doing a poor job of communicating this. By April 1921, they'd decided on a solution, a venture called Science Service, which would be dedicated to providing smart and positive science stories to the public. The organization they formed a century ago would grow into Society for Science, publisher of *Science News*. True science journalism—independent inquiry into the scientific enterprise and the illumination of research with all its wonderfully complex human interactions—would come much later. But with the founding of Science Service, a new profession did take its first steps, albeit somewhat stumbling ones.

Although scientific societies and organizations supported the new service, researchers themselves remained wary of the often flamboyant journalism of the early 20th century. In 1934, a dozen American science writers formed a National Association of Science Writers, in part to build better relationships with their wary sources, promoting it as a way to identify elite, science-savvy writers from the other journalistic riff-raff. Boyce Rensberger, a former director of the Knight Science Journalism Program, once described this alliance between scientists and journalists as the beginning of the "Gee Whiz" period of science journalism, one that he believed led directly to embarrassing fan-boy coverage of the development of nuclear weapons and the post–World War II arms race.

As Rensberger and others also note, the profession reluctantly let that model go. Science writers were sometimes downright hostile when faced with the environmental downsides of technological development that appeared during the 1960s: air pollution, water pollution, Rachel Carson–driven warnings that unchecked use of pesticides was unsafe, and more. The best science stories, one leading journalist argued, resulted from cooperation with "enlightened industries."

Still, journalistic doubts concerning relentlessly cheery science coverage deepened, and emphasis on telling the whole complicated story also deepened as the profession continued to expand, marked by formation of groups like the European Union of Science Journalists' Associations in 1971, and more.

It's in this moment of doubt that science journalism began to come into its independent own. The last two decades of the 20th century saw a new emphasis on professional training, a growing number of female science journalists (although other forms of diversity have been slow to follow), and newly sharp-edged investigative reporting that looked at everything from the politics of HIV research to space shuttle failures to risky chemical contaminants. As Liza Gross, author of *The Science Writers' Investigative Reporting Handbook*, points out, critics who called science journalists just a bunch of "perky cheerleaders" for researchers were gradually being proved wrong.

> "...science reporters... portray research accurately in both its rights and its wrongs..."

The rise of this century's digital era of communication has served to accelerate change, both in the way writers tell stories, employing tools from podcasting to data visualization, and in their visibility. Science journalists now readily cover contentious areas of science—from climate change to vaccines to the long-standing culture wars around evolution—with clarity and, in turn, deal with furious pushback from skeptics on social media and other platforms.

The original, science-boosting mission of Science Service hasn't been lost. Today, countless "science communicators"—from press officers to scientists themselves—work to foster a positive portrait of science. And there's still a place for journalistic stories about the wonders of science. But the past century has proved that this is not the most important contribution of science reporters. Rather, it is to portray research accurately in both its rights and its wrongs and stand unflinchingly for the integrity of the story. Scripps and Ritter were smart men, and there's a strong argument to be made that they would approve this endpoint.

–Deborah Blum

Deborah Blum
is the director of the Knight Science Journalism Program at the Massachusetts Institute of Technology in Cambridge, MA, USA and the publisher of *Undark Magazine*.
dlblum@mit.edu;
@deborahblum

PHOTO: LEISE JONES

10.1126/science.abj0434

14 of 15

U.S. states not requiring people to wear masks in public recorded relatively high rates of new COVID-19 cases from May to October 2020. None of eight states with high mask wearing had high rates of infection. (*PLOS ONE*)

IN BRIEF

Edited by **Jeffrey Brainard**

The Ingenuity helicopter snapped a selfie of its shadow on Mars's surface, its spinning rotor blades frozen.

PLANETARY SCIENCE

Ingenuity takes otherworldly first flight

The Wright brothers' storied flight at Kitty Hawk had a sequel this week more than 288 million kilometers away: Ingenuity, NASA's $80 million minihelicopter, took a 1-minute test hop on Mars, the first controlled flight of a powered aircraft on another planet. The autonomous 1.8-kilogram machine, the size of a tissue box, spun up its 1.2-meter rotors to more than 2500 revolutions per minute before ascending about 3 meters and hovering in the thin martian air. Ingenuity rotated and took a picture before alighting back on the surface. NASA plans to send Ingenuity, which first landed on Mars on 18 February with the Perseverance rover, on four more flights of increasing height and distance and to use the resulting data to build larger, more ambitious helicopters to explore the Red Planet.

Anthropology group renamed

SCIENTIFIC SOCIETIES | The 90-year-old American Association of Physical Anthropologists (AAPA) has rechristened itself in order to separate today's association from the field's racist and colonial past. At AAPA's virtual annual meeting last week, an overwhelming majority of members voted to delete the word "physical" and become the American Association of Biological Anthropologists. They acknowledged that the old name has roots in the 19th century, when early anthropologists helped create damaging concepts of race by quantifying physical differences among people. The new name conveys that anthropology is now a multidisciplinary biological science that deals with the adaptations, variability, and evolution of humans and their living and fossil relatives, as well as their culture and behavior, according to a statement by the current and past AAPA presidents. "Importantly, the change allows us to reflect deeply on issues of racism and colonialism, which, at times, permeated the field of 'physical anthropology,'" they wrote.

California to fly carbon probes

CLIMATE SCIENCE | California and its partners announced plans last week to launch two satellites by 2023 to spot plumes of planet-warming carbon dioxide and methane. The $100 million Carbon Mapper project, financed by publisher Michael Bloomberg and other philanthropists, will advance efforts to track concentrated emissions of greenhouse gases that rise from sources such as fossil fuel power plants and leaky pipelines. Previous satellites have lacked the resolution, sensitivity, and focus to collect the data officials need in order to regulate the emissions effectively. The new spacecraft will rely on "hyperspectral" imaging spectrometers that can record more than 400 visible and infrared wavelengths, whose patterns can reveal the abundances of certain gases in the atmosphere below.

Very few with shots get COVID-19

PUBLIC HEALTH | A tiny fraction of the U.S. residents fully vaccinated against COVID-19 by 14 April have become infected, the U.S. Centers for Disease Control and Prevention (CDC) said last week. The agency said it expected some "breakthrough" infections and that the low numbers support the value of the inoculations. CDC said it received 5814 reports of such infections in 75 million people vaccinated in 43 U.S. states and territories. Of the infected people, 65% were female, 45% were 60 or older, and 29% were asymptomatic. Seven percent were hospitalized, and 1% died, some from causes unrelated to COVID-19. CDC cautioned that the data from the states reporting might be incomplete. Public health specialists say the infections were more likely to have resulted from weak immune responses to vaccination than to mutations in the virus that let it evade those defenses.

PHOTO: NASA/JPL-CALTECH

Small-scale farming, such as in rice terraces in the northern Philippines, can help sustain forest biodiversity.

PHOTO: JAY DIRECTO/AFP/GETTY IMAGES

NATURAL RESOURCES

Ancient Indigenous land use has helped protect biodiversity

Just 19% of Earth's lands are truly wild, with no history of human impact, a new study shows. In other parts of the globe, however, biodiversity hot spots have survived even where humans thrived, thanks in part to millennia of beneficial land management practices by Indigenous people, these researchers conclude. By 10,000 years ago, humans had already spread across three-quarters of the globe, and their controlled burns, small-scale farming, and other practices may have sustained or even improved biodiversity, according to the analysis of past and present land use, published this week in the *Proceedings of the National Academy of Sciences*. The finding sheds light on a long debate between archaeologists, who cited evidence of this lengthy history, and conservationists, who have insisted that humans did not significantly affect biodiversity until intensive agriculture, urbanization, and deforestation began 200 years ago. Because of the present-day overlap between biodiversity hot spots and lands occupied by Indigenous people, the study bolsters the idea that the growing push to help them regain and retain control over their lands might help protect biodiversity.

U.K. approves 'challenge' studies

COVID-19 | Researchers at the University of Oxford will intentionally reinfect people previously infected by the virus that causes COVID-19 to study their immune responses and symptoms. The "human challenge trial," announced on 19 April, will initially re-expose up to 64 volunteers who previously tested positive for the virus and measure what viral dose triggers new infections. A U.K. government ethics panel approved the study and a similar one led by Imperial College London scientists who are evaluating the performance of COVID-19 vaccines. Such experiments may provide results faster than other trial methods allow.

Remains offered for repatriation

ANTHROPOLOGY | More than 1300 skulls held in a museum collection that was used to justify racism will now be available for return to communities of the people's descendants, the University of Pennsylvania said last week. Samuel Morton started the collection in the 19th century and used studies of its contents to support the idea of white superiority. Many of the crania belonged to enslaved Africans and Indigenous people. In a statement, Christopher Woods, director of the university's Museum of Archaeology and Anthropology, where the Morton Cranial Collection is held, apologized for the "unethical possession of human remains." The museum will work to identify descendant communities and accept requests for the return of any crania in the collection. Repatriation of human remains, especially Black and Indigenous ancestors, "is part of a cultural and social reckoning" about how to address anthropology's history of racism, Woods says.

Archaeology talk draws fire

SCIENTIFIC MEETINGS | A talk last week at the virtual annual meeting of the Society for American Archaeology (SAA) sparked criticism for arguing against a key U.S. law giving Native Americans rights to the human remains and artifacts of their ancestors. Many society members were outraged that SAA gave a platform to what they considered a racist and anti-Indigenous presentation. Some note that this incident comes after a sexual harassment scandal at the organization's 2019 conference. In her talk, SAA member and anthropologist Elizabeth Weiss of San Jose State University said archaeologists "have let creationism into the heart of our discipline" because the law, the Native American Graves Protection and Repatriation Act (NAGPRA), allows Indigenous communities to request repatriation of remains, which they may do partly because of religious beliefs. But archaeologists widely support the law, under which many tribes have collaborated with researchers. In response to the criticism, SAA issued a statement encouraging "the rigorous interrogation of diverse views." SAA President Deborah Nichols later told *Science* the organization's

ASTRONOMY

A fresh effort to glimpse black holes

The Event Horizon Telescope (EHT) collaboration, which in 2019 produced the first image of a black hole's shadow, this week completed another observing campaign, its first in 3 years. Organizers hope their network of radio telescopes will reveal more of the dark heart of the nearby M87 galaxy as well as the Milky Way's center and the quasar 3C 273. EHT must synchronize 10 observatories across the globe in good weather, so its observing window each year is short. Three observatories joined the network this year (including the Kitt Peak 12-meter telescope in Arizona, below), which will sharpen images. Researchers gathered data for more than seven full nights over 2 weeks this month, and EHT spokesperson Eduardo Ros called the results "excellent." Now begins a long wait as recorded data are shipped to Boston and Bonn, Germany, for months of processing before an image might be revealed.

board rejects the viewpoint of Weiss and her co-author and supports NAGPRA.

Canada research plan disappoints

POLICY | The relatively modest research investments outlined in Canada's new federal budget could make it difficult for the nation to recruit and retain scientific talent, Canadian science advocates fear. The multiyear spending plan, announced on 19 April, includes CA$2.2 billion in mostly new funding for life sciences, with much of the money aimed at boosting biomedical applications and vaccine development. (Canada will continue to provide other spending for research this year under multiyear budgets approved in 2018 and 2019.) But analysts worry the increases are too modest compared with much larger ones proposed for the United States by President Joe Biden, and that some Canadian scientists will look for work south of the border. Under Canada's budget, three main research councils will share CA$250 million for a new joint biomedical research program, and the Canadian Institutes for Health Research will get an additional CA$250 million to fund clinical trials. Universities and research hospitals will get CA$500 million for infrastructure such as equipment and buildings. Three programs—an existing artificial intelligence program and two new ones in genomics and quantum science—will each receive CA$400 million in new funding.

Middle East ups research papers

PUBLISHING | Egypt, Iran, Turkey, and other countries in the Middle East and North Africa have boosted their share of scholarly articles in international journals and citations to those papers during the past 4 decades, the Clarivate analytics firm said this month. From 1981 to 2019, the region quadrupled its share of research articles and reviews to 8%; among regions and large countries, only China grew by more. Fifteen of the region's 19 countries had a citation score in 2019 higher than the world average, when adjusted for differences across disciplines; in 2000, almost all had scores well below average.

THREE QS

Reckoning with climate blues

Sustainability scientist Kimberly Nicholas of Lund University found herself struggling with feelings of grief as research by her and others revealed how much climate change will harm agriculture, ecosystems, and human communities. And she discovered she is not alone. In her new book, *Under the Sky We Make: How to Be Human in a Warming World*, she offers insight into how people and institutions can respond to those feelings and the climate challenge. (A longer version of this interview is at http://scim.ag/climategrief.)

Q: How does your experience with grief inform your thinking about climate change?
A: Things are changing beyond recognition right now from climate change. To me, grieving is an important part of the process of acknowledging that. It does draw from my experience of losing a dear friend to cancer, who died at 37. It was a kind of wake-up call [that prompted me] to think about my core values and what matters. But it shouldn't take a terminal diagnosis for life on Earth to wake us up to the urgency of working for climate stability.

Q: Students come to you distraught about harm to ecosystems they hope to study. What do you tell them?
A: The main thing is not to shy away from those conversations. It's not really helpful to deny the reality or not equip them with the tools to face that reality. You have to acknowledge that they're running into a house that is on fire.

Q: You argue for a shift from what you call the "exploitation mindset." What's an example?
A: A big wake-up moment for me came at a climate science conference. Pretty much everyone there, including me, had flown in. The presentations were a litany of depressing things happening because of climate change. I felt like I was at this conference of doctors puffing on cigarettes, but telling our patients to quit smoking! I realized we really have an obligation to model the change that we want to see. So, I have pretty much stopped flying for work. It hasn't meant I can't be a productive researcher.

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

PHOTO: ARASH ROSHANINESHAT

eppendorf & Science

PRIZE FOR NEURO BIOLOGY

2020 Winner
Christopher Zimmerman, Ph.D.
Princeton Neuroscience Institute
For research on thirst and drinking behavior

Application Deadline
June 15, 2021

Now It's Your Turn!

Eppendorf & Science Prize for Neurobiology
The annual Eppendorf & Science Prize for Neurobiology is an international prize which honors young scientists for their outstanding contributions to neurobiological research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by *Science*'s Senior Editor, Dr. Peter Stern. If you are 35 years of age or younger and doing great research, now is the time to apply for this prize.

As the Grand Prize Winner, you could be next to receive
> Prize money of US$25,000
> Publication of your work in *Science*
> Full support to attend the Prize Ceremony held in conjunction with the Annual Meeting of the Society for Neuroscience in the USA
> 10-year AAAS membership and online subscription to *Science*
> Complimentary products worth US$1,000 from Eppendorf
> An invitation to visit Eppendorf in Hamburg, Germany

It's easy to apply! Write a 1,000-word essay and tell the world about your work. Learn more at:

eppendorf.com/prize

eppendorf Science AAAS

AAAS® and Science® are registered trademarks of the American Association for the Advancement of Science, USA. Eppendorf® and the Eppendorf Brand Design are registered trademarks of Eppendorf AG, Germany. All rights reserved, including graphics and images. Copyright © 2021 by Eppendorf AG. Photography: Sameer A. Khan.

IN DEPTH

Where debris flowed from a 1980 eruption, plants and animals have slowly returned. But a road may disrupt those ecological communities.

ECOLOGY

Researchers decry planned road near St. Helens

Project could help avert catastrophic flooding, but threatens research plots

By **Robert F. Service**

When Mount St. Helens in Washington state erupted on 18 May 1980, the initial explosion blew sideways, creating a nearly 600-square-kilometer blast zone—and what has become a prized ecological research area. Dozens of groups have tracked life's reemergence there, one lupine and ladybug at a time. Now, many of those research projects may be endangered. Last month, the U.S. Forest Service (USFS), which manages the land, issued a decision to build a road stretching through the heart of the research area to the banks of Spirit Lake, 5 kilometers northeast of the crater. The agency says the road will service a tunnel that drains the lake to prevent a catastrophic flood, a threat to tens of thousands of people in the valley below.

The USFS decision has set off protests from scientists, including a lawsuit aimed at preventing the road's construction and charges that the agency failed to consult with researchers. "We absolutely recognize the public safety issue," says John Bishop, a botanist at Washington State University, Vancouver, who has been working for 31 years in the blast zone, designated a national monument in 1982 in part to encourage long-term ecological research. But, he says, "We think we can protect public safety and protect the heart of the monument. We've got to do both."

When the eruption occurred, the north side of the mountain gave way, producing the largest landslide ever recorded. The avalanche of debris rushed into 890-hectare Spirit Lake, raising its water level by 60 meters and damming the lake's natural outflow. The lake continued to rise, and officials worried it could cause the debris dam to give way, inundating downstream communities and an estimated $3.65 billion in property, including ports on the Columbia River. In 1985, the U.S. Army Corps of Engineers completed a 2.25-kilometer tunnel through the debris dam to allow the lake to drain into nearby Coldwater Creek. But seismic activity has repeatedly damaged the tunnel, requiring repairs.

On 16 March, Eric Veach, USFS supervisor for the region, released a "notice of decision" saying "the current risk situation at Spirit Lake is unacceptable and action is required." The best option, Veach announced, is to build a 5.5-kilometer-long temporary road from an existing road across the blast zone, known as the Pumice Plain, to Spirit Lake. The road will allow the agency to bring in heavy equipment needed to repair the tunnel, replace a damaged gate that regulates the outgoing water flow, and carry out geological drilling studies of the long-term stability of the debris dam.

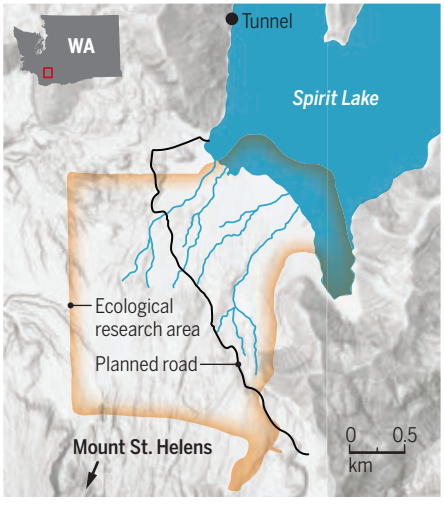
Carri LeRoy, an ecologist at Evergreen State University, Olympia, isn't convinced about the urgency. "They argue it's an emergency," she says. But she and others say USFS could continue its practice of flying repair equipment in on helicopters to maintain the tunnel. "There is no imminent threat."

However, the road's impact on research would be immediate, says Susan Jane Brown, an attorney with the Western Environmental Law Center who filed a complaint with the U.S. District Court of Washington against USFS on behalf of three scientists and four conservation groups. Pumice Plain is home to 992 research plots, investigating how birds, amphibians, insects, and plants recolonized the shattered landscape. "It is a unique opportunity to look at long-term changes after a massive disturbance," says James Gawel, an environmental chemist at the University of Washington, Tacoma, who has worked on the area since 2005.

The road would cut through 25 of the research plots, and USFS says it will directly affect about 4% of the total research area. But LeRoy calls that "a gross oversimplification." The road would also cut across at least

A fractured environment

The U.S. Forest Service plans to build a road through an area scientists treasure for ecology research.



CREDITS: (PHOTO) CARRI J. LEROY; (MAP) N. DESAI/*SCIENCE*; (DATA) U.S. FOREST SERVICE

eight streams formed since the eruption, which drain into Spirit Lake. As a result, it would also impact everything downstream, including the lake, possibly introducing invasive species. "The impact would be massive," Bishop says. "If you want to study the reestablishment of [native] life, that will be lost." LeRoy, who studies the evolution of the newly formed streams, agrees: "There will be no point in studying there anymore."

Veach acknowledges the road would have downsides for science, writing in his decision that the project "reflects a balance between public safety concerns and effects to ongoing research." USFS considered other options, including a shorter road. But it would have required building a new parking lot for trucks and equipment on the lake, adding to the project's ecological footprint. It would also have relied more heavily on helicopters and boats to access repair and drilling sites. The road now planned would extend to a staging site used decades ago for tunnel construction.

USFS's road doesn't represent a final plan for draining Spirit Lake. In 2018, the National Academies of Sciences, Engineering, and Medicine laid out options for a long-term fix, including adding a second tunnel and carving a large channel through the debris dam. The report emphasized the need to include voices of all stakeholders, such as scientists, government agencies, and downstream residents. With the intermediate USFS plan, scientists lament that they have largely been ignored. In April 2020, USFS had issued a finding that the road would have "no significant impact" on local ecology, prompting letters from scientists. Last month's decision dismissed those complaints.

"There has been no public assessment of the different options," Bishop says. He favors continuing to use helicopters to service the tunnel and gate. Gawel agrees: "People fly in drilling equipment to other remote locations all the time. They are building a road only to make it cheap and easy." But Veach's decision noted the helicopter option would pose delays for repairs, as well as risks for crews. USFS declined to make Veach or other personnel available for interviews.

Construction will begin this spring, USFS says. But Brown is hoping the agency will hold off and work with scientists and other stakeholders to come up with other options. Barring that, she says she will file an injunction to block the road's construction.

USFS's announcement may actually make addressing the Spirit Lake threat more difficult by triggering a lawsuit and years of delays, says Robin Gregory, an environmental decision scientist at the University of British Columbia, Vancouver. "That's exactly what you don't want." ■

U.S. POLICY

NIH lifts restrictions on fetal tissue research

Scientists applaud end of intramural research ban and Trump-era review board

By **Kelly Servick**

Researchers who rely on fetal tissue from elective abortions to study human development and disease may now be able to resume work recently stalled by restrictive federal policies. President Joe Biden's administration last week lifted a ban on such studies by researchers within the National Institutes of Health (NIH) and eliminated an ethical review board that had torpedoed funding applications from external researchers.

"I would call it a very welcome return to a socially responsible approach to the use of fetal tissue in research," says Alta Charo, a professor of law and bioethics at the University of Wisconsin, Madison.

The previous policy, announced in June 2019 during then-President Donald Trump's administration, prohibited NIH's in-house scientists from working with human fetal tissue from abortions, which would otherwise be discarded (*Science*, 14 June 2019, p. 1016). The policy also created new hurdles for extramural applicants, including review by an ethics advisory board. That panel was dominated by scientists and ethicists who oppose abortion, and in August 2020, it rejected all but one of 14 applications that scientific reviewers had already deemed worthy of funding.

> **"I would call it a very welcome return to a socially responsible approach."**
> **Alta Charo**, University of Wisconsin, Madison

"It was a farce," says Lawrence Goldstein, a neuroscientist at the University of California (UC), San Diego, and the only member of the advisory board who has spoken out for fetal tissue research. "Highly valuable projects that had been through multiple levels of review already … went to this board to die."

Under the policy change released last week, the board will not convene again. And intramural NIH research involving fetal tissue can resume, an agency spokesperson confirmed. Still in place is a longstanding requirement for getting informed consent from women who donate tissue and a prohibition on financial incentives for donors and clinics that collect tissue.

The tissue has been vital to studies of neurological and infectious diseases and for understanding normal fetal development. Although scientists can model development and disease with other types of cells, fetal tissue is still needed for comparison to validate those models, researchers say. "Humanized mice" containing transplants of fetal tissue can model the human immune system and can be used to test potential treatments for diseases, including COVID-19.

The International Society for Stem Cell Research called the reversal a "return to evidence-based policymaking"—though some Trump-era requirements for justifying the use of fetal tissue in grant applications remain, notes Sean Morrison, chair of the society's public policy committee.

To groups that oppose the use of fetal tissue in labs, however, the remaining restrictions aren't enough. "This reflects a turn to poor ethics as well as poor science," says David Prentice, vice president and research director of the anti-abortion Charlotte Lozier Institute who served on the now-disbanded ethics review board.

For some scientists, the change offers a way forward. Stem cell biologist Hanna Mikkola of UC Los Angeles (UCLA), who studies how blood stem cells emerge during development, had NIH funding for studies with a human fetal tissue component for more than 10 years. But last summer, the new ethical review board recommended against funding a proposal she submitted in August 2019 to map human blood stem cell development.

Mikkola has continued her work with bridge funding from UCLA, but she hasn't hired new postdocs because she couldn't guarantee she'd be able to keep paying their salaries. She is waiting to hear whether NIH will now approve the stalled funding for her work, which could lead to treatments for blood cancers and other disorders. "This is the question I wanted to answer the last 15 years, and now we're getting so close," she says. "We can't stop now." ■

SCIENTIFIC COMMUNITY

Sale of building to Qatar deepens Carnegie rifts

Some scientists see deal and reorganization as "corporatization" of historic institution

By **Meredith Wadman**

A decision by the venerable Carnegie Institution for Science to sell its iconic Washington, D.C., headquarters to the government of Qatar has ignited long-simmering discontent with an imminent restructuring of the 120-year-old organization and with its management.

Earlier this month, after Carnegie President Eric Isaacs announced on 2 April that the nonprofit institution was finalizing the building's sale to Qatar for an undisclosed amount, more than 140 Carnegie scientists, students, and staff members wrote to Isaacs and the board of trustees, noting what they called Qatar's dismal human rights record. They urged them to find another buyer for the 1909 beaux-arts building, a Washington, D.C., landmark 1.5 kilometers from the White House.

The sale was only the beginning of their complaints. "Qatar was the gas thrown on a fire that already was burning," one senior Carnegie scientist says.

"I am very, very concerned" about the cultural change at Carnegie and the top-down nature of the reorganization, says Yixian Zheng, director of Carnegie's embryology department. That plan will consolidate the institute's three life sciences departments into a new, 12,600-square-meter research building in Pasadena, California. "In my opinion, this is a national problem of corporatization of academic institutions. It is destroying people's freedom to do science. And it has played out in a very acute way at Carnegie, which is dear to my heart."

But Isaacs and board Vice Chair David Thompson say the consolidation, which includes a new global ecology–focused collaboration with the California Institute of Technology (Caltech), will propel outstanding science at Carnegie. And they say the sale is essential to funding it. The sale "is part of a multiyear strategic shift for Carnegie," says Thompson, co-founder of Orbital ATK, an aerospace and defense technology company that is now part of Northrop Grumman. "I'm convinced that over the next 5 to 7 years it's going to make a tremendous difference for what we can do."

The departments to be consolidated in Pasadena are plant biology and global ecology, which are currently located at Stanford University, and developmental biology (offically called embryology), now located at Johns Hopkins University. The Carnegie Observatories are already headquartered in Pasadena, and earth and planetary sciences will remain in Northwest Washington, D.C.

Robert Hazen, a Carnegie mineralogist, says the plan makes sense. "I find very compelling the vision that Eric has put forth," he says. "When you start bringing in the concept of ecology and how planets work with individual organisms and how ecosystems work you have a tremendous opportunity … [to do research that is] profoundly important to the way our planet is changing today."

The iconic headquarters of the Carnegie Institution for Science is being sold.

Industrialist Andrew Carnegie founded the institution in 1902 as an independent organization devoted to increasing basic scientific knowledge. Today, 53 Carnegie scientists run labs at five locations; the institution's staff number 400 and its 2021 budget is $87 million.

The letter writers say the impending sale of the building to Qatar betrays the institution's values. "Carnegie … was established to apply 'knowledge to the improvement of mankind,'" one letter read. "Knowing the record of the Qatari government, an oppressive, brutal, and misogynistic regime, we cannot ethically engage in a business transaction with it." An annual State Department report released last month notes that Qatar restricts free expression, criminalizes homosexuality, legally enshrines second-class citizenship for women, and failed in 2020 to protect migrant workers from abuse. "This sale says, particularly to women and minorities, that Carnegie Science is not fully committed to [diversity, equity, and inclusion]," a second letter says.

Isaacs told *Science* that he and the 18-member board of trustees, who unanimously approved the sale to Qatar, "thought long and hard" about Qatar's human rights record. "In the end, the decision was to go with what was best for Carnegie science," he says. "This was a decision to convert as much of our capital into science as possible. … This was a business transaction." The sale will allow Carnegie scientists to do much of their research "with internal funds," rather than grants, Thompson adds.

The Qatari government plans to use the building to house its embassy, according to *Washington Business Journal*. The embassy did not respond to emails seeking comment.

The proceeds from the sale will significantly underwrite the construction of a "brand new, state of the art, biology and chemistry and environmental sciences building" in Pasadena, Isaacs says. Caltech has made the land available free of charge, across the street from the future site of its own new sustainability science building. Carnegie's new building will house up to 25 principal investigators under a new director. Administration staff at its headquarters will primarily relocate to a separate campus in Northwest Washington, D.C.

Isaacs and the board say the consolidation will increase synergy across the departments, that Caltech is enthusiastic, and that both institutions will benefit from collaborations focused on global ecology. The total commitment including construction, additional staff, and moving expenses for employees now located at Stanford and Johns Hopkins will be about $125 million, says Thompson, who is also a Caltech trustee.

But some life scientists, who say the consolidation was planned without their input, are not enthusiastic. "The notion that by just smashing us together in a place near the Caltech campus, brilliance is gonna emerge, where is the evidence for that?" asks Steve

PHOTO: DADEROT/WIKIMEDIA COMMONS

Farber, a molecular physiologist in the department of embryology. "Working on CRISPR and genomes is much different than working on global ecology. It's forcing together people who work at very different scales."

Scientists at the Johns Hopkins Carnegie campus, where embryology is housed today, also worry their crucial access to graduate students will be cut by about 50% because Caltech is a much smaller school with fewer biology and related graduate students.

The Carnegie campus at Johns Hopkins is "a robust place with a long-term future and they are shutting that down," says Susanne Garvey, director of development from 1992 to 2017. She notes that Carnegie will leave behind a 2005 building tailored to the department's needs. "It's like throwing away a piece of clothing that you wore once."

Other alienated scientists say the Pasadena move is in keeping with a series of decisions by Carnegie presidents and the board since 2014. They allege that those decisions undermined the institution's historic culture of small science and independent inquiry.

Some scientists cite a failure to replace outstanding faculty who left Carnegie in response to the changing culture; for instance, the global ecology department now has one principal investigator, down from five in 2014. Others point to the 2017 removal of molecular biologist and former Carnegie President Maxine Singer from the board of trustees. "These are money people whose biggest interest is to increase the endowment," says Wolf Frommer, former director of Carnegie plant biology who left in 2017 and is now at the Heinrich Heine University of Düsseldorf. "They have given up basically listening to the scientists and [are] just following their own path."

In the old Carnegie, "The trustees would come and visit our labs, look in our microscopes and be excited by our work. In the new Carnegie that just stopped," says Marnie Halpern, a developmental neurobiologist who left Carnegie's embryology department last year, after 26 years, to become head of molecular and systems biology at Dartmouth College's Geisel School of Medicine. She left, she says, because she felt she didn't have a voice in institutional decision-making and because "leadership lost sight of [Carnegie's] incredible vision."

But Hazen believes the coming changes are for the best. "The institution is larger than any one building or any subdiscipline," he says. "My sense is that Eric has really listened very, very carefully to everybody and has made some tough decisions. The status quo was not optimizing our use of the endowment. It was not doing the best science we possibly could do." ■

PHOTO: CHRISTOPHE ARCHAMBAULT/AFP/GETTY IMAGES

Researchers at the Pasteur Institute abandoned their top candidate for a COVID-19 vaccine.

BIOTECH

After vaccine failures, France laments biomedical decline

Researchers blame a squeeze in basic research funding and scarce venture capital for biotech startups

By **Tania Rabesandratana**

On 25 January, as France's third pandemic wave gathered force, Christophe d'Enfert, scientific director of the Pasteur Institute, appeared on national TV with a grim duty: explaining how the venerable institute, named after vaccine pioneer Louis Pasteur, had given up on its most advanced COVID-19 vaccine candidate. Around the same time, French drug giant Sanofi said its own contenders were delayed—and that it would cut hundreds of French jobs. Today, France remains the only nation on the U.N. Security Council without a viable vaccine. To d'Enfert, it "brings into question our capacity not only to do very high-level fundamental research, but also to transform this into innovation."

The high-profile failures have cast a spotlight on the problems facing biomedicine in France. Although no one failure could have been predicted, the pattern "is not just bad luck," says Audrey Vézian, a sociologist of biomedicine at CNRS, France's national research center, in Lyon. "It shows that something isn't working in our innovation process." Some experts cite a squeeze in basic research funding and scarce venture capital. Vézian also blames a proliferation of bureaucratic organizations that waste resources and add confusion.

Margaret Kyle, an economist at Mines ParisTech graduate school who co-authored a January study by the Council of Economic Analysis (CAE), a government advisory body, says France ought to be well-positioned to do biomedical research—and to commercialize it. Its education system churns out talented scientists, and it has a national health care system, which provides data that can be deployed in medical research. But recent analyses paint a picture of long-term erosion in public biomedical investment. The CAE study found that public spending on biology and health research has shrunk dramatically since 2011, even as it grew in Germany and the United Kingdom (see chart, p. 332).

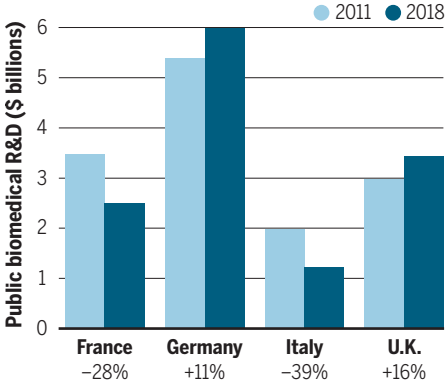
Bruno Canard, a structural biologist who studies coronaviruses at CNRS in Marseille, has felt that decline firsthand. For instance, he says, France only has three of the cryo–electron microscopes (cryo-EMs) that can reveal molecular structures like that of the coronavirus at near-atomic resolution; Germany and the United Kingdom each have about two dozen. And France's national research funding agency, set up in 2005 to provide competitive project-based funding, has seen its budgets drop sharply. Emergency COVID-19 research money has begun to flow, which Canard says has returned his lab's budget to 2003 levels. But by the time pandemic funds were available, he says, "Chinese teams, among others, had already published the first cryo-EM papers in *Science*, *Cell*, and *Nature*."

Biotech startups, critical in pharmaceutical innovation, are also less well funded in France than in its European peers. Funding through France's public investment bank (BPI) and tax rebates can be generous in the early stages of business development, but private funding is too sparse to enable enough companies to grow significantly at later stages. In 2020, French health tech startups each raised only €8 million in venture capital on average, compared with €12 million in the United Kingdom and €25 million in Germany, according to data compiled by the trade group France Biotech.

In the first 5 years after biotechnologist Odile Duvaux co-founded a startup called Xenothera in 2014, she raised €6 million to develop the company's immunotherapies. Things picked up during the pandemic, when BPI gave Xenothera €5.3 million to scale up production of an intravenous COVID-19 antibody treatment and test it in trials; since then, the company raised another €10.3 million and the treatment is being tested in 35 hospitals in France, as well as five other countries.

A sickly budget

Public funding for health research has fallen in France—one factor in the country's slow pace of biomedical innovation.



But Duvaux says those amounts are dwarfed by what U.S. companies are often able to raise. "We're running after nickels and dimes," she says. "We need preorders. That's what the U.S. government does; that's what the U.K. government is doing with [French vaccine company] Valneva: they buy millions of doses before knowing if the products work," and shoulder the risk. In contrast, French leaders tend to be suspicious of biotech companies or oblivious to them, instead favoring well-established academic centers and pharmaceutical firms, Duvaux says. But neither is as nimble as a startup, she says: "A paddleboat or a petrol tanker can't go fast."

French organizations, both public and private, lack "mixed expertise": people with experience with both health and biotech as well as finance, law, and business, according to a 2017 report by the Boston Consulting Group, commissioned by France Biotech. The report blames France's predilection for elite schools that train generalists rather than specialists. France Biotech President Franck Mouthon says administrative burdens and safety procedures, added after health scandals, also weigh down the country's innovation system. "There is money flowing to fund innovation in France, but we need reforms," he says. For instance, members of ethics committees that examine clinical trials applications are drawn randomly to limit conflicts of interest, but that also means they often don't have the relevant expertise, he says.

Changes are afoot. The current administration pledged to reverse what it calls "decades of underinvestment" with a 10-year plan and reform enacted in December 2020. The plan aims to raise R&D spending from 2.2% to 3% of gross domestic product, in line with Germany's, increasing public spending from €15 billion to €20 billion by 2030. It also intends to make research careers more attractive by boosting meager salaries and creating junior tenure track jobs, a novelty in France. (Some researchers have protested the law, arguing that the budget rises are not enough, and that the junior jobs are a downgrade compared with lifelong state employment.)

Change is coming to the startup world as well. Mouthon says the pandemic has smoothed communication with health authorities, helping startups get clarity on regulatory requirements early on. And last summer, the government induced a group of insurance companies and semipublic institutions to pledge €6 billion in tech investment in France through 33 funds; nine of these funds are dedicated to health.

Many scientists hope, cautiously, that the COVID-19 wake-up call will bring lasting improvements. At Pasteur, researchers are pressing on with two other vaccine candidates and other COVID-19–related research, in part thanks to public donations. D'Enfert says the institute is contemplating starting a production unit, like one at the University of Oxford, to make preclinical vaccine batches in house, or adding messenger RNA—the technique behind the Moderna and Pfizer vaccines—to the institute's research portfolio.

D'Enfert hopes the government will beef up research funding and give basic science more "power" and "recognition." "It shouldn't be just about snapping fingers and putting gas in the engine," he says. "It has to be sustained in the long term." ■

PALEONTOLOGY

Ethical controversy swirls around shark fossil from Mexico

Paleontologists raise questions of access to exquisite specimen

By **Rodrigo Pérez Ortega**

About 90 million years ago, an unusual shark cruised the shallow sea covering what is now Vallecillo, in the Mexican state of Nuevo León. After the animal died, conditions on the sea floor were just right to preserve its soft tissue and long, winglike fins. Described last month in *Science*, the shark, named *Aquilolamna milarcae*, made headlines around the world for its surprising similarities to today's plankton-eating manta rays.

But for some paleontologists, the publication—by a team of mostly European scientists—also raises thorny questions about the role of private collections in their field and lingering scientific colonialism. The paper originally said the specimen was purchased by a collector, which is illegal under Mexican law. And a public museum founded by the collector, where the fossil was said to be accessible for study by other researchers, is not scheduled to open until later this year. That could mean other scientists "can't know if what the [authors] described is true or not," says Omar Regalado Fernández, an independent paleobiologist based in the United Kingdom who first raised concerns about the new paper on Twitter.

"A publication with Mexican material is always gratifying," says Felisa Aguilar Arellano, a paleontologist at the National Institute of Anthropology and History (INAH), the federal agency that oversees paleontological research in Mexico. However, she says, "It is unethical to claim the existence of a museum" that is not yet finished, and "it is totally clear that [fossils], considered national heritage, are not marketable."

On 15 April, *Science* published an Erratum noting that starting 1 May, the fossil

CREDITS: (GRAPHIC) A. ALLA *ET AL., LIFECYCLE OF PHARMAEUTICAL INNOVATION*, FRENCH COUNCIL OF ECONOMIC ANALYSIS, ADAPTED BY K. FRANKLIN/*SCIENCE*; (DATA) ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT GOVERNMENT BUDGET ALLOCATIONS FOR R&D

will be housed in the Museo del Desierto in Saltillo, Mexico, until the originally listed public museum is built. For now, the fossil is at the home of Mauricio Fernández Garza, a Mexican businessman, politician, and fossil collector, who says the fossil is accessible to scientists.

Co-author Wolfgang Stinnesbeck, a paleontologist at Heidelberg University, says the team followed all Mexican laws and did not purchase the fossil, which he says is officially registered with INAH, as required.

Fernández Garza says he did not buy the fossil either. The shark turned up in a slab of rock after he purchased it from a fossil-rich limestone quarry in Vallecillo, he says. "I buy the rocks" that might contain fossils, he says; this is legal under Mexican law. He and Stinnesbeck say buying the slabs saves potential fossils from destruction. (Fernández Garza adds that organized crime has recently taken an interest in the Vallecillo site and that other private collectors are now illegally buying its fossils.)

On 8 April, *Science* made an online correction to the paper's Supplementary Materials, updating the fossil's current location and removing the word "purchased" from the Materials and Methods section. "In order to avoid a mix-up and misinterpretation of these circumstances [around the fossil discovery] we removed the word 'purchased,' which—as we know now—caused unexpected problems," Stinnesbeck wrote in an email.

Local preparator Margarito González González, the only Mexican co-author on the paper, cleaned and prepared the fossil. Fernández Garza then allowed Stinnesbeck and his team to study it in 2018. "We obeyed all legislations in Mexico," Stinnesbeck says.

Aguilar Arellano felt that the original paper gave the false impression that buying fossils in Mexico is legal. "We know fossil trade exists, unfortunately," she says. "But we're trying to … mitigate these actions."

For paleobiologist Nussaïbah Raja Schoob of the Friedrich Alexander University of Erlangen-Nuremberg, buying slabs from a quarry known for fossils "seems to be the exploitation of a legal loophole." She adds: "As part of the review process, the journal should already be asking these questions about provenance." Sacha Vignieri, a deputy editor at *Science*, wrote in an email that editors are proactive in asking authors to meet requirements pertaining to legality and ethical use for all paleontology papers.

INAH has other concerns, Aguilar Arellano says. The authors, she says, failed to comply with INAH's protocol in describing a holotype—a specimen defining a new species. Holotypes should be stored in a scientific, public collection, she says, but the fossil is registered to Fernández Garza's private collection. INAH's Paleontological Council says it plans to write to *Science* with its concerns.

Stinnesbeck agrees that holotypes must be stored with public access and says the new arrangements fulfill that requirement. When the paper was submitted, he said, the fossil was destined for the La Milarca Museum, a public museum then scheduled to open at the end of 2020, in time for publication. That museum is meant to be a replica of Fernández Garza's mansion and will house only a couple of fossils from his private collection, according to Fernández Garza. However, the pandemic delayed construction. While dealing with reviewers' queries last year, the museum's construction timetable "slipped our mind," Stinnesbeck wrote in an email. "All of us were convinced that the opening of the museum would happen."

Science's editorial policy requires all fossils described in the journal to be accessible to researchers, but allows fossils to be deposited in private collections as long as "full public access is available." But some paleontologists are not pleased that the fossil is now in a collector's home and will reside in a museum with a private collection in the coming months. In a private collection, "access will always remain at the discretion of the collector," says Raja Schoob. The Society of Vertebrate Paleontology (SVP), in its Code of Ethics, stresses the need for public access to published fossils and has called on journal editors not to publish on privately held specimens, because private collectors can restrict access at any time, says Margaret Lewis, SVP vice president and a paleontologist at Stockton University. The *Journal of Vertebrate Paleontology* does not publish on such specimens, she says, adding that SVP plans to write to *Science* about the issues raised by the paper.

The controversy over *A. milarcae* comes at a moment when some paleontologists are pushing to eliminate the remnants of colonial practices, an issue that inflames the debate over the paper.

In Mexico, before a 1986 law made buying and selling fossils illegal, foreigners "came, they worked, they took things away, and you never heard about it," says paleontologist Gerardo Carbot Chanona at the Eliseo Palacios Aguilera Museum of Paleontology in Tuxtla Gutiérrez, Mexico. He wishes Stinnesbeck's team had collaborated with scientists based in Mexico—as new INAH guidelines, published on 4 February, require foreign researchers to do. He says such partnership is key to following Mexican laws, making sure fossils are collected and housed safely and publicly, and developing Mexico's emerging paleontological community. Lewis echoes that view.

Collaborating with Mexican paleontologists might have helped the authors avoid the issues that arose in this paper, Raja Schoob says.

Eberhard Frey, a paleontologist at the State Museum of Natural History Karlsruhe and a co-author on the shark paper, counters that as long as all local laws are followed, doing research without local scientists "is not problematic." He laments that "everybody is seeking … juristic arguments to get fossils 'nationalized'" rather than making international agreements to study them.

In Mexico, paleontologists say the paper has served as a wake-up call. "It is time" to review ethical codes again, Aguilar Arellano says. For now, the new INAH guidelines requiring scientific collaboration may help avoid new controversies, Carbot Chanona says. "We are not closing the door to foreigners, but we do want them to be regulated in some way." ■

Collector Mauricio Fernández Garza poses with a prized fossil specimen (left) and the cover of the *Science* issue describing the new shark species.

PHOTO: RAÚL PALACIOS/MILENIO

FEATURES

UNLEASHING BIG MUDDY

By sending Mississippi waters on a new course, engineers hope to build new land—and test ways to save a retreating coast

By **Warren Cornwall**

In a swamp at the edge of Louisiana's Barataria Bay, plastic-capped GPS antennas sprout like oversize mushrooms from four small wooden platforms. The gear, which helps scientists monitor changes in the surrounding marsh, is easy to miss in this expanse of water and swampland the size of Delaware. But it represents something even bigger: the beginnings of a grand ecosystem engineering experiment that has been 50 years in the making and could ultimately cost some $50 billion.

If all goes as planned, 2 years from now engineers will punch a massive hole in a nearby levee that holds back the Mississippi River. A 3.5-kilometer-long canal will carry sand and muck from the muddy river into the bay, helping rebuild vast wetlands eroded by sinking land and rising seas. Over 5 decades, researchers forecast that the project—formally known as the Mid-Barataria Sediment Diversion—could move enough sediment to bury the island of Manhattan under 3 meters of muck and create at least 54 square kilometers of new wetlands. The diversion, expected to cost $2 billion, is a critical part of a much larger effort aimed at preventing coastal Louisiana, and the human and wild communities it supports, from slipping beneath the sea.

"There's is nothing like [it] anywhere in the world," says coastal geoscientist Torbjörn Törnqvist of Tulane University, one of a small army of researchers who have helped shape the project through years of fieldwork, computer models, and even the use of a giant replica of the Mississippi River in Massachusetts. "It's going to be completely new."

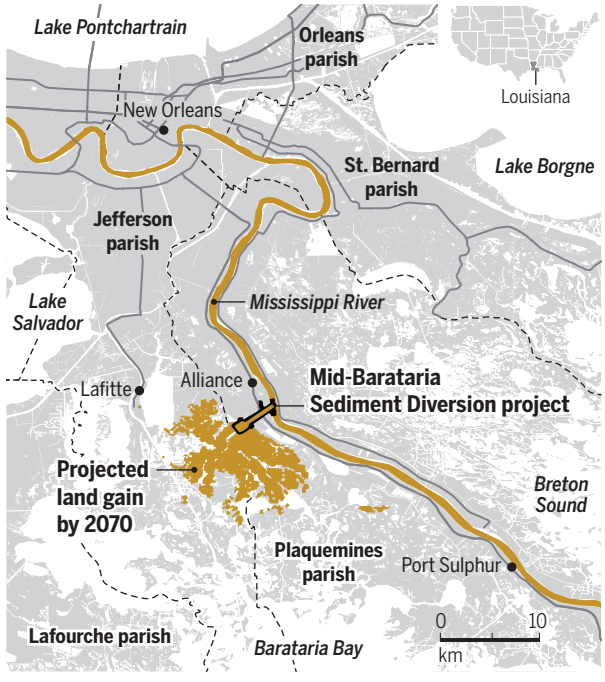
The project, which will be financed with money paid by oil giant BP after the 2010 *Deepwater Horizon* oil spill, last month reached a key regulatory milestone and could begin by 2023. But it is not universally loved. Fishers who make their living pulling oysters and shrimp from nearby waters fear the influx of freshwater could harm their livelihoods. Local officials are questioning the cost. And some scientists are skeptical it will achieve the desired results, given the complexity of trying to tinker with one of the world's largest and most dynamic rivers.

But others are thrilled that the experiment may finally get underway, saying it could help heal the Louisiana coast and shape similar restoration projects else-

PHOTO: GERRIT VYN/NPL/MINDEN PICTURES

From water, land

The state of Louisiana plans to divert sediment from the Mississippi River into Barataria Bay (left), creating an estimated 54 square kilometers of marshland over the next half-century (brown, on map).



where. "There's no question that this is huge," says John Lopez, a coastal scientist who was an early architect of coastwide restoration proposals for Louisiana and now runs a private consulting firm that advises environmental groups about the issue. He hopes it will become "a model for other areas in the world."

THE PROBLEM the Barataria diversion aims to address has been 150 years in the making. For millennia, the Mississippi built coastal Louisiana, depositing dirt and sand as it flooded and flipped from one channel to another. This land building explains why the river's delta pokes into the Gulf of Mexico, a swampy snout teeming with wildlife, marshy grasslands, and watery bayous darkened by ancient cypress trees.

In the late 1800s, people started to build a network of levees along the river's southern reaches to reduce flooding and ease navigation. Today, the Mississippi is little more than a massive canal, straitjacketed by earthen walls and traversed by massive freighters moving cargo bound for places as far north as Wisconsin. As the river pours into the Gulf of Mexico, just a trickle of its waters still flow into surrounding bays such as Barataria to the west and Breton Sound to the east.

Cut off from their steady supply of sediment, these marshlands have been sinking for 100 years. By 2016, the state had lost an estimated 4800 square kilometers of coastland since the 1930s. Nearly one-quarter of that loss occurred in Barataria Bay, and nearly 500 square kilometers in nearby Breton Sound, according to the U.S. Geological Survey.

Decades ago, scientists first suggested trying to restore the marshland by re-engineering part of the Mississippi. In 1975, Sherwood "Woody" Gagliano, a geographer and private environmental consultant who pioneered research into Louisiana's coastal erosion, mapped an elaborate system of levees and canals that could funnel freshwater and sediment into Barataria Bay. "The mistakes that we have made in our coastal areas in the past were largely through ignorance," wrote Gagliano, who died in 2020. "If we … [continue] to make the same errors, destroying the invaluable renewable resources, the action is inexcusable."

For decades, however, little happened. Then, in 2005, Hurricane Katrina slammed into southern Louisiana, killing more than 1800 people, causing $161 billion in damage, and devastating New Orleans, just north of Barataria Bay. The loss of wetlands that had once blunted storms was partly to blame, scientists agreed. And restoration plans previously confined to scientific conferences and bureaucratic meetings suddenly took center stage. In 2007, state lawmakers endorsed the first iteration of what has become known as the coastal master plan, a 50-year initiative to salvage the state's coastline from the forces eating away at it: sinking land, rising seas, and channels dug for the oil and gas industry.

The plan includes a variety of measures: rebuilding barrier islands with new sand, hauling dredged muck to replenish drowning marshes, strengthening levees and flood barriers, and raising buildings above projected flood levels. But the most novel and ambitious piece is an updated version of Gagliano's vision of harnessing the Mississippi's power to build new land. In its latest iteration, the plan calls for creating 11 diversions along the Louisiana coastline. The first and biggest would be the massive mid-Barataria diversion, which would puncture the river's western bank.

From the beginning, money posed a major barrier. Completing all the projects would cost at least $50 billion, the state estimates, with the 11 diversions alone costing $5 billion.

Then BP's *Deepwater Horizon* drilling platform in the Gulf of Mexico exploded, unleashing the world's largest oil spill. The resulting fines promise to pour more than

CREDITS: (MAP) K. FRANKLIN/*SCIENCE*; (DATA) *MID-BARATARIA SEDIMENT DIVERSION ENVIRONMENTAL IMPACT STATEMENT*, U.S. ARMY CORPS OF ENGINEERS, 2021

In a warehouse in Massachusetts, a scale model of the Mississippi River has helped researchers test the design of diversion projects aimed at building wetlands.

PHOTO: TONY LUONG/*THE NEW YORK TIMES*/REDUX

$7 billion into Louisiana's coffers, with much of the money earmarked for environmental restoration. At last, planners not only had a vision, but also the money to start to realize it. Last month, the plans for the mid-Barataria diversion passed a key landmark with the release of the preliminary draft of a massive, 7780-page environmental impact study by the U.S. Army Corps of Engineers (USACE). Although the study found the project could have some negative impacts on fisheries and wildlife, diversion proponents welcomed the finding that it would build significant new marshlands. That could prove critical for winning federal permits.

IN THE MID-1980s, wildlife biologist Andy Nyman witnessed one of the first small experiments to examine whether punching a hole in a Mississippi levee could help rebuild bay wetlands. The state Department of Wildlife and Fisheries excavated gaps in land separating the river from several bays near where the river flows into the Gulf. At first, he recalls, he was unimpressed. After 3 months, there was little evidence of success. "I was like, 'Man, this thing hasn't done anything.' I thought it was a tremendous waste of time."

Twenty years later, however, Nyman, who is now at Louisiana State University (LSU), Baton Rouge, took a boat out to the area near the river's mouth. He discovered more than 100 hectares of reeds, cattails, and seagrasses now blanketed what had been open water. "It was so cool."

Many scientists predict the mid-Barataria diversion will conjure such scenes on a much bigger scale. But other experiments have shown forecasts are tricky. Just downstream of Barataria Bay, near the river's mouth, an effort called the West Bay project has helped create a string of sandy islands covering some 160 hectares. But when work began there in 2003, researchers expected nearly 10 times more land to form by now.

What happened? One problem was that water poured through the diversion so quickly that it flushed much of the new sediment into deeper water, says Alex Kolker, a coastal geologist at the Louisiana Universities Marine Consortium who has studied West Bay. On top of that, the land beneath the bay was sinking unusually fast because of the underlying geology, making it hard for the new sediment to compensate. Land did start to emerge after USACE built artificial islands in the bay with dredged sand. The islands then acted like speed bumps, slowing the diverted water and allowing more sediment to accumulate.

Such complexities make it hard to predict exactly how much land the new Barataria diversion will create. Last year, officials at Louisiana's Coastal Protection and Restoration Authority released a figure of up to 121 square kilometers of land over 50 years, based on hydrologic models developed by scientists in the Netherlands, a country renowned for its water management. In its recent environmental study, however, USACE cut that number by more than half, to just 54 square kilometers. The reduction reflected higher projections for sea level rise.

Even the latest estimate could be overly optimistic, data from the antenna-sprouting platforms have suggested. In 2016, a team led by Törnqvist drilled three core samples, each as wide as a saucer, from the muck beneath the platforms. The deepest plunged almost 40 meters, through 10,000 years of sediments. Scientists then mounted the GPS units on poles cemented into the holes at different depths, to measure whether the marsh is sinking and by how much.

Analyses of sediment cores extracted from the deepest hole suggest buried layers, particularly peat, will compress when weight is added. As a result, new sediment might cause the marsh to sink, erasing much of the gains. Studies of the core samples suggest the land gain over 50 years could be as little as 1 centimeter, or as much as 1.75 meters, says Molly Keogh, a wetland geologist who led the research while earning a Ph.D. at Tulane. The range is large in part because of uncertainty about how much sediment the river will deliver. (Keogh's estimates don't include sea level rise, which could further erode the gains.)

Biology adds an uncertainty of its own. Eugene Turner, a coastal ecologist at LSU, notes that the Mississippi is often loaded with nutrients washed from farm fields farther upstream. As the new canal dumps those nutrients into the bay, he fears they could cause marsh plants to grow too fast to develop root systems hefty enough to hold sediment in place, leading to marsh erosion. "My concern is [the additional nutrients] will overwhelm the benefits," he says.

Tracy Quirk, a wetland plant ecologist at LSU, has been testing that idea. She has been adding different amounts of nutrients and sediments to more than 130 wetland plots in Barataria Bay, each 1 square meter in size. Two years of observations suggest even low levels of added nutrients catalyze root growth, she says. But, so far, she isn't seeing signs that those roots are weaker.

Far from Louisiana, engineers are running experiments aimed at reducing other uncertainties. In a warehouselike building in Massachusetts, they have built a model of the Mississippi and the planned Barataria diversion canal at 1/65th scale. It even includes canoe-size models of the giant freighters that ply the waterway. "Numeric models are good, but they don't get everything perfect," says Dan Gessler, a civil engineer and vice president at the Alden Research Laboratory, the firm that built the model in Holden, Massachusetts.

Using grains of plastic to simulate sand and silt, the researchers have studied how the water will move sediment from the river to the bay. One goal is to fine-tune the design of the canal to cope with the region's relatively flat topography, which could cause the water to slow and drop sediment too soon, clogging the waterway. They also want to be sure the diversion doesn't cause sand bars to form in the main river, disrupting its flow and blocking ships. "When you're spending over a billion dollars," Gessler says, "you want to be really confident that it's going to work exactly the way [you] expect it to."

But only the full-scale experiment will really dispel the questions, Törnqvist says. "The real interesting thing is going to be when the diversion actually gets going," he says. "We just can't wait."

THE UNCERTAINTIES OF POLITICS, however, could get in the way. The overall coastal plan enjoys bipartisan support from state and federal lawmakers. Major environmental groups, including the Environmental Defense Fund and the National Audubon Society, have endorsed the diversions. "This is a critical part of restoring the natural processes that built and sustained Louisiana," says Alisha Renfro, a sedimentologist with the National Wildlife Federation.

But some local governments and fishing interests warn that the mid-Barataria diversion threatens their way of life. This month, the governing council for Plaquemines parish, on the eastern edge of Barataria Bay, voted eight to zero to protest the project. "The whole … concept is going to devastate our fishing and seafood industry," says George Ricks, founder and president of the Save Louisiana Coalition, an alliance of local seafood interests. "There won't be nothing left once they do the diversion."

Scientists don't dispute that adding more freshwater to the bay's ecosystem will bring changes. Marsh plants that like brackish water, and some species of birds and fish, will likely benefit, Nyman says. "But there will be immediate costs" to species that thrive in saltier waters, he adds. For example, oysters and brown shrimp, mainstays of the local fishing industry, could be pushed out of their current ranges. The salinity drop could also cause skin diseases in the bay's 2000 dolphins, reducing their survival by as much as one-third, according to scientists at the National Oceanic and Atmospheric Administration.

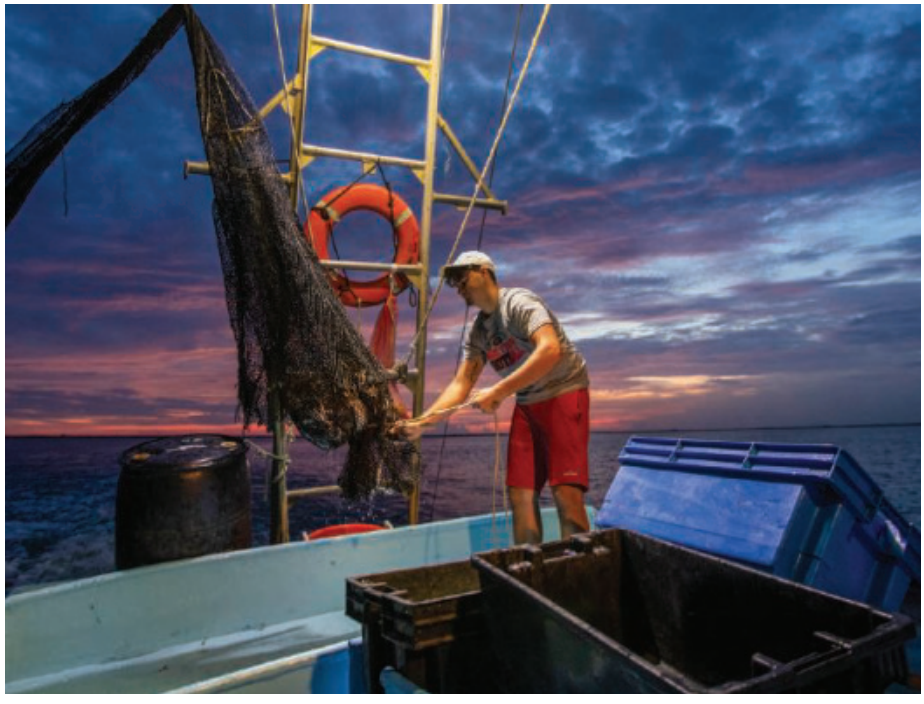
Other critics question the diversion's cost, given the potential economic upheaval. "Holy mackerel! We're going to spend $2 billion for 12,000 acres?" says coastal oceanographer Joe Suhayda, who retired from LSU in 2002 and now consults for several local governments, including Plaquemines parish.

Many observers doubt those downsides will derail the diversion. But even if the project works as promised, they note a sobering reality: Even the most optimistic restoration scenarios see the state losing far more coastline than it will gain.

At the state's coastal restoration agency, officials often point to "the big red map." It forecasts what Louisiana's coastline will look like in 50 years, even if the state completes everything on its $50 billion to-do list. Small patches of green indicate wetlands it expects to save or restore, whereas a thick red fringe along the coast highlights marshes it expects to lose. All told, the state could protect or restore about 2000 square kilometers of wetland, the agency estimates, but other parts of the coast could lose more than 3700 square kilometers. In Barataria Bay alone, wetlands will shrink from 1500 square kilometers today to just 346 square kilometers by 2070 even with the diversion, according to USACE.

"We stopped promising a long time ago that we were ever going to be able to re-create a historical coastline," says Jim Pahl, a plant ecologist and senior scientist at the state coastal authority. Yet diversions can play an important role, he says. Although pumping in sediment with heavy equipment can rebuild patches of marshland, the land will continue to sink and disappear. Because the diversions keep delivering new sand and mud each year, Pahl says, they are critical for the state to "maintain anything resembling a sustainable system."

Törnqvist agrees, even though he concluded in a 2019 study published in *Science* that efforts to save the Louisiana coast are probably doomed in the very long run if seas continue to rise. The mid-Barataria diversion and similar projects, he says, will be worth the cost if they buy time for vulnerable cities in the delta, including New Orleans, to adapt to climate change. "A few more decades can make the difference in the long run," he says, "between managed retreat and complete chaos." ■



Shrimpers oppose plans to divert more freshwater into Barataria Bay because it could reduce their catches.

PHOTO: DREW ANGERER/GETTY IMAGES

EDUCATION

Reframing rankings in educational assessments

Ability and test-taking behavior should be disentangled and jointly reported to improve interpretation and fairness

By **Steffi Pohl[1], Esther Ulitzsch[2], Matthias von Davier[3]**

Educational large-scale assessments (LSAs), such as the Programme for International Student Assessment (PISA), with their widely publicized country rankings, are central sources of information used in policy interventions in education. Their main aim is to assess competencies needed to meet real-life challenges (*1*). Current reporting practices, however, confound differences in test-taking behavior (such as working speed and item nonresponse) with differences in competencies (ability). Furthermore, they do so in a different way for different examinees, threatening the fairness of comparisons, such as country rankings (*2*, *3*). We argue that test-taking behavior is not a nuisance factor that may confound measurement, but an aspect that provides important information on how examinees approach tasks, which is relevant for real-life outcomes. Disentangling and reporting all of these factors as part of a portfolio of performance could result in fairer comparisons across groups and also allow for a better understanding and valid assessment of competencies, as well as for more tailored interventions, targeting possible causes of low performance.

THREATS TO INTERPRETABILITY AND COMPARABILITY

Speed–ability trade-off

It has long been recognized that there is a "speed–ability trade-off," by which an examinee may choose a lower speed (reflected in the time per response on a test) and obtain higher ability (reflected in the correctness of responses at a chosen speed level) or vice versa (*4*). Each examinee may have a different speed–ability trade-off curve and may end up at a different location on that curve depending on the particular testing situation.

If, for example, examinee E1 has a higher ability level for any given speed level as compared to examinee E2, but chooses a higher speed than examinee E2, observed ability for examinee E1 may be lower than for examinee E2. This results in test scores that reflect not only examinees' ability but also their test-taking behavior (i.e., the chosen speed). Analyses of LSAs typically do not take differences in speed into account and falsely imply that examinees who show identical ability are equally proficient, even if they perform at a different speed level.

PHOTO: GARETH FULLER/PA WIRE/AP IMAGES

Scoring of large-scale educational assessments can confound ability and test-taking behavior, threatening the fairness of comparisons across countries.

Missing responses due to time limits

For tests with time limits, slower speed may result in missing values as a result of examinees not reaching the end of the test. Consider examinees who complete all items before time expires (E5), require exactly the allotted time to complete all items (E4), or are unable to complete all items before time expires (E3). If we score these missing values as incorrect, we penalize E3 on the basis of E3's speed level. Thus, the performance score estimated for E3 reflects both ability and speed. For E4 and E5, who do not show any missing values, speed is not incorporated in the overall performance score (i.e., there is no reward for being fast), but solely ability at the chosen speed level is estimated.

Thus, scoring not-reached items as incorrect—as done in some LSAs—results in scores that differ in their meaning, depending on whether examinees do or do not show missing values. This jeopardizes the comparability of performance scores across examinees and, thus, fairness. If missing values are ignored in the estimation—another approach often adopted in LSAs in which items with missing responses are treated as if they had not been administered to the respective examinee—one would wrongly ignore the missing-data mechanism and end up obtaining biased ability estimates for examinees with missing values (*5, 6*).

[1]Freie Universität Berlin, Berlin, Germany. [2]Leibniz Institute for Science and Mathematics Education, Kiel, Germany. [3]Boston College, Chestnut Hill, MA, USA.
Email: steffi.pohl@fu-berlin.de

Item omission

Omitted items are items that examinees decide not to respond to. Consider two examinees, E6 and E7, who operate at the same speed level and show the same ability but differ in their response propensity, i.e., their tendency to respond to versus to omit items. Whereas E6 shows a high response propensity, responding to all items administered, E7 shows a low response propensity, omitting some of the items on the test. If we score omitted items as incorrect, we will opt for different performance scores for E6 and E7, the former being ability and the latter a combination of ability and response propensity (*3*). If we ignore missing values in the estimation, ability estimates will be biased (*5*). Again, both treatments of missing values would result in an unfair comparison.

There is general agreement that such threats to interpretability and comparability should be avoided. It is with the rise of computer-based assessment, the possibility of collecting additional data on test-taking behavior, and the development of models for such data that these threats can be better met.

A PROFILE OF PERFORMANCE

For an adequate interpretation of the scores, it is necessary to clearly define and disentangle different aspects of performance, and estimate and report these aspects in the same way for all examinees. We argue that test-taking behavior is not a nuisance factor but may provide valid information on how persons approach tasks in general. Performance should, thus, be reported as a profile of aspects involved. Note that we do not aim to adjust competence scores for test-taking behavior. Rather, we aim at (i) being transparent in depicting the aspects affecting test performance, (ii) enhancing fairness by using competence scores in the same way for all examinees when making comparisons, and (iii) increasing transparency and validity of assessment results. This allows for a better understanding of performance, as well as for more tailored interventions.

If an overall performance score is of interest, for ensuring fair comparisons across groups (e.g., countries), the different factors influencing test performance need to be incorporated in this overall score in the same way for all examinees. How these different factors are weighted should be decided on the basis of subject matter considerations rather than implicitly by statistics.

The speed–accuracy + omission (SA+O) model (*7*) aims to account for not-reached items owing to time limits and item omissions in the estimation of ability and speed. Besides ability, this model estimates speed and response propensity. We use this model to describe performance as a profile of different aspects. A composite score reflecting all aspects of test performance in one score can still be constructed, but—unlike traditional scoring—(i) we can incorporate our ideas on how this composite score should reflect the different aspects of performance and (ii) it reflects these different aspects of performance in the same way for all examinees.

To illustrate the approach, we used responses and response times to tasks of a mathematical literacy test administered in PISA 2018 in Australia (number of examinees N = 7659), Switzerland (N = 3138), and Italy (N = 6388) (see supplementary materials for details on data and analyses). Examinees from Australia, Switzerland, and Italy spent a median of 69.60, 82.45, and 86.94 s per item, respectively. In Australia, Switzerland, and Italy, only 0.6, 1.3, and 0.7% of item responses were missing as a result of not reaching the end of the test, whereas 6.5, 7.8, and 11.2%, respectively, were omitted.

Applying the PISA scaling procedure (*1*), we recoded item omissions as incorrect and ignored not-reached items. We included the examinee-specific proportion of not-reached items as a predictor of latent ability (*1, 6*). Using the SA+O model, we provided profiles based on the standardized country-level means of ability, speed, and response propensity. To compute composite scores, we combined standardized country-level means of ability, speed, and response propensity for illustrative purposes with two different sets of arbitrary chosen weights.

First, we implemented weights of 0.7, 0.2, and 0.1, respectively, for the three aspects. This reflects a large focus on ability but also accounts for speed and to a small extent for response propensity. Illustrating the impact of weights, in a second analysis, we implemented weights of 0.8, 0.0, and 0.2, respectively, reflecting an even larger focus on ability, none on speed, and some on response propensity. This partly mirrors the implicit ideas of the PISA scoring,

in which omissions are punished (i.e., a higher response propensity is rewarded) and not-reached items (speed) are ignored. However, in the present composite score, the impact of the different aspects is made visible and can directly be decided on.

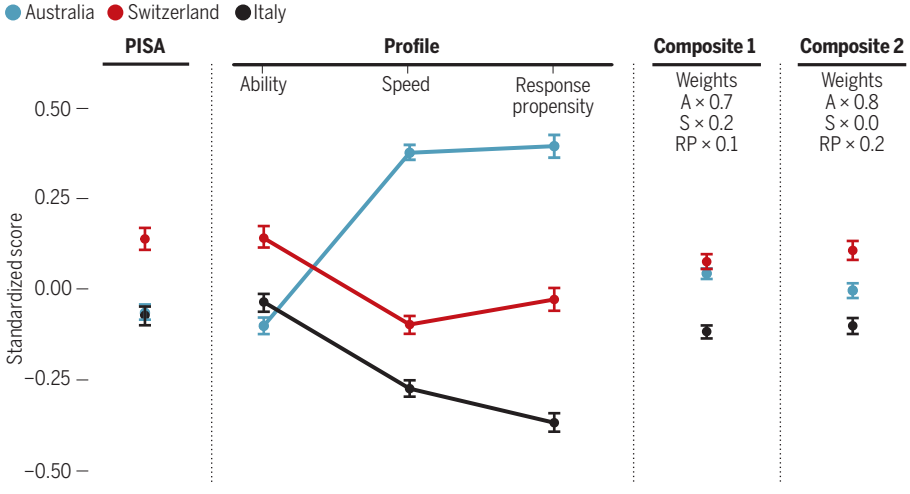
Evaluating the profile of performance aspects for all countries (see the figure), we find that Italian students seem to work more thoroughly on items with the price of being slow and not responding to all of them. By contrast, Australian students take on a more efficient approach by answering most of the items in rather little time, sacrificing accuracy. Swiss students are able to achieve a high accuracy and to respond to most of the items in rather little time, thus combining high efficiency with high accuracy. Whereas ranking Switzerland and Italy is straightforward, as Switzerland outperforms Italy in all three aspects of performance, ranking Australia is more challenging.

Italian and Swiss students show a higher ability than Australian students, but they achieve this by taking more time and responding to fewer tasks. Thus, if we were to focus on ability only, we would rank Switzerland and Italy higher than Australia, neglecting differences in time taken and items omitted. Using the PISA scaling, which in some form combines ability and response propensity in the overall score (as item omissions are scored as incorrect), Australia scores lower than Switzerland and ranks at about the same level as Italy. Under this approach, we would (mis-)interpret the scores as ability scores and would not be able to see that there is variation in test-taking behavior.

In contrast to the PISA scaling, the composite scores take into consideration all three aspects of performance in the same manner for each country. Using the first composite score, which also incorporates speed, Australia and Switzerland rank at a similarly high level. If ability and response propensity are weighted even higher and speed not at all (composite 2), Australia ranks in between the two other countries. Thus, the comparison will depend on how the different aspects of performance are weighted in the composite.

Impact of choice of analysis on country rankings

The figure shows standardized scores derived from the 2018 Programme for International Student Assessment (PISA) mathematical literacy test. Average estimated performance scores with 95% credibility interval are shown when using the PISA scoring approach; a profile of different aspects of performance (ability, speed, response propensity) based on our proposed approach; and composite score estimates with two different schemes to weight ability (A), speed (S), and response propensity (RP). See supplementary materials for details on data and analyses.



IMPLICATIONS FOR REPORTING

Disentangling and reporting different aspects of performance allow for—what we believe—a fairer comparison of groups as well as the derivation of more tailored interventions. Furthermore, these profiles may better reflect performance in situations outside of testing sessions and contain relevant information on examinees' skill sets. For example, nonresponse behavior may predict later real-life outcomes such as educational attainment and employment, supporting the relevance of this aspect of performance in assessments (*8*).

If stakeholders are interested in a single score, a composite score may be used that combines different aspects of performance in the same manner for all examinees. The extracted aspects could be weighted using either substantive considerations or an optimality criterion that maximizes the validity of the composite score—for example, its relation to later school dropout, final grades, or other policy-relevant variables. Decisions about the weights should be made before data collection and with a clear rationale in order to be transparent and to prevent choosing weights to optimize a desired ranking result. Both low speed and low response propensity may reflect different underlying mechanisms, such as low test motivation, low test wiseness, or low competence. To understand which information can be gained by assessing test-taking behavior and to decide on the weighting of the different aspects describing performance, further research is needed to investigate the relevance of these aspects for real-life outcomes.

Although the conceptual idea is clear, further research is needed on the analysis models. The model presented here is just one example for how to disentangle the different aspects of performance. It does not consider all aspects of test-taking behavior. Future research may provide models that incorporate other prominent types of test-taking behavior—such as rapid guessing (*9, 10*), quitting (*11*), number of item revisits (*12*), or varying speed across the test (*13*)—into one model. This requires formulating profound assumptions on test-taking behavior and appropriate methods that enable estimation of such complex models. ■

REFERENCES AND NOTES

1. Organisation for Economic Co-operation and Development, *PISA 2018 Technical Report*, (OECD Publishing, 2020).
2. J. D. Singer, H. I. Braun, *Science* **360**, 38 (2018).
3. S. Pohl, M. von Davier, *Front. Psychol.* **9**, 1988 (2018).
4. W. J. van der Linden, *Psychometrika* **72**, 287 (2007).
5. C. O'Muircheartaigh, I. Moustaki, *J. R. Stat. Soc. Ser. A Stat. Soc.* **162**, 177 (1999).
6. N. Rose, M. von Davier, B. Nagengast, Psychometrika **82**, 795 (2017).
7. E. Ulitzsch, M. von Davier, S. Pohl, *Multivariate Behav. Res.* **55**, 425 (2020).
8. C. Hitt, J. Trivitt, A. Cheng, *Econ. Educ. Rev.* **52**, 105 (2016).
9. E. Ulitzsch, M. von Davier, S. Pohl, *Br. J. Math. Stat. Psychol.* **73** (Suppl 1), 83 (2020).
10. C. Wang, G. Xu, *Br. J. Math. Stat. Psychol.* **68**, 456 (2015).
11. E. Ulitzsch, M. von Davier, S. Pohl, *Educ. Psychol. Meas.* **80**, 522 (2020).
12. U. Bezirhan, M. von Davier, I. Grabovsky, *Educ. Psychol. Meas.* **81**, 363 (2021).
13. J.-P. Fox, S. Marianti, *Multivariate Behav. Res.* **51**, 540 (2016).

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) under grant PO 1655/3-1. We thank T. Götz, S. Krumm, M. Brunner, T. Asalti, and anonymous reviewers for helpful comments.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/338/suppl/DC1

10.1126/science.abd3300

GRAPHIC: N. CARY/*SCIENCE*

ENGINEERING

Distributed manufacturing for and by the masses

Networked systems, artificial intelligence, and an engaged public can enable mass customization of local products

By **Chinedum E. Okwudire** *and* **Harsha V. Madhyastha**

Distribution and democratization represent two complementary paradigms that are gaining increasing attention in manufacturing. Distributed manufacturing (DM) allows for geographically dispersed production, often at small scales and near the end user. Democratization enables large populations to engage in manufacturing. Massively distributed manufacturing (MDM), which combines these paradigms, is performed on demand by a large network of people located anywhere. Rather than rely on mass production in centralized factories, MDM promises to improve the responsiveness and resilience of manufacturing to urgent production demands (such as emergencies like pandemics); promote mass customization and cost-effective, low-volume production; gainfully employ many informally trained citizens in manufacturing (such as through the gig economy); and reduce the environmental footprint of manufacturing by producing items near their points of use. The Fourth Industrial Revolution will play an important role in enabling MDM by way of cyber-physical operating systems (CPOSs).

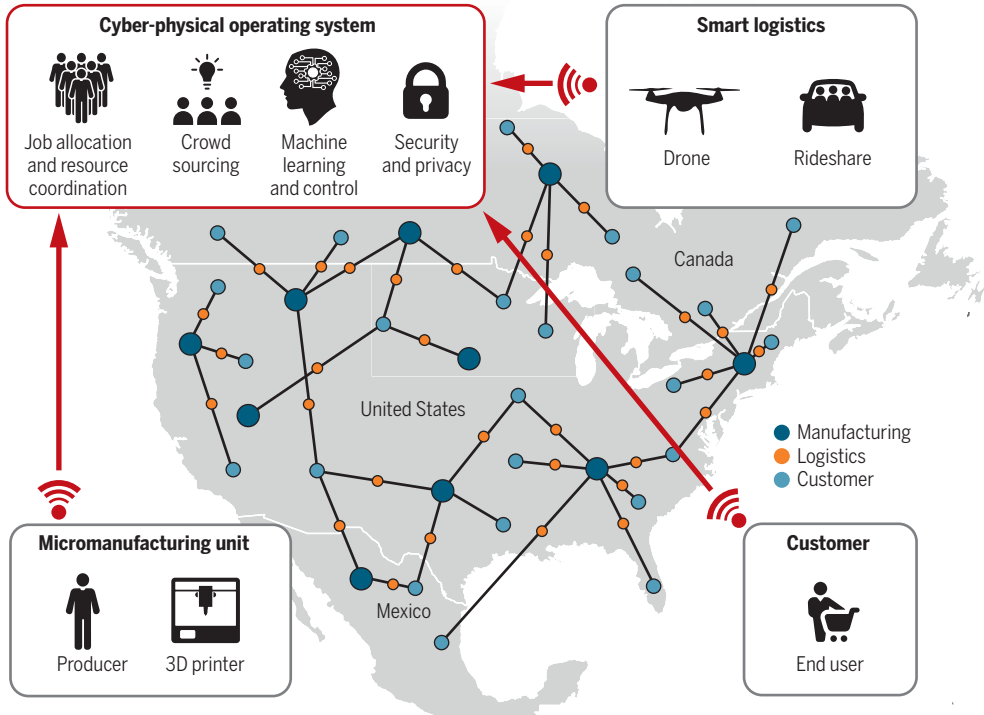
From the First Industrial Revolution in the 18th century onward, manufacturing has been carried out predominantly through mass production in centralized factories, often far from the end user. Mass production enables large quantities of products to be produced with standardized quality, high productivity, and low cost. However, in the face of urgent demands or disruptions, it lacks flexibility, agility, and resilience and cannot readily provide consumers with personalized products in small quantities (mass customization) (*1*). Moreover, its environmental footprint is large, mainly because it often requires raw materials and finished goods to be transported over long distances.

During the past decade, there has been growing interest and activity in distributed and democratized manufacturing as alternative or complementary paradigms to mass production. DM has been emphasized by the United Nations International Development Organization (*2*), the World Economic Forum (*3*), and other major agencies (*4*, *5*) as critical to the future of manufacturing. Several companies engaged in DM, such as 3D Hubs, 3Diligent, Fast Radius, and Xometry, have sprouted. Xometry, for example, enables its customers to access the manufacturing capacity of a network of >5000 carefully curated partners—typically small- and medium-sized enterprises—distributed across the world. In terms of democratization (*6*), perhaps the most compelling example is the proliferation of desktop three-dimensional (3D) printers, which currently retail on average for ~$1000 (*7*), which is within the purchasing power of large portions of the population. In 2019, >700,000 desktop 3D printers were sold globally (*7*). These printers can now be found in homes, offices, schools, maker spaces, public libraries, and other facilities, and people can use them for prototyping and small-scale or micromanufacturing without extensive technical training.

However, distributed and democratized manufacturing are still far from the goal of MDM (*8*–*10*), in which products are manufactured by a large, diverse, and geographically dispersed but coordinated network of individuals and organizations with agility and flexibility, but with near–mass-production quality, productivity, and cost effectiveness. For example, a company like Xometry would need to engage millions of users in micromanufacturing across the globe, similar to what companies like Uber and Lyft have achieved with transportation. The latent potential of MDM was evident during the early days of the COVID-19 pandemic, when personal pro-

Networked systems linking producers to consumers

Massively distributed manufacturing uses a cyber-physical operating system and artificial intelligence tools to connect and coordinate consumers with producers. Producers in micromanufacturing units can use three-dimensional (3D) printing to fabricate customized products. Smart logistics such as drones and rideshare services enable the physical product delivery.



GRAPHIC: C. BICKEL/*SCIENCE*

Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA. Email: okwudire@umich.edu

tective equipment (PPE) were in short supply. Mass production was too slow to react to the sudden demands for PPE, including demands for simple but vital plastic products like face shields. Worldwide, thousands, if not millions, of people, many of whom did not have experience in making these products, organized themselves into small networks to produce millions of face shields and other PPE using desktop 3D printers and other small-scale manufacturing equipment (*11*). This effort exposed key challenges of MDM in terms of standardizing production requirements, guaranteeing quality and reliability, and attaining high production efficiencies that can rival those of mass production.

This example illustrates the important role of technology in enabling MDM.

"Mass production…lacks flexibility, agility, and resilience and cannot readily provide consumers with personalized products…"

The First, Second, and Third Industrial Revolutions, driven by mechanization, electrification plus assembly lines, and digital computing, respectively, paved the way for the Fourth Industrial Revolution (or Industry 4.0), undergirded by networked cyber-physical systems and artificial intelligence. For example, Xometry leverages cloud computing and machine learning to power its instant quoting engine that enables customers to receive pricing, expected lead times, and manufacturability feedback within seconds. Similarly, 3Diligent uses cloud computing to enable manufacturers in its network to route jobs across their shop floors and track quality.

With advances in Industry 4.0, manufacturing machines (including low-cost 3D printers) are increasingly equipped with sensors and cloud connectivity (*12*). The large amounts of data generated by these sensors are being used in machine-learning algorithms to provide predictive and corrective actions (*13*). Advanced cloud-based controllers are being developed to improve the quality and productivity of the machines (*14*). These advances in technology and automation can converge into a cloud-based CPOS for MDM.

An inspiration for CPOSs is the central coordinator used in distributed computing to automate the allocation and execution of large-scale computing tasks on distributed networks of computers. The central coordinator has enabled Folding at Home (*15*), a distributed computing cluster that leverages the idle capacity of >100,000 personal computers to run simulations that help scientists to understand how proteins fold. Similarly, a CPOS will intelligently, efficiently, and securely coordinate large networks of cloud-connected, autonomous, and geographically-dispersed manufacturing resources. It will optimally allocate manufacturing jobs to the resources connected to it and leverage distributed and democratized delivery systems, such as shared vehicles and drones, for logistics (see the figure). It will apply machine learning to the data gathered from sensors to help assure and improve quality and to optimize operations. Furthermore, CPOSs will leverage the ingenuity of humans through the crowdsourcing of ideas to improve manufacturing operations across networks of manufacturers as well as cybersecurity measures to protect intellectual property and the privacy of participants. CPOSs will thus allow the collaboration of large, autonomous, heterogeneous, and geographically dispersed networks of manufacturers to rapidly respond to production demands and disruptions with agility and flexibility, while ensuring the high quality, productivity, and cost effectiveness of MDM. ■

REFERENCES AND NOTES

1. B. J. Pine II, *Mass Customization: The New Frontier in Business Competition* (Harvard Business School Press, 1993).
2. C. López-Gómez, E. O'Sullivan, M. Gregory, A. C. C. Fleury, L. Gomes, *Emerging Trends in Global Manufacturing Industries* (United Nations Industrial Development Organization, 2013).
3. B. Meyerson, *Top 10 Emerging Technologies of 2015* (World Economic Forum, 2015).
4. Foresight, *The Future of Manufacturing: A New Era of Opportunity and Challenge for the UK Summary Report* (The Government Office for Science, London, UK, 2013).
5. European Factories of the Future Research Association, *Factories of the Future: Multi-annual Roadmap for the Contractual PPP under Horizon 2020* (European Commission, 2013).
6. MForesight, *Democratizing Manufacturing: How to Realize the Promise of the Maker Movement* (2017).
7. Wohlers Associates, *Wohlers Report 2020: 3D Printing and Additive Manufacturing: Global State of the Industry* (Wohlers Associates, 2020); https://wohlersassociates.com/2020contents.htm.
8. J. S. Srai *et al.*, *Int. J. Prod. Res.* **54**, 6917 (2016).
9. H. Stewart, J. Tooze, *Making Futures* **4**, 1 (2015).
10. P. Jiang, J. Leng, K. Ding, P. Gu, Y. Koren, *Proc. Inst. Mech. Eng. B* **230**, 1961 (2016).
11. J. M. Pearce, *J. Manuf. Mater. Process.* **4**, 49 (2020).
12. C. E. Okwudire, S. Huggi, S. Supe, C. Huang, B. Zeng, *Inventions* **3**, 56 (2018).
13. T. Wuest, D. Weimer, C. Irgens, K. D. Thoben, *Prod. Manuf. Res.* **4**, 23 (2016).
14. C. E. Okwudire, X. Lu, G. Kumaravelu, H. Madhyastha, *Robot. Comput.-Integr. Manuf.* **62**, 101880 (2020).
15. Folding at Home, https://foldingathome.org/.

ACKNOWLEDGMENTS

C.E.O. is a founder of Ulendo, which has licensed research in advanced cloud-based 3D printer control algorithms.

10.1126/science.abg4924

GENOMICS

Expanding cell-to-cell interactions

A new single-cell sequencing approach reveals astrocyte and microglia cross-talk during inflammation

By **Aymeric Silvin**[1] *and* **Florent Ginhoux**[2,3,4]

High-dimensional single-cell technologies have revolutionized the study of cell biology by unraveling the cellular heterogeneity and molecular complexity that underlies human health and disease. Most recently, these techniques have been used to answer intriguing questions about the cells of the central nervous system (CNS). For example, this approach was used to describe a protective type of microglia present in the brains of patients with Alzheimer's disease (*1*) and an astrocyte population that represses protective antioxidant and anti-inflammatory transcriptional programs in a neuroinflammatory mouse model and in multiple sclerosis (MS) patients (*2*). On page 360 of this issue, Clark *et al.* (*3*) take current single-cell technologies to the next level by dissecting cell-to-cell cross-talk in the CNS and beyond.

Our knowledge of the immune system in particular has been taken to unprecedented levels by the use of single-cell approaches, most notably single-cell RNA-sequencing, but these findings have generated their own questions and challenges. Among the most pressing is the need to precisely map the cross-talk between newly identified cell populations in vivo in real time and to understand their relative locations within a tissue in both homeostatic and disease states. Deciphering these key phenomena at the single-cell level will be critical to our understanding of molecular disease pathophysiology and will likely underpin the de-

[1]INSERM U1015, Gustave Roussy Cancer Campus, Villejuif 94800, France. [2]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore. [3]Shanghai Institute of Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [4]Translational Immunology Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore. Email: aymeric.silvin@gustaveroussy.fr; florent_ginhoux@immunol.a-star.edu.sg

velopment of the next generation of therapeutics for a wide range of conditions.

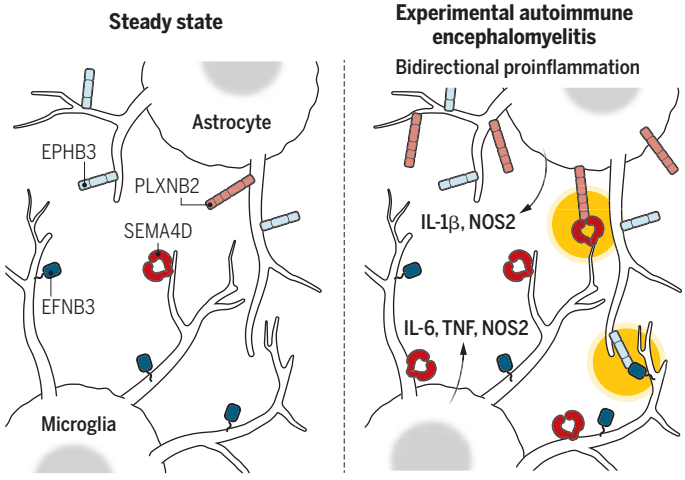
Traditionally, cell-to-cell interactions have been studied by using sorted bulk cell populations that were cultured together in vitro and then underwent bulk RNA-sequencing. The resulting gene expression data were then "mined" with available algorithms to predict possible interactions. However, using bulk cell populations often dismisses the sometimes critical contribution of minor subpopulations within the whole, and the in vitro environment excludes the importance of the physiological niche in which cells are intimately in contact. Additionally, spatial transcriptomics can now combine, for example, techniques such as the use of photoactivatable fluorescent reporters with single-cell RNA-sequencing (NICHE-seq) to infer the cellular and molecular composition of immunological niches (*4*). However, until now, no high-throughput technological pipeline that can comprehensively profile cell-to-cell interactions at the single-cell level within tissues has been available.

Understanding the intercellular interactions that shape and maintain the CNS is particularly challenging because these tissues are relatively inaccessible for experimental manipulations or difficult to model in vitro, and a small subset of cells within them may mediate highly potent effects. Thus, the cellular cross-talk that shapes the responses of specific cell types within the brain remains unknown, particularly during neuroinflammation, which is observed in many brain diseases. This knowledge gap drove the development of rabies barcode interaction detection followed by sequencing (RABID-seq) by Clark *et al.*, which combines molecular barcoding, viral tracing, and single-cell RNA-sequencing in vivo in an unbiased manner, expanding the toolkit for analysis of cell-to-cell interactions at the single-cell level. The authors investigated astrocyte-interacting cells, enabling the reconstruction of both interaction and transcriptional maps in vivo, all at single-cell resolution.

Microglia-astrocyte interactions are important during CNS development, homeostasis, and diseases such as MS (*5*, *6*), Alzheimer's disease (*7*), and prion disease (*8*). However, a comprehensive understanding of these interactions and how they shift during neuroinflammation is lacking.

Microglia-astrocyte cross-talk

Microglia-astrocyte cross-talk in EAE is revealed by RABID-seq. Microglia SEMA4D and EFNB3 interact with PLXNB2 and EPHB3 on astrocytes. Activation of SEMA4D-PLXNB2 and EFNB3-EPHB3 signaling induces the expression of proinflammatory factors (IL-6, TNF, IL-1β, NOS2) by both cell populations.



EAE, experimental autoimmune encephalomyelitis; EFNB3, ephrin-B3; EPHB3, ephrin type-B receptor 3; IL-6, interleukin-6; NOS2, nitric oxide synthase 2; PLXNB2, plexin B2; RABID-seq, rabies barcode interaction detection followed by sequencing; SEMA4D, semaphorin 4D; TNF, tumor necrosis factor.

Using the RABID-seq approach, Clark *et al.* compared the interactions of astrocytes in mice with induced experimental autoimmune encephalomyelitis (EAE), which may model human MS. The authors demonstrated the specific interaction between microglia and astrocytes but also detected astrocyte-astrocyte interactions, as well as interactions with microglia and other cells in the steady state and at the peak of EAE in affected mice. In the context of EAE, the authors uncovered astrocyte interaction networks that were more diverse than anticipated and included interactions with peripheral cells recruited to the inflamed CNS such as T cells, as well as with known interaction partners such as microglia.

This study revealed that microglia-astrocyte cross-talk was modified during EAE, becoming a bidirectional interaction that induced a local proinflammatory environment. RABID-seq allowed the authors to track specifically those microglia that were in close contact with astrocytes, enabling the identification of substantial differences in gene expression between microglia and astrocytes from naïve versus EAE mice, which revealed important details of the pathways implicated in their interactions. Further pathway and gene expression analyses suggested two main axes of interactions that promoted CNS inflammation in EAE: microglia-astrocyte semaphorin 4D (SEMA4D)–plexin B2 (PLXNB2) signaling and microglial Ephrin-B3 (EFNB3) and astrocyte ephrin type-B receptor 3 (EPHB3) (see the figure), pathways already implicated in the process of inflammation in various diseases.

The authors validated the roles of this receptor signaling in the process of CNS inflammation during EAE. Inactivation of the receptor *Sema4d* in microglia or its ligand *Plxnb2* in astrocytes resulted in EAE amelioration, whereas ablation of *Ephb3* expression in astrocytes and *Efnb3* in microglia led to reduced recruitment of proinflammatory monocytes to the CNS during EAE. These results demonstrate that EPHB3 receptor signaling boosts the proinflammatory activities of astrocytes, whereas EFNB3 signaling boosts nuclear factor–κB (NF-κB)–driven responses in microglia during EAE. Thus, EPHB3 signaling in astrocytes promotes CNS inflammation and is a candidate target for therapeutic intervention.

Looking to MS, the authors analyzed a publicly available RNA-sequencing dataset of brain samples from MS patients and matched controls and found evidence of relatively increased interactions between microglia and astrocytes through the SEMA4D-PLXNB2 axis in patients. Moreover, immunolabeling revealed higher expression of SEMA4D in microglia and PLXNB2 in astrocytes of MS patients, underlining the relevance of this work for human MS.

Together, the study of Clark *et al.* presents a new approach (RABID-seq) that allows scientists to recreate small networks of in vivo single cell–to–single cell interactions more accurately and efficiently than before. Although the technique currently does not formally incorporate spatial information, it could be modified to do so using "seqFISH," a technique that uses fluorescent RNA probes, allowing visualization of cell localization in a tissue expressing a particular RNA (*9*). In addition, although the assessment of cell interactions itself is highly time efficient once the model is established, at present each research team wishing to use the approach will need to generate the appropriate mouse models. ■

REFERENCES AND NOTES

1. H. Keren-Shaul *et al.*, *Cell* **169**, 1276 (2017).
2. M. A. Wheeler *et al.*, *Nature* **578**, 593 (2020).
3. I. C. Clark *et al.*, *Science* **372**, eabf1230 (2021).
4. C. Medaglia *et al.*, *Science* **358**, 1622 (2017).
5. V. Rothhammer *et al.*, *Nature* **557**, 724 (2018).
6. S. A. Liddelow *et al.*, *Nature* **541**, 481 (2017).
7. A. U. Joshi *et al.*, *Nat. Neurosci.* **22**, 1635 (2019).
8. K. Hartmann *et al.*, *Acta Neuropathol. Commun.* **7**, 83 (2019).
9. C. L. Eng *et al.*, *Nature* **568**, 235 (2019).

10.1126/science.abh2810

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

WATER RESOURCES

The hidden crisis beneath our feet

Disappearing groundwater requires action to prevent widespread water scarcity

By **James S. Famiglietti**[1,2] *and* **Grant Ferguson**[1,2,3,4]

Most of the world's unfrozen freshwater is invisible to humanity. Ninety-six percent of it (*1*) is stored beneath the land surface as groundwater in soil and rock layers called aquifers. However, groundwater's unobtrusive nature belies its critical importance to global water and food security while simultaneously subjecting it to massive overexploitation. Groundwater is the primary water source for billions of people and for nearly half of irrigated agriculture, yet its inconspicuous presence has allowed groundwater to elude effective governance and management in countless regions around the world (*2*, *3*). Consequently, more than half of the world's major aquifers are being depleted, some of them at an alarming pace (*4*). On page 418 of this issue, Jasechko and Perrone (*5*) show that millions of the wells that are used to pump the disappearing groundwater are at risk of running dry.

Jasechko and Perrone assembled a dataset of nearly 39 million groundwater wells from 40 different countries and territories. The dataset, which includes the locations, depths, purposes, and construction dates of the wells, allowed the authors to reach two main conclusions. They found that up to 20% of the wells they analyzed are at risk of running dry because of long-term groundwater decline, seasonal variation in water levels, or both. In addition, they identified that new wells tend to be drilled to greater depths. However, the authors note that the deeper wells are not necessarily located in regions of substantial groundwater depletion (*6*), with the implication that many of the new wells are just as likely to run dry as the old wells.

Jasechko and Perrone implicitly deliver a timely warning that universal access to groundwater is fundamentally at risk. As groundwater levels decline around the world, only the relatively wealthy will be able to afford the cost of drilling deeper wells and paying for the additional power required to pump groundwater from greater depths. Lower-income families, poorer communities, and smaller businesses, including smaller farms, will experience progressively more limited access in the many regions around the world where groundwater levels are in decline. This scenario is already playing out, for example, in California's heavily agricultural Central Valley, where deeper wells are drawing down groundwater levels such that farm workers' domestic wells are running dry. Without intervention, the gap between the water "haves" and "have nots" (*3*) will only widen further.

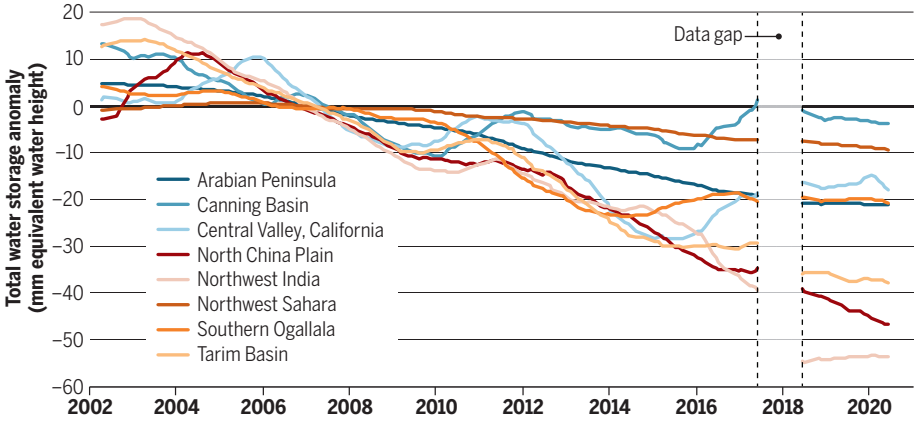
Beyond dwindling access to groundwater, the consequences of millions of wells running dry, and perhaps millions more in the decades to come, would be severe and unparalleled at such a scale in human history. They include major threats to food production, the health and livelihoods of the millions to billions of people affected, and the environment. Disappearing groundwater resources may act as a trigger for violent conflicts and have the potential to generate waves of climate refugees. Avoiding such a scenario is clearly paramount to human security. This requires considerable discovery-driven and engaged research, combined with broad and inclusive stakeholder engagement, water diplomacy, and advocacy to federal governments and international bodies.

Maintaining water levels above the intake portions, called well screens, is a critical measure of whether groundwater is being developed sustainably but often has been ignored in favor of recharge rates and residence times (*7*). Predicting changes in water levels remains a key challenge in hydrology owing to the difficulty in characterizing the hydraulic properties of the subsurface (*8*). Breakthroughs in characterization and monitoring of groundwater systems and new modeling approaches are required to understand where and when wells might go dry.

Satellites like NASA's Gravity Recovery and Climate Experiment (GRACE) mission can identify large-scale (>150,000 km^2) trends in groundwater storage changes (*4*–*6*), which are important for understanding aquifer-scale behavior and identifying excessive rates of decline (see the figure). However, satellites cannot peer beneath Earth's surface to measure water-table heights and fluctuations. National and international collaborative efforts are required to systematically explore and characterize Earth's hydrogeology (*3*), comprehensively monitor groundwater lev-

Watching groundwater depletion from the sky

Changes in total water storage are shown for regions within several mid-latitude aquifer systems from the NASA Gravity Recovery and Climate Experiment (GRACE) and follow-on (GRACE-FO) satellite missions. Vertical dashed lines indicate the data gap between the GRACE and GRACE-FO missions. Monthly storage changes are reported as anomalies for the period from March 2002 to June 2020 with 24-month smoothing.



GRAPHIC: DAVID FERRIS, UNIVERSITY OF SASKATCHEWAN. ADAPTED BY N. DESAI/*SCIENCE*

[1]Global Institute for Water Security, University of Saskatchewan, Saskatoon, SK S7N 3H5, Canada. [2]School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK S7N 5C8, Canada. [3]Department of Civil, Geological, and Environmental Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada. [4]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ 85721-0011, USA. Email: jay.famiglietti@usask.ca

els and groundwater quality (*9*), develop and test large-scale groundwater models and coupled land surface-groundwater models (*10*), and openly share the data (*9*).

To ensure that groundwater remains a reliable component of water supplies, several other formidable challenges must be addressed. Inclusive governance (*2*, *11*), effective management (*12*), and an agreed-upon definition of groundwater sustainability (*7*) are all essential elements of what must be a multifaceted strategy. Few aquifer systems around the world are managed within such an ideal framework, and the capacity to do so does not typically exist.

New institutions and networks must be fostered to raise awareness of these urgent needs, encourage and coordinate stakeholder participation, and help governments build the political will to protect groundwater as a key element of water security. Because most large groundwater systems are regional and transboundary, regional collaboration is essential, as is a water diplomacy strategy that treats groundwater as a vehicle for cooperation rather than conflict. A global coordinating body may well be required to share knowledge, tools, and best practices across regions and raise the profile of groundwater within the United Nations Sustainable Development Goals (SDGs), in particular, in SDG6, ensuring the availability and sustainable management of water and sanitation for all.

Jasechko and Perrone have helped make the invisible visible. The authors have demonstrated that millions of groundwater wells are at risk of running dry, illuminating that climate resilience is at considerable risk. Digging deeper wells is a supply-side solution that only exacerbates the problems described. The time is now for key research and exploration and for science-informed governance and policy that address the demand for groundwater and eliminate its overexploitation. ■

REFERENCES AND NOTES

1. I. Shiklomanov, in *Water in Crisis: A Guide to the World's Fresh Water Resources*, P. H. Gleick, Ed. (Oxford Univ. Press, 1993).
2. K. G, Villholth, K. I. Conti, in *Advances in Groundwater Governance*, K. G. Villhoth *et al*., Eds. (CRC Press, 2018).
3. J. S. Famiglietti, *Nat. Clim. Chang.* **4**, 945 (2014).
4. A. S. Richey *et al*., *Water Resour. Res.* **51**, 5127 (2015).
5. S. Jasechko, D. Perrone, *Science* **372**, 418 (2021).
6. M. Rodell *et al*., *Nature* **557**, 651 (2018).
7. T. Gleeson, M. Cuthbert, G. Ferguson, D. Perrone, *Annu. Rev. Earth Planet. Sci.* **48**, 431 (2020).
8. S. M. Gorelick, C. Zheng, *Water Resour. Res.* **51**, 3031 (2015).
9. International Groundwater Resources Assessment Centre; www.un-igrac.org/.
10. L. E. Condon, R. M. Maxwell, *Sci. Adv.* **5**, eaav4574 (2019).
11. J. Hoogesteger, P. Wester, *Environ. Sci. Policy* **51**, 117 (2015).
12. J. J. Butler Jr., D. O. Whittemore, B. B. Wilson, G. C. Bohling, *Geophys. Res. Lett.* **43**, 2004 (2016).

10.1126/science.abh2867

NEUROSCIENCE

A spotlight on the elusive striatal cholinergic interneuron

A new role for the ubiquitous integrated stress response pathway in striatal-dependent learning and memory

By **Anna E. Ingebretson** *and* **Julia C. Lemos**

The brain of the mouse, one of the most commonly used animal models in biomedical research, is composed of about 70 million neurons all working in concert to control behavioral outputs and promote survival. Yet, on page 361 of this issue, Helseth *et al*. (*1*) demonstrate that just one signaling pathway in striatal cholinergic interneurons, a cell type that makes up ~1% of neurons in a brain region of ~120,000 cells, has a profound regulatory role in shaping habitual learning and memory, a type of learning and memory that allows us to drive to work or get to the cereal aisle at the grocery store without even thinking about it.

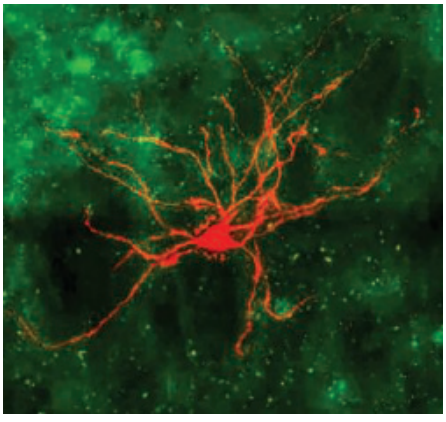
The dorsal striatum is a central brain region in a constellation of nuclei known as the basal ganglia. The basal ganglia plays an essential role in movement, action planning, reward learning, and goal-directed and habitual behaviors, as well as procedural learning (*2, 3*). Furthermore, several brain diseases are linked to basal ganglia dysfunction, including Parkinson's disease, Huntington's disease, depression, substance use disorders, Tourette's syndrome, and obsessive-compulsive disorder (*3, 4*). The majority of neurons in the dorsal striatum are inhibitory medium spiny projection neurons that are characterized by spiny protrusions where inputs from other neurons make contact (*5*). Medium spiny neurons have been a major focal point within the basal ganglia field because they make up ~95% of the neurons in the striatum. By contrast, cholinergic interneurons comprise only 1 to 2% of all neurons in the region (*6*) and as such have historically been more elusive to study. Cholinergic interneurons are defined by their large non-spiny morphology, pacemaker-like action potential firing, and production of the neurotransmitter acetylcholine (*6*). Despite constituting only 1 to 2% of neurons in the striatum, their many axonal divisions and wide branching structures allow a single neuron to supply a large number of neighboring neurons with acetylcholine (*6*). Accordingly, cholinergic interneurons have powerful, wide-reaching effects that allow them to act as master regulators of striatal output.

Cholinergic neurotransmission is necessary for normal striatal function; however, owing to the relative scarcity of this population, it has been difficult to identify the precise mechanisms through which cholinergic interneurons regulate other neuronal populations. New cutting-edge neuroscience tools have allowed researchers to identify and experimentally manipulate these cells. For example, recent studies determined that the pattern of cholinergic interneuron firing—not simply the overall rate—plays an important role in behavior and cognition. Striatal cholinergic interneurons display two different firing modes: a continuous, pacemaker-like firing pattern and an intermittent firing pattern in which a burst of action potentials is followed by a pause in firing (*7*). Behavioral studies showed that pauses in cholinergic firing occur at motivationally important moments. Pauses are critical for making associations between cues in the environment and specific outcomes in classical conditioning paradigms (*8*) and are also necessary for updating behavioral strategies in the face of changes in previously learned associations (*9*). Although several different cellular and circuit mechanisms can generate pauses in firing, one well-established mechanism is activation of inhibitory dopamine type 2 (D2) receptors located on cholinergic interneurons (*10, 11*). Dopaminergic and cholinergic systems are reciprocally connected within the striatum: Cholinergic transmission can both trigger and modulate dopamine release (*12, 13*), and dopamine released from terminals emanating from the midbrain into the dorsal striatum activates D2 receptors on cholinergic interneurons, which briefly inhibits firing (*11*).

Helseth *et al*. demonstrate that cholinergic regulation of dopamine signaling in the dorsal striatum relies on the integrated stress response (ISR) pathway, a ubiquitous cell signaling pathway that is transiently engaged in response to different homeostatic challenges (*14*). Because the ISR pathway

Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, USA. Email: jlemos@umn.edu

regulates protein synthesis and is therefore a major contributor to long-term synaptic plasticity, learning, and memory, it has emerged as a central focus of pharmacotherapeutic development. Phosphorylation of eukaryotic initiation factor 2 alpha (eIF2α), a factor required for initiating protein translation, is required for activating the ISR pathway (*14*). To track how this pathway is engaged, Helseth *et al.* generated a new tool called selective phospho-eIF2α ORF (open reading frame) Tracking light (SPOTlight) that uses fluorophores to report when phosphorylated eIF2a expression is "on" or "off" (*1*). Using this tool,



The integrated stress response pathway is constitutively activated in a mouse striatal cholinergic interneuron (red) to support learning and memory.

they made a surprising discovery: Despite its canonically transient expression, the ISR pathway was constitutively active in cholinergic interneurons. Dampening the firing activity of cholinergic interneurons reduced ISR pathway activation, indicating that the basal activity of these neurons promotes constitutive activation of the ISR pathway. Notably, disruption of the ISR pathway switched the function of D2 receptors on cholinergic interneurons from inhibitory to excitatory, fundamentally altering a key component of the "pause." This same manipulation also disrupted D2 regulation of dopamine release. Consistent with the behavioral role of striatal cholinergic interneurons, disruption of the ISR pathway in cholinergic interneurons disrupted skill and habitual learning assayed by a Morris water maze and operant conditioning tasks, but left motor coordination intact. Intriguingly, the authors found that increased performance vigor, defined as the amount of energy and speed used in a task, influenced the behavioral results. Because dopamine signaling contributes to performance vigor, these findings suggest that blocking the ISR pathway in cholinergic interneurons affects local dopamine dynamics to shape striatal neuron output to brain regions supporting motor and cognitive functions.

In addition to its substantial technical innovation, the study of Helseth *et al.* puts forth an entirely new and unrecognized role for the ISR pathway that not only is important for cholinergic interneuron activity but also supports D2 receptor–mediated inhibition of cholinergic interneurons. D2 receptor–mediated inhibition of cholinergic interneurons alters excitatory transmission at medium spiny neuron synapses, ultimately affecting the net output of the striatum (*15*). It is rare to find evidence that basal activation of a signaling pathway, such as the ISR pathway, not only is critical to receptor function, but without it, the functional sign of the receptor is switched. Another surprising finding is that some of the behavioral consequences of ISR pathway disruption were only apparent in male mice.

The findings of Helseth *et al.* reveal new topics for exploration in basal ganglia function. It is not yet clear what mechanisms drive constitutive activation of the ISR pathway in this sparse population of cells. Another question is how disruption of this pathway modifies D2 receptor function within striatal circuitry. Probing these mechanisms could reveal how firing patterns of cholinergic interneurons encode conditioned associations that support learning. In addition, the sexually divergent behavioral consequences of ISR pathway disruption in cholinergic interneurons highlight avenues for investigating key sex differences. These findings could have profound consequences for the therapeutic utility of treatments that target the ISR pathway. ■

REFERENCES AND NOTES

1. A. R. Helseth *et al.*, *Science* **372**, eabe1931 (2021).
2. J. Cox, I. B. Witten, *Nat. Rev. Neurosci.* **20**, 482 (2019).
3. D. M. Lipton, B. J. Gonzales, A. Citri, *Front. Neurosci.* **13**, 1 (2019).
4. A. B. Nelson, A. C. Kreitzer, *Nat. Rev. Neurosci.* **37**, 117 (2014).
5. J. M. Tepper, E. D. Abercrombie, J. P. Bolam, Eds., *GABA and the Basal Ganglia: From Molecules to Systems* (Elsevier Science, 2007), vol. 160, pp. 3–7.
6. K. K. Gonzales, Y. Smith, *Ann. N. Y. Acad. Sci.* **1349**, 1 (2015).
7. Y. F. Zhang, S. J. Cragg, *Front. Syst. Neurosci.* **11**, 80 (2017).
8. P. Apicella, *Neuroscience* **360**, 81 (2017).
9. V. F. Prado, H. Janickova, M. A. Al-Onaizi, M. A. Prado, *Neuroscience* **345**, 130 (2017).
10. J. B. Ding, J. N. Guzman, J. D. Peterson, J. A. Goldberg, D. J. Surmeier, *Neuron* **67**, 294 (2010).
11. N. Chuhma, S. Mingote, H. Moore, S. Rayport, *Neuron* **81**, 901 (2014).
12. S. Threlfell *et al.*, *Neuron* **75**, 58 (2012).
13. D. B. Lester, T. D. Rogers, C. D. Blaha, *CNS Neurosci. Ther.* **16**, 137 (2010).
14. K. Pakos-Zebrucka *et al.*, *EMBO Rep.* **17**, 1374 (2016).
15. S. M. Augustin, J. H. Chancey, D. M. Lovinger, *Cell Rep.* **24**, 2883 (2018).

ACKNOWLEDGMENTS

The authors are supported by US National Institute of Mental Health grants MH R00109627 and MH R01122749.

10.1126/science.abi4907

TISSUE REGENERATION

Healing without scarring

A gene regulatory circuit directs tissue regeneration and limits fibrosis

By **Piotr Konieczny**[1,2,3] *and* **Shruti Naik**[1,2,3]

Rapid repair and restoration of tissue integrity after injury is vital for organismal survival. Some lower vertebrates and invertebrates possess the extraordinary ability to regenerate functional tissues throughout their lifetime (*1*). In mammals, however, tissue regeneration is largely restricted to gestation. In the terrestrial environment, evolutionary pressures to heal have favored the rapid formation of fibrotic tissue (or scar tissue) instead of recreating original tissue architecture (*2*). Scarring-related tissue dysfunction remains a major health care challenge with limited treatment options. Thus, scar prevention and regeneration of functional tissues is a coveted goal of tissue repair. On page 362 of this issue, Mascharak *et al.* (*3*) point to a therapeutic target for tissue regeneration by revealing how specific fibroblasts produce scars in response to skin wounding.

A subset of mammalian dermal fibroblasts that express the transcription factor Engrailed-1 (En-1) was previously identified as responsible for scar deposition, raising the tantalizing possibility that modulating this lineage of En-1–positive fibroblasts (EPFs) may limit scarring (*4*). Mascharak *et al.* now report the presence of subsets of "profibrotic" EPFs and "proregenerative" fibroblasts that do not express En-1 [En-1–negative fibroblasts (ENFs)] in adult murine skin wounds. They further identified mechanotransduction—that is, sensing and converting mechanical stimuli into biochemical signals—through the transcription factor Yes-associated protein (YAP) as a key signal involved in the conversion of wound ENFs to EPFs. Genetic and pharmacological inhibition of YAP was sufficient to block EPF and scarring in mice. Notably, abrogating the EPF scar response led to ENF-mediated restoration of skin appendages, tensile strength,

[1]Department of Pathology, New York University School of Medicine, New York, NY 10016, USA. [2]Department of Medicine, New York University School of Medicine, New York, NY 10016, USA. [3]Ronald O. Perelman Department of Dermatology, New York University School of Medicine, New York, NY 10016, USA. Email: shruti.naik@nyulangone.org

PHOTO: CHARLES R. GERFEN/NIMH

and extracellular matrix architecture that matches that of unwounded skin (see the figure).

The discovery of diverse fibroblast populations across different organ systems and between anatomical sites within the same organ has underscored highly specialized roles for these cells in health and disease. For instance, functionally distinct populations of fibroblasts reside in different layers of the dermis and contribute to skin development and wound responses (*5*). Although embryonic and wounded skin are replete with scar-forming EPFs, healthy postnatal skin conspicuously lacks these cells. Then, an open question is where scar-forming postnatal wound EPFs come from. Are they generated by a resurgence of embryonic EPFs or by the direct conversion of adult ENFs to EPFs in the wound bed? To address this, Mascharak *et al.* fractionated ENFs from the mouse upper or papillary dermis, middle or reticular dermis, and lower or hypodermis and transplanted these populations into a secondary wound environment in genetically matched host animals. Consistent with previous studies highlighting the importance of a specific subset of reticular fibroblasts in fibrosis (*5*), the authors observed that ENFs from the reticular dermis preferentially up-regulated *En1* expression and converted to EPFs after transplantation. By contrast, papillary dermal ENFs failed to generate EPFs, and ENFs demonstrated a propensity toward the adipocyte lineage. How this potential is imposed on reticular ENFs by developmental and/or microenvironmental cues awaits investigation.

Fibroblasts are particularly sensitive to mechanical cues, and increased wound tension triggers profibrotic responses (*5*). Mascharak *et al.* examined if and how mechanical forces promote En-1 expression in ENFs by culturing ENFs on plastics of varying stiffness. Reticular ENFs converted to EPFs on high-stiffness plastic but failed to do so on soft hydrogel. This ENF-to-EPF transition was blocked by inhibiting Rho kinase, a canonical mediator of mechanotransduction. When mouse skin was artificially stretched, both unwounded and wounded skin had increased expression of fibroblast YAP and alpha smooth muscle actin (α-SMA), two key indicators of the fibrotic response (*6*). Moreover, YAP, a mechanosensitive transcription factor, showed increased nuclear localization in ENFs but not EPFs. Pharmacological inhibition of YAP in mechanically loaded wounds limited the generation of EPFs, suggesting that YAP controls the switch to a "profibrotic" EPF fate. A transcriptomics time course of reticular ENFs revealed an enrichment of integrin signaling and collagen-related pathways, which were reversed by YAP inhibition. Whether En-1 is a direct transcriptional target of YAP or is circuitously activated by other factors is yet to be determined. Additionally, En-1 is a homeobox transcription factor with a central role in midbrain development (*7*). Untangling the complex gene regulatory circuits involving YAP and En-1 that specify EPF fate and function warrants further study.

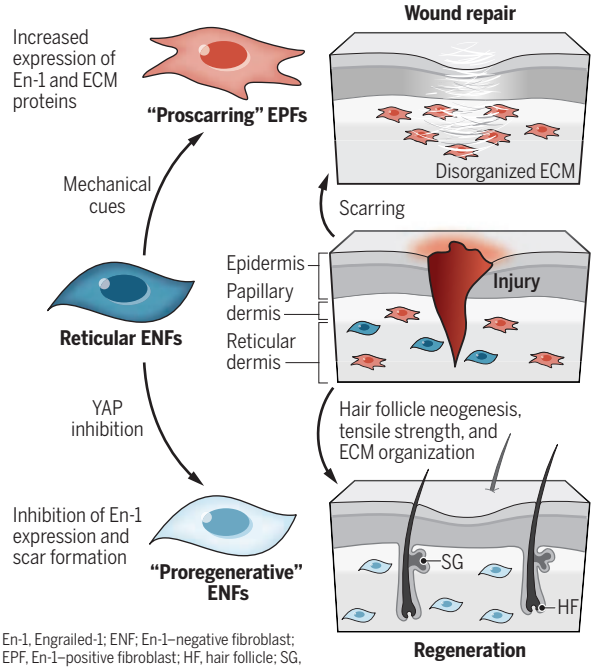
To formally confirm that EPFs were responsible for scar formation, Mascharak *et al.* genetically ablated these cells from wounds. The absence of EPFs not only limited scar formation but also unexpectedly resulted in the regeneration of hair follicles and appendages. Both chemical inhibition and genetic deletion of YAP in fibroblasts also limited scarring and recapitulated the regenerative phenotype observed after genetic ablation of EPFs. Notably, an unbiased machine-learning analysis of tissue ultrastructure revealed that the organization of the extracellular matrix after YAP inhibition more closely resembled that of unwounded skin. Previous attempts to modulate tension reduced scarring but failed to regenerate tissue (*8*). Thus, simply reducing collagen (a constituent of the extracellular matrix) deposition by fibroblasts may not be sufficient to promote regeneration, which likely requires complex remodeling of tissue architecture and other secondary cues driven by loss of YAP. YAP mechanosignaling underlies the pathology of liver cirrhosis and idiopathic pulmonary fibrosis that compromise organ function (*6*, *9*). Similarly targeting profibrotic fibroblast populations could be a therapeutic strategy to limit scarring and restore tissue function in a range of fibrotic diseases.

In the study of Mascharak *et al.*, ENF-mediated skin regeneration (as the result of YAP inhibition) is defined by three key criteria: regeneration of epidermal appendages (hair follicles and glands), reestablishment of extracellular matrix architecture identical to that in healthy skin, and restoration of tensile strength to healthy skin levels. Perhaps regeneration is a default state that is outpaced by scarring in mammalian repair. If so, what are the molecular mediators of regeneration, and do distinct cues govern each facet of regeneration, for example, hair follicle neogenesis versus tensile strength? Previous studies have attributed the sparse de novo regeneration of hair follicles in large wounds to the resurgence of developmental signals (*10*). Whether similar mechanisms are at play or whether modulating mechanotransduction engages distinct pathways that are tailored to adult regeneration begets further investigation.

Tissue scarring is a leading cause of morbidity and mortality worldwide (*11*). The findings by Mascharak *et al.* thus hold great promise not only for antifibrotic therapies but also for the simultaneous activation of the skin's regenerative properties. ■

Specific fibroblasts orchestrate tissue regeneration in adult mammalian skin

Classic wound healing has favored the rapid formation of fibrotic tissue or scar tissue. Blocking mechanotransduction by YAP inhibition prevents the transition of proregenerative fibroblasts (ENFs) to proscarring fibroblasts (EPFs) and simultaneously promotes tissue regeneration, hair follicle neogenesis, tensile strength, and extracellular matrix (ECM) organization.



En-1, Engrailed-1; ENF; En-1–negative fibroblast; EPF, En-1–positive fibroblast; HF, hair follicle; SG, sebaceous gland; YAP, Yes-associated protein.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. G. C. Gurtner, S. Werner, Y. Barrandon, M. T. Longaker, *Nature* **453**, 314 (2008).
2. S. A. Eming, P. Martin, M. Tomic-Canic, *Sci. Transl. Med.* **6**, 265sr266 (2014).
3. S. Mascharak *et al.*, *Science* **372**, eaba2374 (2021).
4. Y. Rinkevich *et al.*, *Science* **348**, aaa2151 (2015).
5. R. R. Driskell *et al.*, *Nature* **504**, 277 (2013).
6. J. J. Gokey *et al.*, *JCI Insight* **3**, e122490 (2018).
7. P. S. Danielian, A. P. McMahon, *Nature* **383**, 332 (1996).
8. H. I. Harn *et al.*, *Exp. Dermatol.* **28**, 464 (2019).
9. S. Alsamman *et al.*, *Sci. Transl. Med.* **12**, eaay8798 (2020).
10. M. Ito *et al.*, *Nature* **447**, 316 (2007).
11. A. Bayat, D. A. McGrouther, M. W. Ferguson, *BMJ* **326**, 88 (2003).

10.1126/science.abi5770

VIEWPOINT: COVID-19

Addressing racial inequities in medicine

The COVID-19 pandemic highlights the need to dismantle structural racism that perpetuates health inequities

By **Malika A. Fair**[1,2] *and* **Sherese B. Johnson**[1]

COVID-19 inequitably affects marginalized racial and ethnic populations across the world. A review of more than 17 million adult patients in the United Kingdom revealed a nearly twofold risk of death from COVID-19 among Black and Asian populations compared with white populations (*1*). Black people comprise 12.5% of the US population, yet they account for more than 18% of COVID-19 associated deaths (*2*). Although Black and Latinx populations in the US experience higher rates of infection, hospitalization, and deaths compared with white populations, they have similar case fatality rates, suggesting that there is no innate vulnerability or susceptibility to COVID-19 (*3*). Persistent COVID-19 racial and ethnic inequities are likely caused by structural racism that results in an increased risk of exposure and inadequate health care access in communities of color (*3*).

Structural racism, or the discriminatory policies, practices, and systems that reinforce an unequal distribution of power and resources in social institutions, is a considerable driver of health inequities (*4*). The COVID-19 pandemic has exacerbated the socioeconomic disadvantages that have led to an overrepresentation of marginalized groups in service-industry jobs and excessive financial insecurities that affect how people live and their ability to access health care (*4*). Communities of color are more likely to reside in shared and congregate housing, use public transportation for commuting, and support themselves and their families with low-income jobs. These factors increase COVID-19 exposure, because of the likelihood for more person-to-person contact, and limit the ability to self-isolate. Additionally, these social conditions increase a person's risk for chronic conditions, including hypertension, diabetes, cardiovascular disease, and obesity, that are associated with more severe COVID-19 outcomes (*5*).

[1]Association of American Medical Colleges, Washington, DC, USA. [2]Department of Emergency Medicine, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA.
Email: mfair@aamc.org; sjohnson@aamc.org

A broad, multisector community-engaged response is required to address the social determinants of health and achieve health equity with COVID-19 and other diseases. The medical and scientific community must be a willing partner in dismantling structural racism in society, which includes an emphasis on addressing the socioeconomic, environmental, and behavioral factors that influence most health outcomes. Although these efforts are critical, there is also an urgent need and opportunity to dismantle structural racism within the traditional functions of the medical and scientific community (see the figure). This necessitates several actions in clinical practice, medical education, and research: acknowledging that race (group categorizations based on physical traits such as skin color) and ethnicity (groups defined by shared language, history, religion, and culture) are social and political constructs and are a poor proxy for ancestry (inherited genetic variations traced to geographical origins of a person's ancestors); uprooting the racism that is deeply embedded in perspectives, policies, and practices in health systems; and co-creating a new system that incorporates the voices of marginalized populations who are unjustly suffering with unmet social needs and adverse health outcomes related to COVID-19 and other diseases.

"...the medical and scientific communities must make bold changes for the world to overcome two public health crises—racism and COVID-19."

Quality improvement activities are a core element of health care, but an equity lens has not routinely been included (*6*). Annual data from US hospitals and health systems show that substantial inequities remain in the quality of care between racial and ethnic groups (*7*). Marginalized racial and ethnic groups in the US are less likely to receive clinically necessary and routine treatments for several diseases, including cancer and cardiovascular disease, as well as mental health disorders, compared with their white counterparts (*7*). To begin effectively operationalizing an equity lens in quality improvement, clinics, hospitals, and health systems should collect and stratify data by self-identified racial and ethnic categories and make these data transparent to leaders and clinicians. This will allow previously unnoticed differences in quality of care, due to implicit or explicit bias, to be addressed with system-level changes that facilitate quality improvement.

Providing quality, equitable care also requires a reexamination of how clinical algorithms are used in medicine. The use of race in clinical decision support tools that provide guidance on medications, treatments, or procedures can delay access to care and further exacerbate health inequities (*8*). For example, in pediatrics, a calculator has been developed to predict the likelihood of a urinary tract infection (UTI). If the child is Black, the calculator predicts a lower risk of UTI, which may delay diagnostic tests and treatments for Black patients (*8*). This tool provides neither an evidence-based biological nor social explanation for why race has been added. Additionally, some hospitals have removed race from the calculation of kidney function estimates, or glomerular filtration rate (GFR). In a recent study, removing race from the GFR could allow up to a third of Black patients to be reclassified to a more severe stage of kidney disease and prompt appropriate referrals for dialysis and transplant (*9*). There is evidence of variation in the GFR within racial categories, suggesting differences based on ancestry and not race (*10*). Within these algorithms, clinicians and scientists should consider opportunities to include more specificity, such as country of origin, an explanation of the genetic association from a geographic region, and the social factors linked to identifying as a particular race in various regions around the world. Without this context, the sole inclusion of race serves to reinforce the false narrative of genetic or biological racial differences and disregard the socioeconomic conditions that largely influence health outcomes.

Beliefs about biological differences between races are often explicitly or implicitly embedded in medical education (*11*). Some medical students and residents believe that Black people are not as sensitive to pain or are more prone to blood clots (*12*). Other physicians incorrectly associate diseases with race instead of ancestry. For example, it is common to hear an association between sickle cell disease and Black people versus examining it as an ancestrally associated disease common in populations at risk for

malaria (*11*). Challenging beliefs and assumptions about biological differences requires increased education about the implications of racism and racial categorizations. White Coats for Black Lives, a US medical student group, developed a racial justice report card to evaluate curricula, policies, and practices to dismantle racism in medical school. Their evaluation includes an analysis of the pedagogy on race as a social and political construct, rather than solely based in science (*13*). These students and other faculty champions across the US are helping to catalyze changes in medical school curricula.

Historically accurate and transformative education for learners, physicians, and scientists should be made available to counteract the subtle and unfounded teachings that have persisted along the medical education continuum. Medical and scientific professionals should have a critical understanding of the unjust social conditions that contribute to inequities as well as their root in structural racism and other systems of oppression. Historical injustices, both in society and in medical institutions, should be taught and opportunities provided to participate in reversing oppressive systems-level policies and practices. Foundational and continuing education should also be made available on population genetics, migration patterns, and expression of genes due to environmental exposures related to their impact on diseases in the global community. Evidence-based education that describes both the social impact of race and genetic ancestry will help the medical and scientific community identify the underlying critical factors that cause health inequities.

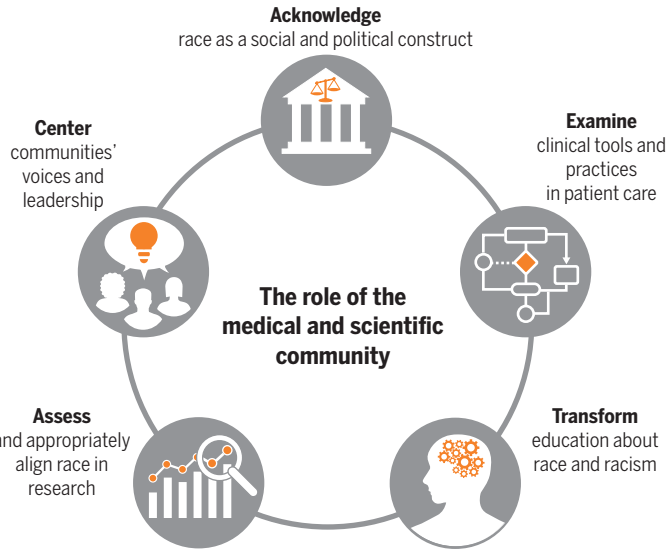
In research studies, race is often used as a variable with imprecise definitions and without questioning what the variables measure (*14*). For example, a research study comparing Black people to other racial and ethnic groups could include three people who self-identify as "Black" with varying backgrounds: someone who is biracial with a white parent while living in the US, someone who is multiracial with an Asian parent living in the Caribbean, or someone with two parents from and living in West Africa. Depending on their individual decisions and the communities they identify with, those three people may change which racial category they identify with over time. Grouping the individuals into the same analysis that studies a "Black" population may overlook extensive genetic variation between them. Research questions, analyses, and conclusions made solely on race should account for the complex and imprecise socially constructed categories of race (*14*). Analyses are incomplete if they do not also reflect the impact of unfair social advantage and disadvantage experienced by racial and ethnic groups worldwide. Several studies show an association between perceived racism and poor health indicators such as high blood pressure, coronary artery calcification, and breast cancer incidence (*15*). Research investigations should include defining the racial and ethnic categories and clarifying the impact of both social factors on what is being studied and whether conclusions should be based on ancestry instead of race. There is value in continuing to include race in research to elucidate root causes and helpful interventions to address health inequities if more specificity and rigor are added to what race is serving as a proxy for in the study. For this research to have greater impact and advance health equity, efforts should be made to diversify the biomedical and scientific workforce, include multidisciplinary scholars in study designs, and increase participation of marginalized populations in research studies.

One of the most important lessons from the COVID-19 pandemic is that the medical community waited too long to become trustworthy partners with marginalized communities. The systems were not in place to hear their concerns, respond quickly in areas that were disproportionately affected by COVID-19 cases, or promptly build vaccine confidence. To address both the health inequities in local communities and the health care inequities within hospitals and clinics, the medical and scientific community must listen and adopt shared leadership models that prioritize communities in health interventions, including COVID-19 vaccine development and distribution. This begins with recruiting and retaining a diverse health care workforce that reflects the local community and ensuring broader access to care before, during, and after public health crises.

Institutions must prioritize and support these efforts in clinical care, medical education, and research and embed their work through the lens of the communities served. Processes for continuous review and modification of systems-level policies and practices should be integrated into every aspect of health care to eliminate inequities. As global partners and leaders, the medical and scientific communities must make bold changes for the world to overcome two public health crises—racism and COVID-19. The greatest of these is to resist any further passivity and delay toward remedying social injustices and the structural racism that has proven beyond the COVID-19 pandemic that the burden is too heavy to further ignore it. ■

Remedying racial inequities

Taking action in the medical and scientific community requires a rigorous analysis of the use of race in medical education, clinical care, and research; continuous institutional assessment using an equity lens; and partnering with local communities to advance health equity.



GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. E. J. Williamson *et al.*, *Nature* **584**, 430 (2020).
2. J. A. W. Gold *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1517 (2020).
3. K. Mackey *et al.*, *Ann. Intern. Med.* **174**, 362 (2021).
4. L. E. Egede, R. J. Walker, *N. Engl. J. Med.* **383**, e77 (2020).
5. S. Singu, A. Acharya, K. Challagundla, S. N. Byrareddy, *Front. Public Health* **8**, 406 (2020).
6. Institute of Medicine (US) Committee on Quality of Health Care in America, *Crossing the Quality Chasm: A New Health System for the 21st Century* (National Academies Press, 2001).
7. Agency for Healthcare Research and Quality, "2019 national healthcare quality and disparities report" (U.S. Department of Health and Human Services, 2020); www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/2019qdr.pdf.
8. D. A. Vyas, L. G. Eisenstein, D. S. Jones, *N. Engl. J. Med.* **383**, 874 (2020).
9. S. Ahmed *et al.*, *J. Gen. Intern. Med.* **36**, 464 (2021).
10. B. W. Teo *et al.*, *Adv. Chronic Kidney Dis.* **25**, 41 (2018).
11. C. Amutah *et al.*, *N. Engl. J. Med.* **384**, 872 (2021).
12. K. M. Hoffman, S. Trawalter, J. R. Axt, M. N. Oliver, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4296 (2016).
13. White Coats for Black Lives, Racial justice report card (2019); https://whitecoats4blacklives.org/rjrc/.
14. R. R. Hardeman, J. Karbeah, *Health Serv. Res.* **55**, 777 (2020).
15. D. R. Williams, S. A. Mohammed, *J. Behav. Med.* **32**, 20 (2009).

10.1126/science.abf6738

RETROSPECTIVE

José Manuel Baselga (1959–2021)

Transformative oncologist who laid the groundwork for precision medicine

By **Maurizio Scaltriti**

José Manuel Baselga, renowned oncologist, died on 21 March at the age of 61 at his home in La Cerdanya, a beautiful Catalan county surrounded by the Spanish and French Pyrenees. Possibly the greatest drug developer of the past two decades, José was famous for, in his own words, "going to places and changing things." Throughout his career, he transformed oncology practice and made the development of new treatments his mission.

Born in Barcelona, Spain, on 3 July 1959, José was raised close to the Vall d'Hebron Hospital, an institution on which he would later leave his mark. He graduated from Universitat Autònoma de Barcelona in 1982 with an MD and then fulfilled his military service duty before starting his residency at the Vall d'Hebron Hospital in 1985. With a year left in his residency, José moved to the United States to work at Kings County Hospital in Brooklyn, New York. According to his sister Eulalia, he lived in a modest apartment with windows that would not close in winter, but the accommodations were good enough given that he was rarely home. José would spend his weekends and vacation time at Memorial Sloan Kettering Cancer Center (MSK) running lab experiments. He then moved full time to MSK under the mentorship of oncologists John Mendelsohn and Larry Norton. Together, they worked on newly identified antibodies targeting epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2), the predecessors of the anticancer therapies cetuximab, approved for the treatment of both colorectal and head and neck cancers, and trastuzumab, which drastically changed the prognosis of HER2-positive breast cancer patients. His work at MSK also earned him a PhD from Universitat Autònoma de Barcelona in 1992.

In 1996, José returned to Vall d'Hebron Hospital with plans to implement the MSK research model, specifically the intermingling of basic science with translationally informed oncology practice. According to some at that time, he was making a mistake moving from one of the best cancer hospitals in the United States to an unknown public hospital in Spain with no tradition in oncology research. To prove his naysayers wrong, José focused on what would later become a recurring pattern in his leadership: hiring the best people. Many of those he hired and trained went on to leadership roles in the oncology community. José knew that he also needed laboratories for basic and translational research, so he convinced Joaquin Arribas, a senior postdoc from the MSK lab of Joan Massagué, to follow him and help transform an empty 70-m^2 room in the middle of the oncology practice facilities into a top-notch research center. In just a few years, José put Barcelona and Vall d'Hebron Hospital on the cancer research map, building a bed-to-bench culture that we now call "precision medicine."

I met José in 2004 at Vall d'Hebron. After 40 minutes in his office, he offered me a job, but in typical fashion gave me only 72 hours to decide and begin work. I accepted in 48 hours but managed to negotiate a more leisurely start date. That decision marked the beginning of a 16-year relationship that would transform my professional and personal life.

In 2010, we moved together to Boston, Massachusetts. José accepted the position of chief of oncology and codirector of the Massachusetts General Hospital (MGH) Cancer Center, and I followed him to build his laboratory. His vision was clear to everyone: a clinical investigation program worthy of the name of one of the world's best hospitals. Once again, José was not afraid of betting on young horses. Dejan Juric, a director at MGH who was then still in fellowship, became the heir-apparent to the MGH Phase I program. Astounded by José's drive, energy, and commitment, Dejan said to me at the time, "Mauri, I thought I was ready for anything, but I was not ready for José."

In January 2013, we again moved together, this time to MSK, where José made sure that in addition to managing his laboratory, I got my own independent group. What José did there in 6 years exceeded all expectations. I was in his office when, after a few days in the job, he made a phone call to someone asking the price for a biopsy (critical for translational work in clinical trials). In his way, he informed whoever was on the call that from that moment, they needed to cut the price in half. He founded and then pushed the Center for Molecular Oncology and the Early Drug Development Service to become the gems of precision medicine they are today. José expanded his vision by hiring physicians and scientists to build MSK-IMPACT, the targeted exome sequencing platform capable of genomically characterizing virtually every patient who walks into the clinic. Meanwhile, he empowered a young oncologist named David Hyman with the mandate and the resources to develop the best Phase I unit possible.

In September 2018, José resigned from MSK after it was revealed that he had failed to disclose ties to drug companies. Although Jose's scientific achievements were never called into question, this incident highlighted the need for a long-overdue debate about the uniformity and clarity of conflict disclosures in medical and scientific journals.

In what would become his final challenge, José brought his experience and stamina to AstraZeneca to reform oncology R&D in his singular manner. He immediately hired new talented clinicians and researchers to enhance what was already a successful team and built a perfectly oiled machine to test and approve new cancer drugs.

José was direct, tough, demanding, and unforgiving, but he was also funny and caring. He would often display many or all of these characteristics in the same day. He hated mediocrity and pushed everyone around him to strive for perfection. Directly or indirectly, José helped the development and approval of a dozen drugs. In doing so, he saved and improved the lives of hundreds of thousands of cancer patients around the world. ■

AstraZeneca, Cambridge, UK.
Email: maurizio.scaltriti@astrazeneca.com

10.1126/science.abi7805

PHOTO: JULIANA THOMAS/MEMORIAL SLOAN KETTERING CANCER CENTER

Every American detonation, such as this one in 1946, offered clues to other potential bomb builders.

PHOTO: FRITZ GORO/THE LIFE PICTURE COLLECTION/GETTY IMAGES

HISTORY OF SCIENCE

Crafting a culture of secrecy

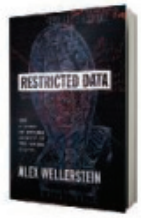
Clandestine research defined America's atomic age

By **Kate Brown**

Secrecy was a defining aspect of the creation of the atomic bomb and, 75 years later, nuclear secrecy remains a feature of American democracy. In *Restricted Data*, Alex Wellerstein examines the health of democracy in the face of big science, big government, and big weapons. "The bomb," he points out, "may have been born in secrecy, but that secrecy was always controversial and always contested."

Wellerstein conceives of nuclear secrecy as a "regime," or a bundle of thoughts, activities, and organizations that seek to normalize the performance of secrecy. Seeing secrecy as a set of practices enables him to research widely into not only censorship and clearance but also media releases, publicity, propaganda, and the trade and control of material objects.

Restricted Data explores the discovery of spies within the Manhattan Project, the denial of Robert Oppenheimer's security clearance, and the 1950s paradox of promoting peaceful uses of nuclear power while keeping nuclear weapons secret. The book adds to these histories an analysis of the "anti-secrecy" campaigns wherein Americans sought to break the Cold War code of classified nuclear knowledge.



Restricted Data
Alex Wellerstein
University of Chicago Press, 2021. 528 pp.

Wellerstein shows that the scientists who produced classified materials were consistently among the chief opponents of secrecy. They grasped, better than nonscientists did, that the bomb itself was its own greatest promoter. Every American detonation offered to Soviets an intelligence windfall in the form of radioactive debris that delivered information about the fuel used, the ratios of fission and fusion, and aspects of the bomb's design.

But not all secrets are equal. The essence of the Teller-Ulam hydrogen bomb (H-bomb) design, for example, was so compact that it could be handed over on the back of a napkin. Wellerstein shows how, in pursuit of the secret of the H-bomb, America shifted its security efforts from stopping the infiltration of foreign entities to scouring for double agents, scrutinizing insiders, and seeking continual affirmations of loyalty.

As the American security regime metastasized, the desire to control classified information fell in step with desires to control ever-larger portions of the globe. With ballooning security budgets, American security officials clandestinely influenced international organizations, such as the International Atomic Energy Agency and nonprofit foundations such as the Rockefeller and Ford Foundations, while engaged in wars and coup d'états in Asia and Latin America. After the release of the Pentagon Papers, Americans demanded transparency.

Wellerstein explores the frustrated efforts of those who actively sought to establish "anti-secrecy" practices aimed at exposing restricted data, revealing how such individuals could fall victim to convoluted logic in much the same way as those who pursued censorship in the name of democracy and freedom. Peppered throughout this section are many quotable passages on the metarelationship of government, societies, and individuals to state secrets. If knowledge is power, Wellerstein writes, "then nuclear knowledge is quite a lot of power."

One comes away with the impression that the only important historical actors in the 20th-century United States were white men who were scientists, administrators, or politicians. There is little in the way of class, race, or gender in this book. This omission takes as a matter of course the question of who decides what is "secret," who gets clearance, and who acts on behalf of everyone else, while failing to mention that the people making these decisions were almost always white, male professionals. Declassification campaigns, meanwhile, were often waged by nonwhites, members of the working class, Indigenous people, and women.

By following only a small portion of the population, Wellerstein omits a major facet of nuclear security. The vast majority of people directly exposed to radioactive fallout were soldiers, prisoners, children, minorities, and colonial subjects. This was perhaps one of the most guarded secrets the US government held—that the public it had vowed to protect and defend was under daily bombardment from radioactive debris from the testing and production of nuclear weapons.

Carl Schmitt, a political theorist and prominent Nazi, noted famously that during emergencies, the public hands over its rights to the state, and that the state rarely gives them back. *Restricted Data* illustrates that insight in spades. "It is easier," Wellerstein concludes, "to imagine the elimination of nuclear weapons than an elimination of the secrecy surrounding them." ■

The reviewer is at the Program in Science, Technology, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: brownkl@mit.edu

10.1126/science.abh2029

ROBOTICS

Man's (next) best friend

Our relationships with animals offer insight into the future of human–robot relations

By **Joshua C. Gellers**

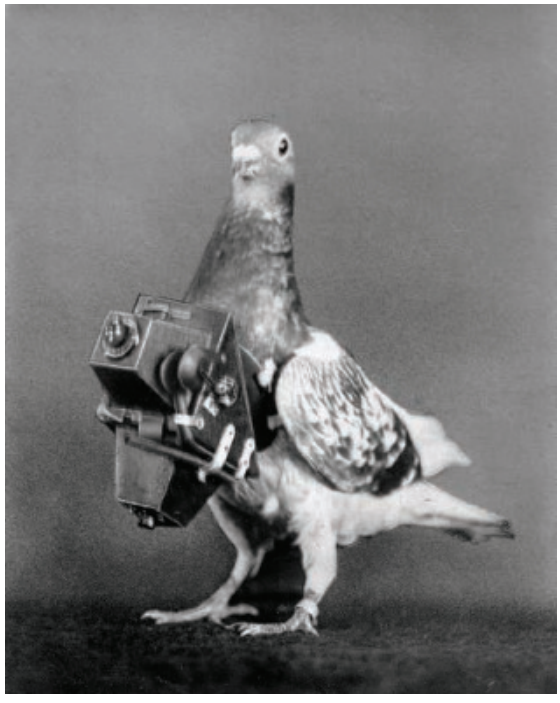
The New Breed: What Our History with Animals Reveals about Our Future with Robots
Kate Darling
Henry Holt and Co., 2021. 256 pp.

Long hidden from public view on factory floors, robots today are increasingly located in spaces where they exist alongside humans. This development has generated considerable hand-wringing. Will robots eventually take our jobs? Will they replace human romantic partners? In *The New Breed*, Kate Darling argues that equating robots with humans deprives us of our ability to determine the fate of humanity in a robot-filled world. She argues that our relationships with animals present an analogy more suitable for predicting how society will evolve as robots become part of it.

Darling begins by illustrating how animals have long held roles that only seem to elicit concern when robots enter the conversation. We have domesticated donkeys to plow fields, brought dogs into battle, and directed birds to deliver messages. Robots, she argues, like animals before them, will augment human abilities, not replace humans altogether.

Whether humans become obsolete is a choice, not an inevitability, and this decision must be placed in the context of the institutions that structure our lives. Crucially, Darling identifies capitalism as the culprit responsible for encouraging "short-term corporate profits rather than long-term investments in humans." Our "moral panic" about the robot invasion is misplaced, she argues. Instead, we should devote our energy to "holding our governments and corporations accountable and demanding that our economic and political systems do better for people."

The book's second section deals with companionship between humans and robots. Here Darling's own research shines brightly. In opposition to those who contend that humans should treat robots like tools or that roboticists should not design intelligent machines to look like humans, Darling demonstrates that we are hardwired to perceive nonhumans capable of movement and other abilities as possessing vitality. "Our desire to anthropomorphize…isn't going away," she notes. This is an important point too often missed by critics of humanoid robotics who mistakenly believe that we can circumvent questions about moral status by ignoring the innate human tendency to project ourselves onto other beings.



Like this pigeon outfitted with a camera (circa 1910), robots will likely augment human abilities, not replace humans altogether.

Darling's chapter on animal companions is an enjoyable one that even those who have no interest in robots would likely relish. Here, she reminds us that humans have a long history of establishing meaningful relationships with animals. Therefore, in principle, there is no reason why we could not also forge important ties with robots. The next chapter adds that, following our trajectory with animals, robots "will become a new type of relationship altogether" and, despite our anthropomorphic urges, "we're unlikely to confuse them with the alternative."

Darling believes we should be less worried about forming relationships with robots and more worried about how these relationships could be exploited. If we fail to take an "intentional" approach to robot design, she insists, we leave ourselves vulnerable to corporate coercion, reinforcing biases, and invasions of privacy. Strong regulations and enforcement mechanisms will be needed to avoid these potential pitfalls.

The book's final section addresses how humans should treat robots. Here the literature on animal rights makes a late but essential entrance. While our treatment of animals is "rife with inconsistencies," empathy, she maintains, might be the key to understanding our obligations toward robots. Perhaps to the disappointment of some, Darling suggests that our emotions, not our reason, might best guide the design of legal protections for intelligent machines.

The New Breed offers readers an energetic and witty overview of how our relations with animals can deliver useful insights into how robots might be incorporated into human society, but a couple of weaknesses might catch the attention of specialists. Darling's emphasis on human needs and empathy, for example, reinforces the kind of anthropocentric thinking that has produced animal suffering and ecological devastation. In addition, her exclusive focus on Western philosophy and law gives short shrift to important ideas about the relational personhood of nonhuman entities and the fundamental interconnectedness of all life forms that are articulated in Eastern and Indigenous worldviews. Despite these shortcomings, this book succeeds in arresting the alarmism that has pervaded recent popular writing on robots.

Darling ultimately makes a strong case that while our future will indeed include robots, it remains up to us to decide how to adjust our systems to accommodate our new companions. By examining the past and present of our relationships with animals, she shows how we might learn lessons that will help us shape our technological future for the better. ■

10.1126/science.abg8074

The reviewer is at the Department of Political Science and Public Administration, University of North Florida, Jacksonville, FL 32224, USA, and the author of *Rights for Robots: Artificial Intelligence, Animal and Environmental Law* (Routledge, 2020). Email: josh.gellers@unf.edu

PHOTO: BOYER/ROGER VIOLLET/GETTY IMAGES

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

中国空间技术研究院
China Academy of Space Technology
中国航天

Space: Science & Technology

Space: Science & Technology is an online-only, Open Access journal published in affiliation with **Beijing Institute of Technology (BIT)** and distributed by the **American Association for the Advancement of Science (AAAS)**. BIT cooperates with China Academy of Space Technology (CAST) in managing the journal. The journal aims to promote the exploration and research of space worldwide, to lead the rapid integration and technological breakthroughs of interdisciplinary sciences in the space field, and to build a high-level academic platform for discussion, cooperation, technological progress and information dissemination among professional researchers, engineers, scientists and scholars.

Submit your research to *Space: Science & Technology* today!

Learn more at **spj.sciencemag.org/space**

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

Visit us at **spj.sciencemag.org**

@SPJournals **@SPJournals**

OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED UNTIL JULY 2023

LETTERS

The effect of COVID-19 vaccine dosing regimens on the emergence of new variants depends on a range of factors.

Edited by **Jennifer Sills**

Partial immunity and SARS-CoV-2 mutations

In their Research Article, "Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes" (this issue, p. 363; published online 9 March), C. M. Saad-Roy *et al.* write that severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccine dosing regimens generating intermediate levels of immunity could accelerate the emergence of new variants that are capable of wholly or partially escaping immunity induced by prior infection or vaccination—immune-escape variants. Their argument assumes that such variants are most likely to arise through de novo mutation and selection in partially immune hosts (i.e., those people who have not mounted a strong immune response to infection or vaccination). However, for a pathogen like SARS-CoV-2 that typically transmits in the early stage of infection after relatively few cycles of replication (*1*), there is little opportunity for adaptive mutants to be generated and rise to a frequency that makes onward transmission likely. Simultaneously, there is good reason to think that intermediate levels of immunity should both reduce the probability of infection and limit the supply of adaptive mutations by restricting the viral population size within each vaccinated host (*2*, *3*). Furthermore, viral loads appear lower in infections 12 to 28 days after a single dose of vaccine than in unvaccinated individuals, which likely translates to decreased transmission by vaccinated individuals (*4*).

Because of limited opportunities for both mutation and selection within a host, the presence of immune-escape variants may be more important at the point of infection of immune individuals (*2*, *3*). If so, what matters is the standing diversity of such variants in the population as a whole. Reducing numbers of infections through vaccination should then reduce opportunities for such diversity to be generated, transmitted, and selected. In short, vaccination is likely to make evolution slower, and this benefit may be realized by initially optimizing breadth of vaccine coverage, rather than strength of immunity.

We do not advocate for delayed dosing strategies without further clinical evidence, and it is important to consider the issues raised by Saad-Roy *et al.*, which will depend on the properties of the vaccines involved. However, anxiety about the potential of vaccination to increase the emergence rate of immune-escape variants should be tempered by the low probability of the confluence of mutation, selection, and transmission as well as the enormous public health benefits of widespread vaccination.

William P. Hanage[1] and Colin A. Russell[2]
[1]Department of Epidemiology, Harvard University T. H. Chan School of Public Health, Boston, MA 02115, USA. [2]Department of Medical Microbiology, Amsterdam University Medical Center, Amsterdam, Netherlands.
Email: w.hanage@hsph.harvard.edu; c.a.russell@amsterdamumc.nl

REFERENCES AND NOTES

1. X. He *et al.*, *Nat. Med.* **26**, 672 (2020).
2. D. H. Morris *et al.*, *eLife* **9**, e62105 (2020).
3. S. Cobey, D. B. Larremore, Y. H. Grad, M. Lipsitch, "Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination," preprint, Harvard University (2021); https://dash.harvard.edu/handle/1/37366988.
4. M. Levine-Tiefenbrun *et al.*, *medrxiv* 10.1101/2021.02.06.21251283v1 (2021).

COMPETING INTERESTS

W.P.H. was paid to provide expert witness testimony in a litigation matter related in part to the expected duration of the COVID-19 pandemic.

10.1126/science.abi4727

Response

Hanage and Russell conjecture [see also (*1*)] that intermediate levels of immunity may sufficiently restrict within-host viral population size and thus limit adaptive mutations. The evolutionary model in our Research Article encompasses a wide range of scenarios, including the optimistic one Hanage and Russell propose (see scenario 1 in Fig. 4A of the Research Article). However, we also argued that uncertainties in immunodynamics, and in particular evolutionary dynamics of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), dominate our ability to project key scenarios. We therefore produced generic models in an effort to encompass this variability while providing a useful framework for considering possible future outcomes.

PHOTO: RICARDO ARDUENGO/AFP/GETTY IMAGES

We would also like to clarify that the calculations for evolutionary potential in our Research Article involve only infections after natural or vaccinal immunity has waned [i.e., I_S (secondary infection after waned natural immunity), I_{S1} (infection after waned one-dose vaccinal immunity), and I_{S2} (infection after waned two-dose vaccinal immunity), see Fig. 1]. Our accompanying interactive online application allows for the exploration of the effect of an even wider range of parameters on immunological and epidemiological outcomes.

We stress that the understanding of viral phylodynamics (the intersection of the epidemiological and evolutionary dynamics of pathogens) is still in its infancy, especially for novel pathogens like SARS-CoV-2. We therefore believe that exploring the resulting uncertainty is key. Understanding the dynamics of different classes of individuals experiencing infections after the waning of natural or vaccinal immunity will be crucial for teasing out drivers of viral immune escape (i.e., the ability of a virus to evolve to evade host immune factors); this is at the heart of our model. Along these lines, and as we highlight in the Research Article, an important area of future work will be to develop phylodynamic models with explicit within-host dynamics (*2*).

The other major source of uncertainties is in epidemiological outcomes. In their Letter, Hanage and Russell emphasize the likely epidemiological importance of widespread vaccine deployment. This echoes a conclusion of our Research Article, wherein we stress that short-term dose sparing deployment of a vaccine reduces infections and buys much time in public health planning. However, we also stressed the range of uncertainties that may modulate the longer-term outcomes of this strategy; in particular, less robust immunity could lead to more complex epidemiological and evolutionary outcomes.

Chadi M. Saad-Roy[1]*, Sinead E. Morris[2], C. Jessica E. Metcalf[3,4], Michael J. Mina[5], Rachel E. Baker[3,6], Jeremy Farrar[7], Edward C. Holmes[8], Oliver G. Pybus[9], Andrea L. Graham[3], Simon A. Levin[3], Bryan T. Grenfell[3,4,10]*, Caroline E. Wagner[11]*

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540, USA. [2]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA. [3]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [4]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA. [5]Departments of Epidemiology and Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [6]High Meadows Environmental Institute, Princeton University, Princeton, NJ 08544, USA. [7]The Wellcome Trust, London NW1 2BE, UK. [8]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences, and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia. [9]Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK. [10]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [11]Department of Bioengineering, McGill University, Montreal, QC H3A 0C3, Canada.

*Corresponding author.
Email: csaadroy@princeton.edu; grenfell@princeton.edu; caroline.wagner@mcgill.ca

REFERENCES AND NOTES

1. S. Cobey, D. B. Larremore, Y. H. Grad, M. Lipsitch, "Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination," preprint, Harvard University (2021); https://dash.harvard.edu/handle/1/37366988.
2. D. H. Morris *et al.*, *eLife* **9**, e62105 (2020).

10.1126/science.abi6719

PAST AS PROLOGUE

A new frontier for mi familia

With sleek black hair, finely arched eyebrows, and deep brown eyes that sparkle, my feisty, 5-foot-tall *Abuelita* (grandmother) is my last link to my ancestors, who spoke a unique dialect of New Mexican Spanish. As the last generation of a disappearing New Mexican lineage, whose *güerita* (light-skinned) appearance belies my heritage, I feel a responsibility to carry on the stories and traditions of my family; yet I feel ill-equipped to do them justice. Every life decision I make carries the weight of the unspoken question: How do I move forward and remember who I am when my family roots are being erased?

The author's great-great-grandparents pose for a photo in Tecolote, New Mexico, in the early 20th century.

PHOTO: RAVEN DELFINA OTERO-SYMPHONY

Immersed in the aromas of her homemade tortillas and red or green chile, Abuelita teaches me her native language in our effort to preserve my identity. Never a homeowner, Abuelita lives in a run-down studio apartment in the center of "the warzone," an impoverished area of Albuquerque adjacent to the heavily militarized Kirtland Air Force Base. "*No te preocupes, mi hijta*," she tells me, motioning to the SWAT vehicles that are a regular sight. "The more police, the safer the place!" Local research proves this to be untrue.

My childhood was filled with upheaval, and although school was always my refuge, I struggled to find where I belonged. My search came to an end with El Puente Research Fellowship, the Maximizing Access to Research Careers (MARC) scholars program, and the 2020 Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS) conference. In these communities, I found diverse scientific research that could answer that ever-present question. I believe that solving the multifaceted issues facing my hometown, from mass incarceration to severe substance addiction, lies within biomedical health research. I hope one day to use scientific evidence to influence policy in these areas, giving families like mine a better chance of stability and allowing their stories to continue unbroken. Science is my pursuit of a new frontier for my family: one of identity, belonging, and community.

My Abuelita is a free spirit whose unflinching resistance to discrimination is often misunderstood as truculence. As I pursue science to move myself and my family history forward together, it is my hope that I will someday come close to her humble survivorship and the resilience of the frontier generations that came before me.

Raven Delfina Otero-Symphony
MARC Scholars Program, Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA. Email: rsymphony32@unm.edu

10.1126/science.abi5545

Call for Submissions Past as Prologue is an occasional feature highlighting the role of family history in the life of scientists. What role did your family background play in your decision to pursue science, your field, or your career? Submit your story to www.submit2science.org.

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

China's climate change mitigation strategies include closing or adapting coal power plants for carbon capture.

PHOTO: JASON LEE/REUTERS

CLIMATE ECONOMICS

Change in the air

The 2016 Paris Agreement set the ambitious goals of keeping global temperature rise this century below 2°C, or even better, 1.5°C above preindustrial levels. Substantial interventions are required to meet these goals, particularly for industrialized countries. Duan *et al.* projected that China will need to reduce its carbon emissions by more than 90% and its energy consumption by almost 40% to do its share in reaching the 1.5°C target. Negative emission technology is an essential element of any plan. China's accumulated economic costs by 2050 may be about 3 to 6% of its gross domestic product. —HJS *Science*, this issue p. 378

EPIGENOME

Replication timing organizes epigenome

The temporal order of DNA replication is conserved from yeast to humans, but its biological significance remains unclear. Klein *et al.* eliminated the protein RIF1, a master regulator of replication timing, in several human cell lines. RIF1 loss during the G_1 phase of the cell cycle resulted in a heterogeneous, nearly random replication timing program from the first S phase that persisted even in stable RIF1-null clones. Altered replication timing was followed by replication-dependent redistribution of active and repressive histone modifications and alterations in genome architecture. These results support a model in which replication timing orchestrates the epigenetic state of newly replicated chromatin. —DJ

Science, this issue p. 371

QUANTUM DEVICES

A junction for single-photon detection

Josephson junctions are simple superconducting devices comprising an insulator or semiconductor separating two superconducting regions. They form the workhorse of superconducting technologies and are exquisitely sensitive to magnetic field. One long-sought proposal has been to use these devices to detect light. Walsh *et al.* have realized a photosensitive Josephson junction based on graphene that is capable of sensing single infrared photons. Such a photosensitive Josephson junction is expected to operate as a high-speed, low-power consumption optical interconnect for communication between superconducting-based supercomputers and quantum computers. —ISO

Science, this issue p. 409

CORONAVIRUS

Backtracking a pandemic

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) may have had a history of abortive human infections before a variant established a productive enough infection to create a transmission chain with pandemic potential. Therefore, the Wuhan cluster of infections identified in late December of 2019 may not have represented the initiating event. Pekar *et al.* used genome data collected from the early cases of the COVID-19 pandemic combined with molecular clock inference and epidemiological simulation to estimate when the most successful variant gained a foothold in humans. This analysis pushes human-to-human transmission back to mid-October to mid-November of 2019 in Hubei Province, China, with a likely short interval before epidemic transmission was initiated. —CA

Science, this issue p. 412

STROKE

Astrotherapy for stroke

Astrocytes, axons, and myelin are subjected to major damage during subcortical white matter stroke (WMS). Llorente *et al.* used a model of WMS in mice to investigate the transplantation of glial-enriched progenitor cells differentiated from human induced pluripotent stem cells (hIPSCs) shortly after stroke.

These cells matured into a mature astrocyte phenotype and had a therapeutic effect on axonal damage, demyelination, and cognitive impairments greater than that of hIPSC-derived neuronal precursor cells. —MM

Sci. Transl. Med. **13**, eaaz6747 (2021).

ORGANIC CHEMISTRY

Adding methyl groups with good timing

In pharmaceutical research, swapping out hydrogens for methyl groups is a frequent strategy to optimize small-molecule properties. Vasilopoulos *et al.* report a versatile, convenient, and comparatively safe method for methylation of carbon centers adjacent to nitrogen or aryl rings. Under carefully optimized conditions, di-*tert*-butyl peroxide plays a dual role as oxidant and methyl source. Cleaving the O–O bond through photosensitization produces butoxyl radicals, some of which cleave substrate C–H bonds, whereas others release methyl radicals that a nickel catalyst delivers to those activated substrates. —JSY

Science, this issue p. 398

HEAT TRANSPORT

Broadband thermal beaming

Thermal radiation emits over a wide range of wavelengths and over a wide range of angles. Xu *et al.* constructed a material that allows a range of wavelengths to emit over a much narrower range of angles. This property allowed the authors to beam thermal energy preferentially in one direction. The strategy requires carefully exploiting stacks of epsilon-near-zero films in which the angular range of thermal emission is controlled by the film thickness. This design could be useful in thermal camouflage and passive radiative cooling applications. —BG

Science, this issue p. 393

IMAGES: (LEFT TO RIGHT) XU *ET AL.*; NASA/JPL-CALTECH/ASU

Thermal camera image revealing changes in apparent temperature at different angles for a material with directional emissivity.

IMMUNOLOGY

Systemically inflamed from afar

The spleen is thought to be the major source of the proinflammatory cytokine tumor necrosis factor (TNF) during systemic inflammation. However, Fonseca *et al.* found that the liver and lungs produced more TNF than did the spleen in response to systemic inflammation induced by the bacterial cell wall component lipopolysaccharide (LPS) in rats. In addition, much of the spleen-dependent, LPS-induced increase in circulating TNF depended on the production of TNF by Kupffer cells, the resident macrophages of the liver. —AMV

Sci. Signal. **14**, eabb0969 (2021)

NEURODEVELOPMENT

Neuronal identities

Neurons of the mouse spinal cord can be identified by any of several metrics, including what neurotransmitters they use, what cells they connect to, where they are located, and what neuroprogenitor gave rise to them. Osseward *et al.* generated a different metric, genetic signatures, and identified classes of local and projection neurons that were otherwise heterogeneous by other classification systems. With this focus on a cell's genetic signature, its neurotransmitter phenotype, which is accessible by a variety of transcriptional routes, can be seen as a parallel to convergent evolution in development. —PJH

Science, this issue p. 385

IN OTHER JOURNALS

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

Artist's conception of the metal-rich asteroid Psyche

ASTEROIDS

Weighing Psyche

Remote observations have shown that the asteroid (16) Psyche has a metallic surface, usually interpreted as being the exposed iron core of a protoplanet that was disrupted by a past giant impact. Siltala and Granvik analyzed astrometric observations of 10 smaller asteroids that experienced close encounters with Psyche between 1974 and 2019. The gravitational perturbations of these asteroids' orbits allowed the authors to derive the mass of Psyche. Combined with its known diameter, this leads to a bulk density that is far too low for a solid iron body. The results support an alternative interpretation: that Psyche's surface formed by volcanic eruptions of liquid iron sulfide. The asteroid is due to be visited by a spacecraft (also named Psyche) in 2026. —KTS *Astrophys. J.* **909**, L14 (2021).

HEMATOLOGY

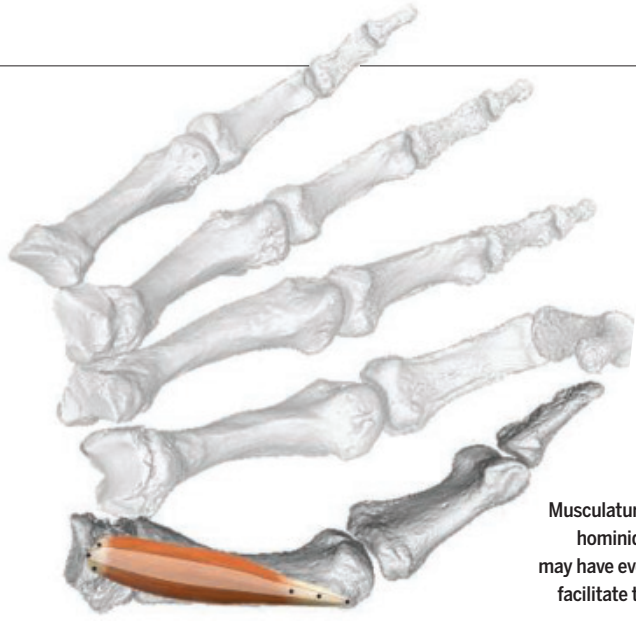
Hemoglobin in the balance

The inherited blood disorder β-thalassemia is caused by loss-of-function mutations in β-globin, a key component of adult hemoglobin. In addition to impaired oxygen delivery caused by the deficiency of β-globin, patients with severe forms of β-thalassemia suffer from a toxic buildup of excess α-globin, which normally pairs with β-globin. Cromer *et al.* developed a gene-editing strategy that addresses both problems at once by replacing one of the two α-globin genes with β-globin, allowing cells to produce fully functional adult hemoglobin and maintain a balance between the key globin types. The authors applied their gene-editing technique to hematopoietic stem cells from human

EVOLUTION

The opposable thumb

Hand bones of australopithecine hominid fossils dated to over 2 million years ago have a long, slender thumb that may have added dexterity. Karakostis *et al.* add muscle to the story of hand evolution with their analysis of torque at the thumb joint. A model validated on the hand movements of modern humans and chimpanzees was used to estimate the opposable thumb strength of fossil hominins. Torque at the thumb joint of early australopithecines, reflecting biomechanical efficiency, was comparatively low, but around 2 million years ago, the efficiency of the thumb joint began to increase. Further dexterity and thumb strength may have facilitated greater complexity in tool use and food acquisition as the genus *Homo* evolved. Tool-using skills could have evolved not so much in conjunction with big brains as much as with stronger, more dexterous hands. —PJH *Curr. Biol.* **31**, 1317 (2021).



Musculature of the hominid thumb may have evolved to facilitate tool use.

patients with β-thalassemia, and then successfully engrafted the edited cells into mouse models. —YN

Nat. Med. 10.1038/s41591-021-01284-y (2021).

NEUROSCIENCE

Unlearning reward responses

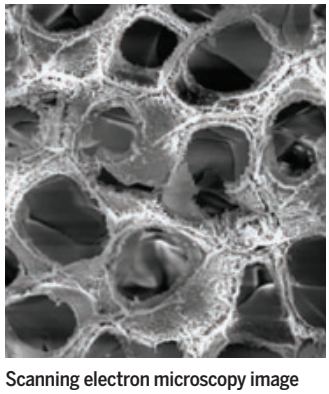
We continually update associations between new stimuli and behavioral responses, but this also requires that old information is subject to a complementary process called extinction learning. We do not fully understand whether this means that an old response is suppressed before a new one can be established, or if a new behavior has to compete with old associations. The rodent medial prefrontal cortex (mPFC) is involved in extinguishing reward-seeking behavior. By means of mPFC recordings, Russo *et al.* examined the temporal relation between neural activity and behavior in response to alcohol. Even when experimental conditions and behavioral responses were stable, recordings in the mPFC were not. However, shortly before the previously learned reward memory was suppressed, the authors found that changes in mPFC activity were highly coordinated across the whole reward network. This pushed the network into a new internal state that drove the extinction of the previous reward-seeking behavior. —PRS

J. Neurosci. **41**, 2406 (2021).

HYDROGELS

Pufferfish-inspired water purification

Water purification is an energy-intensive, multistep process used to remove contaminants such as salt, heavy metals, oils, or biological pathogens. Xu *et al.* designed a hydrogel that purifies water using nothing more than sunlight. Polydopamine, a melanin-based polymer, was added to sodium alginate on top of a microporous poly(N-isopropylacrylamide) (PNIPAm) hydrogel. Upon immersion in water, the hydrogel swells up with water while rejecting the contaminants, in part enhanced by the sodium alginate. Sunlight absorbed by the polydopamine forces the PNIPAm to heat up above its lower critical solution temperature, causing it to collapse, become hydrophobic, and expel the clean water. The process can be cycled, as the authors showed in a field test using lake water. —MSL



Scanning electron microscopy image showing the macroporous architecture of a bioinspired elastic hydrogel.

Adv. Mater. 10.1002/adma.202007833 (2021).

PHYSICS

Muons against the standard model

The standard model (SM) of particle physics has remained frustratingly reliable when tested in experiments. One exception is a 15-year-old measurement of the anomalous magnetic moment of the muon, which shows a discrepancy of 3.7 SDs with respect to the current SM prediction. Abi *et al.* (the Muon *g*−2 Collaboration) set out to verify this result using positive muons. Measuring the anomalous precession frequency of the muons in a magnetic field, they indeed found agreement with the previous result. Together, these two measurements deviate even more from the SM prediction, with a combined discrepancy of 4.2 SDs. Future experiments and calculations of the SM prediction will further test the discrepancy and determine what it means for theoretical extensions of the SM. —JS

Phys. Rev. Lett. **126**, 141801 (2021).

NEURODEGENERATION

How missing APOE4 protects

The apolipoprotein E (*APOE*) gene is the strongest genetic risk factor for Alzheimer's disease. Alzheimer's pathology involves both β-amyloid aggregation and phosphorylated tau–dependent neurodegeneration. The APOE4 variant is associated with worse tauopathy and neurodegeneration. In the brain, apoE is produced and secreted primarily by astrocytes and activated microglia. Wang *et al.* produced a transgenic tauopathy mouse model in which APOE4 (or APOE3 as a control) could be specifically down-regulated in astrocytes after disease onset. Removing astrocytic apoE4 reduced tau-mediated neurodegeneration and suppressed disease-associated gene expression in neurons, oligodendrocytes, astrocytes, and microglia. Furthermore, removal of astrocytic apoE4 decreased tau-induced synaptic loss and phagocytosis of synaptic elements by microglia. These findings shed light on the potential mechanisms underlying how and why APOE4 exacerbates Alzheimer's disease. —SMH

Neuron 10.1016/j.neuron.2021.03.024 (2021).

IMAGES (LEFT TO RIGHT): X. XU *ET AL., ADV. MATER.* 10.1002/ADMA.202007833 (2021); F. A. KARAKOSTIS *ET AL., CURR. BIOL.* **31**, 1318 (2021)

REVIEW SUMMARY

PLANT SCIENCE

Fluctuations shape plants through proprioception

Bruno Moulia*, Stéphane Douady*, Olivier Hamant*

BACKGROUND: Plants constantly experience fluctuating internal and external mechanical cues, ranging from nanoscale deformation of wall components, cell growth variability, nutating stems, and fluttering leaves to stem flexion under tree weight and wind drag. The plants' reactions seem ambivalent. Indeed, fluctuations may appear as noise that may confuse any sensing pathway, jeopardize the reproducibility of plant morphogenesis and development, or even result in mechanical hazards (e.g., stem windbreaks or growth cracks resulting from mechanical conflicts between differentially growing cells). At the same time, plants seem to enhance and even generate some fluctuations to remodel their structure (from the subcellular to the tissue scale), prime their morphogenesis, and control its outcome.

ADVANCES: In this Review, we explore the instructive role of fluctuations. At the organ scale, stems and leaves are one-dimensional (1D) or 2D thin structures embedded in a 3D space, and they display a continuous active degree of freedom along their growing tissues. They display large rotations during growth (e.g., tropism, primordia emergence, hook formation, etc.), bringing geometric nonlinearities and even mechanical elastic instabilities (buckling) or mechanical hazards caused by large bending and possible resonance under the drag of turbulent wind flow. At the cellular level, cells are also endowed with continuous growth that leads to many possibilities of mechanical conflicts with neighboring cells caused by differential growth and possible buckling out-of-plane instabilities. Finally, at the intracellular level, the cytoskeleton is made of thin microtubules and actin filaments, which undergo an autonomous dynamic of assembly-disassembly, leading to larger self-organized behaviors and instabilities. Thus, plant morphogenesis is the outcome of highly nonlinear, dissipative processes at all scales.

At first sight, this may make the effect of fluctuation even worse. However, developing plants use such fluctuations to monitor and channel their own shape and growth, through a form of proprioception. Proprioception (from the Latin words *proprius*, meaning own, and *capere*, meaning to take or to feel) is a sensory mechanism that involves the perception of cell- and tissue-shape changes linked to changes in body configuration and involves the sense of tension and deformation. We present evidence of proprioception in plants, and we show that this perception can feedback on growth control, allowing it to dampen the noisy effect of external and internal fluctuations. This is illustrated at the organ or global scale in stem and leaf tropisms and at the cellular or local scale in the control of cell growth heterogeneities in meristems and sepals.

Fluctuations in mechanical cues may also be actively enhanced, producing oscillating behavior in tissues. For example, proprioception through leaf nastic movements and active leaf fluttering may promote organ flattening. Because certain shapes are more amenable to fluctuations, proprioception may also help plant shapes to reach self-organized criticality to support such adaptability.

OUTLOOK: Fluctuations in plant shape should thus not be disregarded as noise with respect to the display of a genetic program. Instead, fluctuations promote proprioception at all scales. It follows that plants, like all living organisms, should be considered to be like active and self-sensing materials and structures: They perceive the fluctuations in their own shape originating from external or internal sources to monitor and channel their growth. Moreover, proprioception can be tuned flexibly to either dampen or enhance fluctuations, channeling shape reproducibility or, instead, priming symmetry breaking. The latter is made easier for highly symmetric shapes, and hence dynamically critical shapes, such as the ones that are reached through proprioceptive dampening. At such critical points, fluctuations occur at all scales. Coarser-grained fluctuations even span the whole system and trigger global changes and thus further morphogenetic steps. The balance between dampening and priming proprioception and the interplay between internal and external fluctuations is a target for further inquiries. ■

PHOTO CREDIT: JULIEN DERR AND RENAUD BASTIEN, LABORATOIRE MATIÈRES ET SYSTÈMES COMPLEXES

Fluctuation-based plant morphogenesis. Exogenous and endogenous fluctuations in curvature and growth are perceived at organ and cellular scales. Such proprioception can be ambivalent, either dampening or enhancing fluctuations, channeling shape reproducibility, or instead priming symmetry breaking. Proprioception may channel the formation of organs with shapes at critical points, which are also more amenable to fluctuations, thereby enhancing proprioception in a positive feedback loop.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: olivier.hamant@ens-lyon.fr (O.H.); bruno.moulia@inrae.fr (B.M.); stephane.douady@univ-paris-diderot.fr (S.D.)
Cite this article as B. Moulia *et al.*, *Science* **372**, eabc6868 (2021). DOI: 10.1126/science.abc6868

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abc6868

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Barcoded viral tracing of single-cell interactions in central nervous system inflammation

Iain C. Clark†, Cristina Gutiérrez-Vázquez†, Michael A. Wheeler†, Zhaorong Li, Veit Rothhammer, Mathias Linnerbauer, Liliana M. Sanmarco, Lydia Guo, Manon Blain, Stephanie E. J. Zandee, Chun-Cheih Chao, Katelyn V. Batterman, Marius Schwabenland, Peter Lotfy, Amalia Tejeda-Velarde, Patrick Hewson, Carolina Manganeli Polonio, Michael W. Shultis, Yasmin Salem, Emily C. Tjon, Pedro H. Fonseca-Castro, Davis M. Borucki, Kalil Alves de Lima, Agustin Plasencia, Adam R. Abate, Douglas L. Rosene, Kevin J. Hodgetts, Marco Prinz, Jack P. Antel, Alexandre Prat, Francisco J. Quintana*

INTRODUCTION: Glial cells of the central nervous system (CNS), including astrocytes and microglia, play critical roles in development, tissue repair, and homeostasis. However, dysregulated astrocyte and microglia responses contribute to the pathogenesis of neurologic diseases. Indeed, environmental chemicals, microbial metabolites, and cell-cell interactions have been shown to modulate disease-promoting responses in astrocytes and microglia in the context of multiple sclerosis (MS) and its model, experimental autoimmune encephalomyelitis (EAE). In particular, although astrocyte interactions with microglia are known to play important roles in the pathology of MS and other neurologic diseases, the pathways that facilitate astrocyte-microglia cross-talk are poorly understood, and consequently, few therapeutic inventions are available to target them.

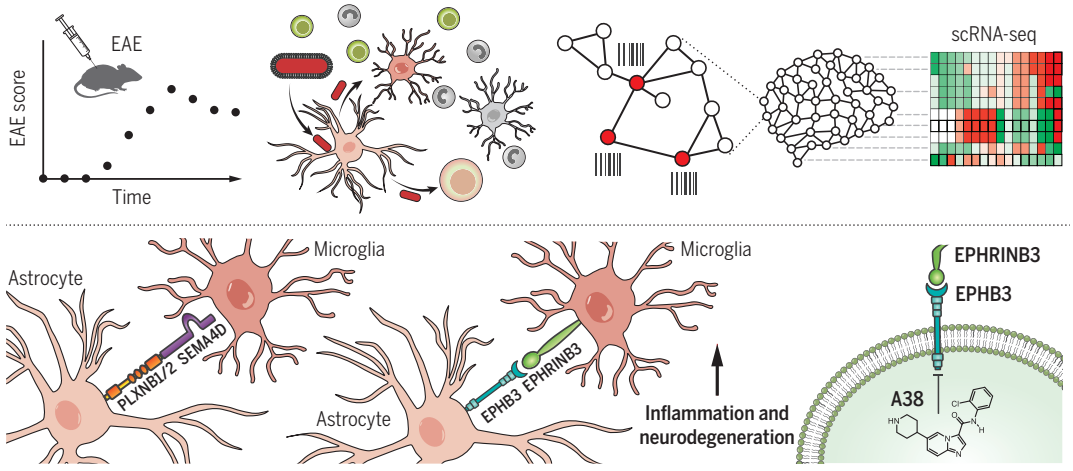
RATIONALE: Understanding the complexity of astrocyte-microglia cross-talk in CNS inflammation requires the study of precise neuroimmune interactions in vivo, but methodologies for defining the specific cell types, pathways, and molecules that mediate these interactions are limited. We developed a virus-based barcoding method for the identification of thousands of CNS cell interactions in vivo and the simultaneous analysis of the transcriptome of interacting cells with single-cell resolution. We applied this technique, named rabies barcode interaction detection followed by sequencing (RABID-seq), to the study of microglia-astrocyte communication in the context of CNS inflammation in EAE and MS.

RESULTS: To develop RABID-seq, we engineered an mRNA-barcoded library in glycoprotein G–deficient pseudorabies virus (RabΔG-BC), which spreads between interacting cells but can only replicate in cells that transgenically express viral glycoprotein G. We pseudotyped the RabΔG-BC plasmid library using envelope protein of subgroup A (EnvA) packaging. Thus, the pseudotyped virus only infects cells that transgenically express the EnvA receptor, TVA. After its replication in cells that express TVA and viral glycoprotein G, RabΔG-BC infects interacting cells, labelling them with the virus-encoded barcode. To study RABID-seq astrocyte interactions in vivo during CNS inflammation in the EAE model of MS, we used transgenic mice expressing glycoprotein G and TVA in astrocytes under the control of the *Gfap* promoter. These studies identified several axon guidance molecules as critical mediators of microglia-astrocyte interactions in the context of inflammation. By combining RABID-seq with genetic perturbation studies in vivo, validation with primary mouse and human cells in vitro, and the analysis of MS patient samples by immunostaining and single-cell RNA-seq, we established that microglia-astrocyte interactions mediated by Sema4D-PlexinB1, Sema4D-PlexinB2, and Ephrin-B3–EphB3 promote CNS pathology in EAE—and potentially MS. Notably, Ephrin-B3–EphB3 participated in forward and reverse signaling, which boosted both microglia and astrocyte pathogenic activities via the regulation of nuclear factor κB and mammalian target of rapamycin, respectively. Finally, we demonstrated that a CNS-penetrant small-molecule inhibitor of the kinase activity of the EphB3 intracellular domain ameliorates EAE in both acute and chronic progressive models.

CONCLUSION: We developed RABID-seq, a novel approach for the simultaneous investigation of cell interactions and the transcriptome of interacting cells in vivo with single-cell resolution. RABID-seq identified signaling pathways controlled by the axon guidance molecules Sema4D-PlexinB1, Sema4D-PlexinB2, and Ephrin-B3/EphB3 as mediators of microglia-astrocyte interactions that promote CNS pathogenesis and also as candidate targets for therapeutic intervention in neurologic disorders. ■

The list of author affiliations is available in the full article online.
†These authors contributed equally to this work.
***Corresponding author. Email: fquintana@bwh.harvard.edu**
Cite this article as I. C. Clark *et al.*, *Science* 372, eabf1230 (2021). DOI: 10.1126/science.abf1230

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf1230



Elucidation of microglia-astrocyte interactions by rabies barcode interaction detection followed by sequencing (RABID-seq). Pseudotyped rabies virus expressing barcoded mRNA targets *Gfap*$^+$ astrocytes, where it replicates before infecting neighboring cells, leaving a barcoded trace. Single-cell RNA sequencing reads both cellular mRNAs and viral barcodes, allowing for the reconstruction of in vivo cell interactions and the transcriptional analysis of interacting cells with single-cell resolution.

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Cholinergic neurons constitutively engage the ISR for dopamine modulation and skill learning in mice

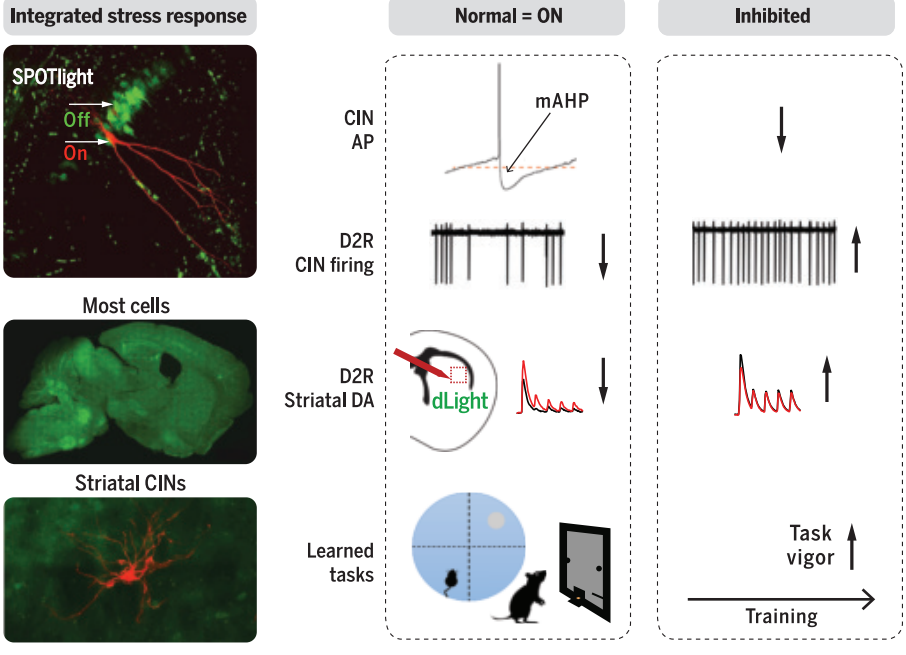
Ashley R. Helseth†, Ricardo Hernandez-Martinez†, Victoria L. Hall, Matthew L. Oliver, Brandon D. Turner, Zachary F. Caffall, Joseph E. Rittiner, Miranda K. Shipman, Connor S. King, Viviana Gradinaru, Charles Gerfen, Mauro Costa-Mattioli, Nicole Calakos*

INTRODUCTION: The integrated stress response (ISR) is a highly conserved biochemical pathway that, upon activation, markedly shifts which proteins are synthesized. Its roles in proteostasis, synaptic plasticity, learning, and memory have made the pathway an attractive therapeutic target for systemic and brain diseases. Preclinical studies showing the capacity of small-molecule ISR inhibitors to enhance some forms of learning and memory have further highlighted its translational potential. Despite strong and accumulating evidence for the ISR as a potent modifier of plasticity, learning, and memory in diverse behavioral paradigms, the cellular sites of action and time course of ISR engagement are less well understood.

RATIONALE: To better understand ISR roles in the brain, we developed a two-color reporter of ISR activation state, SPOTlight, and delivered it through a viral vector for brainwide imaging with cellular resolution. SPOTlight was designed to differentially translate green or red fluorescent proteins from a single transcript based on ISR activation state. We first established that reporter readouts corresponded to known manipulations of the ISR using immunohistochemical analyses. SPOTlight signal indicating high ISR activation at steady state in striatal cholinergic interneurons (CINs) was validated using immunohistochemical analyses. To understand the factors driving ISR activation in CINs, we inhibited CIN activity chemogenetically and assessed the ISR state. Selective pharmacological and genetic manipulations were combined with electrophysiological CIN recordings to establish the mechanisms by which ISR state influences CIN physiology. Fluorescent dopamine reporter imaging was used to examine the circuit-level effects of CIN ISR state on evoked striatal dopamine transients in acute brain slices. Finally, task training assays were used to measure the behavioral effects of CIN ISR activity.

RESULTS: Pharmacological ISR activators and inhibitors delivered in vivo differentially modified levels of SPOTlight-encoded fluorescent proteins and corresponded to immunohistochemical markers of ISR activation in the mouse brain. As expected, in the normal mature mouse brain, SPOTlight revealed only sparse cells with evidence of ISR activation. Unexpectedly, we also found a class of neurons that showed population-wide ISR activation: tonically firing striatal CINs. Chemogenetic inhibition of CIN firing reduced ISR activation, indicating an activity-dependent component. CINs also appeared to be distinctive in ISR engagement; a survey of SPOTlight in other cell types with tonic or high-firing properties did not show similarly high and population-wide ISR engagement. In CINs, manipulations inhibiting the ISR inverted the normal type 2 dopamine receptor (D2R) response from slowing to increasing CIN firing through a mechanism that reduced small-conductance calcium-activated potassium channel currents. Cell-autonomous ISR inhibition in CINs also inverted D2R modulation of evoked striatal dopamine and altered skill learning by increasing the velocity of movement in two learned tasks.

CONCLUSION: Our study defines a distinct role for the ISR in brain, neuromodulation, which expands our understanding of how the ISR influences learning and memory. We show that steady-state ISR activation in striatal cholinergic interneurons determines their response to dopaminergic modulation, shapes circuit-level dopamine release, and regulates learned skill performance. In this context, ISR activation is activity dependent and influences CIN cellular excitability. As ISR-inhibiting drugs move toward the clinical setting, our results highlight an unappreciated mechanism for their effects on learned behaviors. Our results also prompt further examination of the sites of ISR action in various forms of synaptic plasticity given the importance of cholinergic and dopaminergic neuromodulation in this process. Finally, SPOTlight provides a tool with which to explore when and where the ISR is activated across diverse contexts, including developmental, learning-related, and diseased states. ■



The ISR in striatal CINs. Using a two-color viral reporter of ISR activation state, striatal CINs were distinguished from other cell types by showing steady-state ISR activation. ISR inhibition altered CIN excitability, inverted the sign of dopamine modulation of CIN firing rate and evoked dopamine release, and increased movement vigor in learned tasks. mAHP, medium afterhyperpolarization; AP, action potential; DA, dopamine.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: nicole.calakos@duke.edu
†These authors contributed equally to this work.
Cite this article as: A. R. Helseth *et al.*, *Science* **372**, eabe1931 (2021). DOI: 10.1126/science.abe1931

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abe1931

RESEARCH ARTICLE SUMMARY

REGENERATION

Preventing *Engrailed-1* activation in fibroblasts yields wound regeneration without scarring

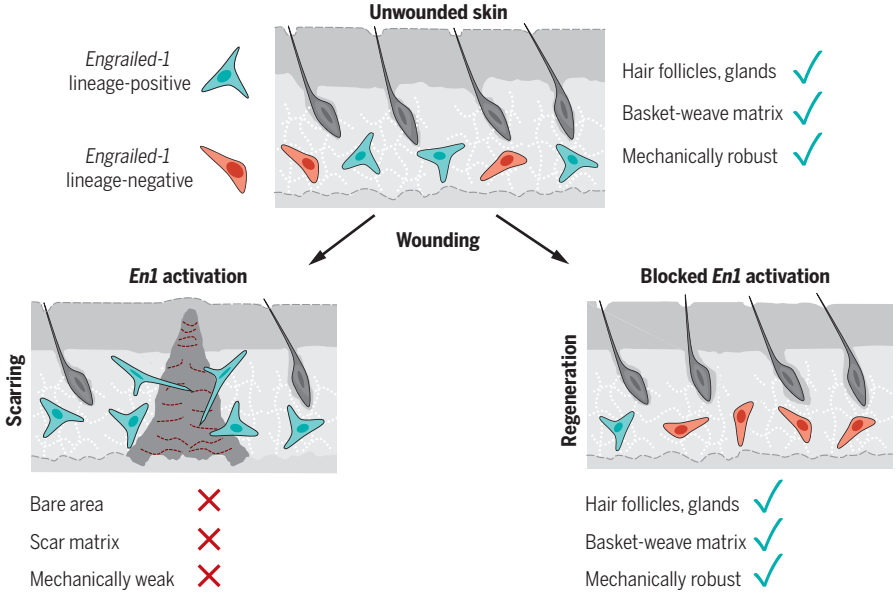
Shamik Mascharak, Heather E. desJardins-Park, Michael F. Davitt, Michelle Griffin, Mimi R. Borrelli, Alessandra L. Moore, Kellen Chen, Bryan Duoto, Malini Chinta, Deshka S. Foster, Abra H. Shen, Michael Januszyk, Sun Hyung Kwon, Gerlinde Wernig, Derrick C. Wan, H. Peter Lorenz, Geoffrey C. Gurtner*, Michael T. Longaker*

INTRODUCTION: Skin wounds generally heal by scarring, a fibrotic process mediated by the *Engrailed-1* (*En1*) fibroblast lineage. Scars differ from normal unwounded skin in three ways: (i) They lack hair follicles, sebaceous glands, and other dermal appendages; (ii) they contain dense, parallel extracellular matrix fibers rather than the "basket-weave" pattern of uninjured skin; and (iii) as a result of this altered matrix structure, they lack skin's normal flexibility and strength. A successful scar therapy would address these three differences by promoting regrowth of dermal appendages, reestablishment of normal matrix ultrastructure, and restoration of mechanical robustness. However, little is known about the cellular and molecular mechanisms blocking a regenerative healing response in postnatal skin, or whether these mechanisms can be bypassed by modulating specific fibroblast lineages.

RATIONALE: We asked whether scarring fibroblasts are derived purely from expansion of existing *En1* lineage–positive fibroblasts present in unwounded skin, or whether *En1* scar fibroblasts could arise de novo by activation of *En1* expression in postnatal, *En1* lineage–negative fibroblasts within the wound niche. We used fibroblast transplantation as well as transgenic mouse models to trace *En1* expression in a spatiotemporally defined fashion. Next, we studied fibroblast responses to mechanical forces in vitro and in vivo to establish a mechanotransduction mechanism linking skin tension to postnatal *En1* expression. Finally, we used chemical (verteporfin) and transgenic inhibition of mechanotransduction signaling [diphtheria toxin ablation of *En1*-expressing fibroblasts, floxed Yes-associated protein (YAP) knockout] to modulate *En1* expression during wound healing. Experimental wounds were compared to unwounded skin and scars (phosphate-buffered saline control) by RNA sequencing, quantitative histopathological comparison (using a custom image-processing algorithm), and mechanical strength testing.

RESULTS: Fibroblast transplantation and lineage-tracing studies reveal that *En1* lineage–negative fibroblasts (ENFs) of the reticular (deep) dermis activate *En1* in the wound environment, generating ~40 to 50% of scar fibroblasts. This phenomenon depends on mechanical cues: ENFs cultured on soft substrates or treated with chemical inhibitors of mechanical signaling proteins (e.g., YAP) do not activate *En1*. Comparison of ENFs with *En1*-expressing and *En1* knockdown (short hairpin RNA) fibroblasts by RNA sequencing suggests that *En1* regulates a wide array of genes related to skin fibrosis. In healing wounds, YAP inhibition by verteporfin blocks *En1* activation and promotes ENF-mediated repair, yielding skin regeneration in 30 days with recovery of functional hair follicles and sebaceous glands. Quantitative comparison of scars and regenerated skin shows that YAP inhibition induces recovery of normal dermal ultrastructure, which in turn confers restoration of normal mechanical breaking strength. Diphtheria toxin–mediated ablation of postnatal *En1*-expressing fibroblasts and fibroblast-targeted transgenic YAP knockout similarly promoted recovery of normal skin structures, which suggests that modulation of *En1* activation, whether direct or indirect, can yield wound regeneration.

CONCLUSION: By delineating how physical stimuli provoke ENFs to contribute to fibrosis, we identify YAP and *En1* as possible molecular targets to prevent scarring. Furthermore, we have shown that inhibition of YAP signaling prevents *En1* activation during wound healing, thus encouraging ENF-mediated wound repair without fibrosis and with regeneration of secondary skin elements (hair follicles, sebaceous glands). Our findings suggest that residual ENFs in postnatal mammalian skin retain a capacity for skin regeneration if the mechanically driven propensity for fibrosis can be blocked. We have demonstrated fully regenerative skin healing in a postnatal mammal that normally scars; this finding has translational implications for the tens of millions of patients each year who develop scars and other fibroses. ■



***Engrailed-1* activation in skin fibroblasts drives scarring.** After injury, a subset of dermal fibroblasts activates *Engrailed-1* (*En1*) to contribute to scarring (left). Inhibiting postnatal *En1* activation, either directly (by ablating *En1*-activating cells) or indirectly (by blocking mechanical signaling with verteporfin), promotes skin regeneration by *En1* lineage–negative fibroblasts, with full recovery of normal hair follicles, glands, matrix ultrastructure, and mechanical strength. Green cells, *En1* lineage–positive fibroblasts; red cells, *En1* lineage–negative fibroblasts.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: longaker@stanford.edu (M.T.L.); ggurtner@stanford.edu (G.C.G.)
Cite this article as S. Mascharak *et al.*, *Science* **372**, eaba2374 (2021). DOI: 10.1126/science.aba2374

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aba2374

RESEARCH ARTICLES

CORONAVIRUS

Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes

Chadi M. Saad-Roy[1]*, Sinead E. Morris[2], C. Jessica E. Metcalf[3,4], Michael J. Mina[5], Rachel E. Baker[3,6], Jeremy Farrar[7], Edward C. Holmes[8], Oliver G. Pybus[9], Andrea L. Graham[3], Simon A. Levin[3], Bryan T. Grenfell[3,4,10]*, Caroline E. Wagner[11]*

Given vaccine dose shortages and logistical challenges, various deployment strategies are being proposed to increase population immunity levels to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Two critical issues arise: How timing of delivery of the second dose will affect infection dynamics and how it will affect prospects for the evolution of viral immune escape via a buildup of partially immune individuals. Both hinge on the robustness of the immune response elicited by a single dose as compared with natural and two-dose immunity. Building on an existing immuno-epidemiological model, we find that in the short term, focusing on one dose generally decreases infections, but that longer-term outcomes depend on this relative immune robustness. We then explore three scenarios of selection and find that a one-dose policy may increase the potential for antigenic evolution under certain conditions of partial population immunity. We highlight the critical need to test viral loads and quantify immune responses after one vaccine dose and to ramp up vaccination efforts globally.

As the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) betacoronavirus pandemic continues, the deployment of safe and effective vaccines presents a key intervention for mitigating disease severity and spread and eventually relaxing nonpharmaceutical interventions (NPIs). At the time of writing, 11 vaccines have been approved by at least one country (*1*). We focus mainly on the vaccines from Pfizer/BioNTech, Moderna, and Oxford/AstraZeneca. The first two elicit adaptive immunity against SARS-CoV-2 in response to the introduction of mRNA molecules that encode the spike protein of SARS-CoV-2 (*2*) and appear to offer greater than 95% [Pfizer/BioNTech (*3*), approved in 60 countries] and 94% [Moderna (*2*), approved in 38 countries] protection against symptomatic COVID-19. Both of these mRNA vaccines were tested in clinical trials according to a two-dose regime with dose spacing of 21 and 28 days for the Pfizer/BioNTech and Moderna platforms, respectively. The Oxford/AstraZeneca vaccine uses a nonreplicating adenovirus vector and has also been tested in clinical trials according to a two-dose regime with a target 28-day interdose period (although for logistical reasons some trial participants received their second dose after a delay of at least 12 weeks). Clinical trials indicated 62 to 90% efficacy for this vaccine depending on the specific dose administered (*4*). Although we base our parameter choices and modeling assumptions on these three vaccines, our results are generalizable across platforms.

As these vaccines have been distributed internationally, several countries, including the UK (*5*) and Canada (*6*), have chosen to delay the second dose in an effort to increase the number of individuals receiving at least one dose or in response to logistical constraints (*7*). Although a number of participants dropped out after a single dose of the vaccine in the Pfizer/BioNTech and Moderna trials, these studies were not designed to assess vaccine efficacy under such circumstances, and Pfizer has stated that there is no evidence that vaccine protection from a single dose extends beyond 21 days (*5*), although other data paint a more optimistic picture (*8*, *9*). The Oxford/AstraZeneca clinical trials did include different dose spacings, and limited evidence suggests that longer intervals (2 to 3 months) did not affect, and may even have improved, vaccine efficacy (*4*, *5*). Ultimately, the consequences of deviating from manufacturer-prescribed dosing regimes at the population scale remain unknown but will hinge on immune responses.

Although there has been considerable progress in quantifying host immune responses after infection (*10*–*12*), substantial uncertainty regarding the strength and duration of both natural and vaccinal SARS-CoV-2 immunity remains. Previous work suggests that these factors will play a central role in shaping the future dynamics of COVID-19 cases (*13*). Future cases also create an environment for the selection of novel variants [e.g., (*14*–*16*)]. Of particular concern is the possibility of antigenic drift [e.g., for influenza (*17*) and for the seasonal human coronavirus 229E (*18*)] via immune escape from natural or vaccinal immunity. For example, immune escape might be especially important if vaccinal immunity elicited after the complete two-dose regime is highly protective, whereas a single vaccine dose provides less effective immunity. Consequently, the longer-term epidemiological and evolutionary implications of these different SARS-CoV-2 vaccine dosing regimes are not yet clear; the immediate need for effective mass vaccination makes understanding them critical for informing policy (*19*).

Here, we explore these epidemiological and evolutionary considerations with an extension of a recent immuno-epidemiological model for SARS-CoV-2 dynamics (*13*), depicted schematically in Fig. 1. Without vaccination, our model reduces to the susceptible-infected-recovered (-susceptible) [SIR(S)] model (*13*, *20*), where individual immunity after recovery from primary infection may eventually wane at rate δ, leading to potentially reduced susceptibility to secondary infections, denoted by the fraction ϵ relative to a baseline level of unity. This parameter ϵ is thus related to the (transmission-blocking) strength of immunity and titrates between the SIR (lifetime immunity, $\epsilon = 0$) and SIRS (hosts regain complete susceptibility, $\epsilon = 1$) paradigms. Quantifying ϵ is challenging because it requires measuring reinfection rates after the waning of immunity. Some studies have made substantial progress in this direction (*11*, *12*); however, uncertainties remain, particularly related to quantifying the average duration of immunity $1/\delta$. In this model extension (Fig. 1 and materials and methods), we incorporate two vaccinated classes; V_1 accounts for individuals who have received one dose of a SARS-CoV-2 vaccine, and V_2 tracks individuals who have received two doses. In the short term, we assume that both dosing options decrease susceptibility by fractions $(1-\epsilon_{V_1})$ (one dose) and $(1-\epsilon_{V_2})$ (two doses), inferred from the clinical trial data (although the nature of the infecting variant may influence susceptibility). We also assume that I_V tracks infection after vaccination. We allow for vaccinal immunity to wane at separate rates [ρ_1 (one dose) and ρ_2 (two doses)],

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540, USA. [2]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA. [3]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [4]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA. [5]Departments of Epidemiology and Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA. [6]High Meadows Environmental Institute, Princeton University, Princeton, NJ 08544, USA. [7]The Wellcome Trust, London NW1 2BE, UK. [8]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences, and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia. [9]Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK. [10]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [11]Department of Bioengineering, McGill University, Montreal, QC H3A 0C3, Canada.
*Corresponding author. Email: csaadroy@princeton.edu (C.M.S.-R.); grenfell@princeton.edu (B.T.G.); caroline.wagner@mcgill.ca (C.E.W.)

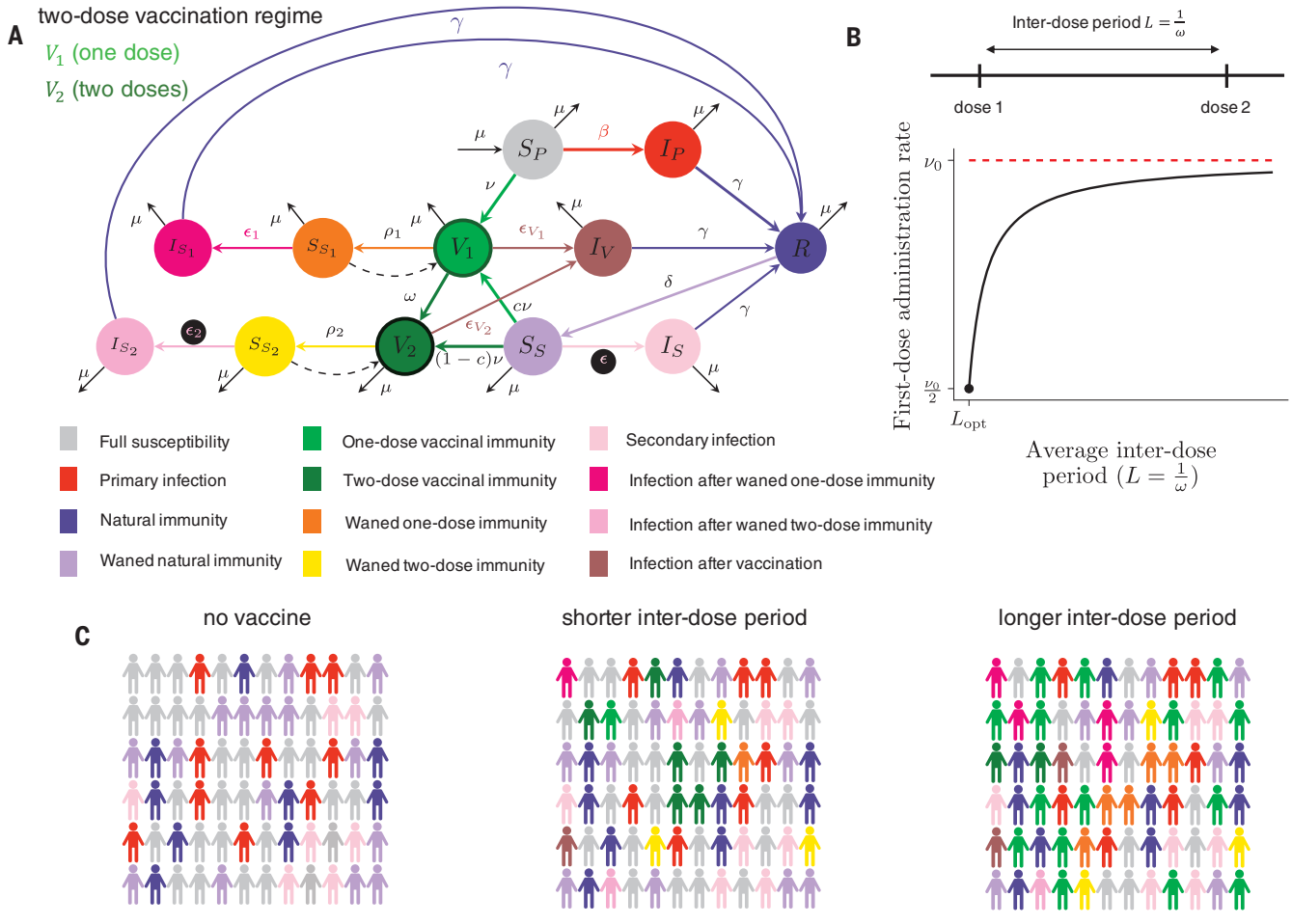


Fig. 1. Description of the extended immuno-epidemiological model with one- and two-dose vaccination regimes. Based on (*13*). (**A**) Model flow chart depicting transitions between immune classes (see main text and materials and methods for a full description of the immune classes and parameters). (**B**) Diagram of the interdose period ($\frac{1}{\omega}$) between the first and second vaccine doses and its relationship to the rate of administration of the first vaccine dose ν. The maximum achievable rate is ν_0 for a fully one-dose strategy, and ν is assumed to decrease exponentially to its lowest value $\nu_0/2$ when a fully two-dose strategy with interdose period corresponding to the clinical recommendation (L_{opt}) is used. (**C**) Representative schematic of societal composition of various immune classes for the SIR(S) model with no vaccination (left), the extended model with a short interdose period (middle), and the extended model with a long interdose period (right).

moving individuals to the partially susceptible immune classes S_{S_1} and S_{S_2} characterized by (possibly different) levels of immune protection ϵ_1 and ϵ_2. Infection after waned one-dose or two-dose vaccinal immunity is tracked by the immune classes I_{S_1} and I_{S_2}, respectively. We consider a continuous spectrum for the interdose period ($\frac{1}{\omega}$), with an infinite value corresponding to a one-dose strategy, and model the rate of administration of the first dose ν as an increasing function of the interdose period (Fig. 1 and materials and methods) to reflect the increase in available doses due to a delayed second dose. Thus, dosing regimes with longer interdose periods allow for higher coverage with the first dose.

We begin by projecting the epidemiological impacts of the different dosing regimes on medium-term temporal dynamics of COVID-19 cases. We then examine the potential evolutionary consequences of each dosing regime by calculating a time-dependent relative net viral adaptation rate (*17*). This term is related to the strength of natural and vaccinal immunity (either via inducing selection through immune pressure or suppressing viral replication) as well as the sizes of classes of individuals experiencing infections after immune waning.

Epidemiological impacts

As a base case, we consider a high-latitude European or North American city with initial conditions that qualitatively correspond to early 2021 (see supplementary materials and figs. S5 and S6 for other scenarios, e.g., a high initial attack rate or almost full susceptibility), in addition to a seasonal transmission rate (*21*) with NPIs (see materials and methods). Given immunological and future control uncertainties, we are aiming to project qualitatively rather than formulate quantitative predictions for particular locations. The UK and Canadian policy is for a delayed second dose; they are not aiming for an exclusively one-dose policy. However, we explore the one-dose strategy as an extreme case for the two-dose vaccines; this strategy also encompasses a pessimistic situation of waning public confidence in vaccination and individuals' own decisions to forgo the second dose. This one-dose policy could also capture vaccines that only require a single dose, e.g., the Johnson & Johnson vaccine.

In Fig. 2, we present potential scenarios for medium-term SARS-CoV-2 infection and immunity dynamics contingent on vaccine dosing regimes. We start by assuming that vaccination occurs at a constant rate, and we also assume a relatively optimistic maximum rate of administration of the first dose of ν_0 = 2% of the population per week (see supplementary materials for other scenarios). Figure 2A and Fig. 2B correspond, respectively, to scenarios with weaker (and shorter) and stronger (and

longer) natural and vaccinal adaptive immune responses. Thus, the former represents a scenario with higher secondary susceptible density than the latter. In each panel, the top and bottom sections consider poor and robust one-dose vaccinal immunity, respectively. The leftmost column represents a one-dose vaccine policy (captured in the model by infinite dose spacing), with dose spacing decreasing to 4 weeks in the rightmost column (i.e., a strict two-dose policy with doses separated by the clinical trial window corresponding to Moderna's recommendations for their vaccine, hereafter referred to as the "recommended" two-dose strategy).

As expected, we find that broader deployment of widely spaced doses is beneficial. Specifically, a one-dose strategy (or a longer interdose period) may lead to a substantially reduced first epidemic peak of cases after the initiation of vaccination (compare the leftmost top panels of Fig. 2, A and B, with the no-vaccination scenarios in fig. S1, A and B). This result applies even if immunity conferred by one vaccine dose is shorter and weaker than that conferred by two doses (top panels of Fig. 2, A and B). However, under these conditions of imperfect immunity, an exclusively one-dose strategy then leads to an earlier subsequent peak owing to the accumulation of partially susceptible individuals. When the rate of administration of the first dose is very high (fig. S4, ν_0 = 5% per week), this subsequent infection peak may be larger than that expected in the scenario with no vaccination. In general, the accumulation of partially susceptible individuals with waned one-dose vaccinal immunity can be mitigated by implementing a two-dose strategy and decreasing the time between doses. Thus, in situations of a less effective first dose where the second dose is delayed, it is important to ensure individuals eventually do obtain their second dose.

In line with intuition, longer and stronger immunity elicited after a single dose heightens the benefits of a one-dose strategy or of delaying the second dose (compare the top and bottom leftmost panels of Fig. 2, A and B). Additionally, the protective effects of adopting these strategies instead of the two-dose regime are maintained in the medium term, with decreased burden in all future peaks. These effects are further summarized in Fig. 3A via the cumulative number of total and severe cases (right and left panels, respectively) over ~4 years from the time of vaccine initiation, normalized by the burdens predicted when no vaccines are administered; these ratios are plotted as a function of the interdose period and the one- to two-dose immune response ratio x_e (see Fig. 3 caption for details). When the immune response conferred by a single dose is similarly robust to that conferred by two doses, total case numbers (Fig. 3A, right panel) can be substantially reduced by delaying the second dose. However, for smaller values of x_e, larger interdose periods are associated with more cases. The reduction in the cumulative burden of severe cases is even more sizeable (Fig. 3A, left panel) owing to the assumed reduction in the fraction of severe cases for partially immune individuals. When vaccination rates are substantially lower (fig. S2, ν_0 = 0.1% per week; and fig. S3, ν_0 = 1% per week), the benefits of a single-dose strategy diminish even for an effective first dose, as an insufficient proportion of the population is immunized. The short-term effect of the vaccine on case numbers is sensitive to when it is introduced in the dynamical cycle (figs. S7 and S8), highlighting the critical interplay between the force of infection and the level of population immunity (see supplementary materials for further details).

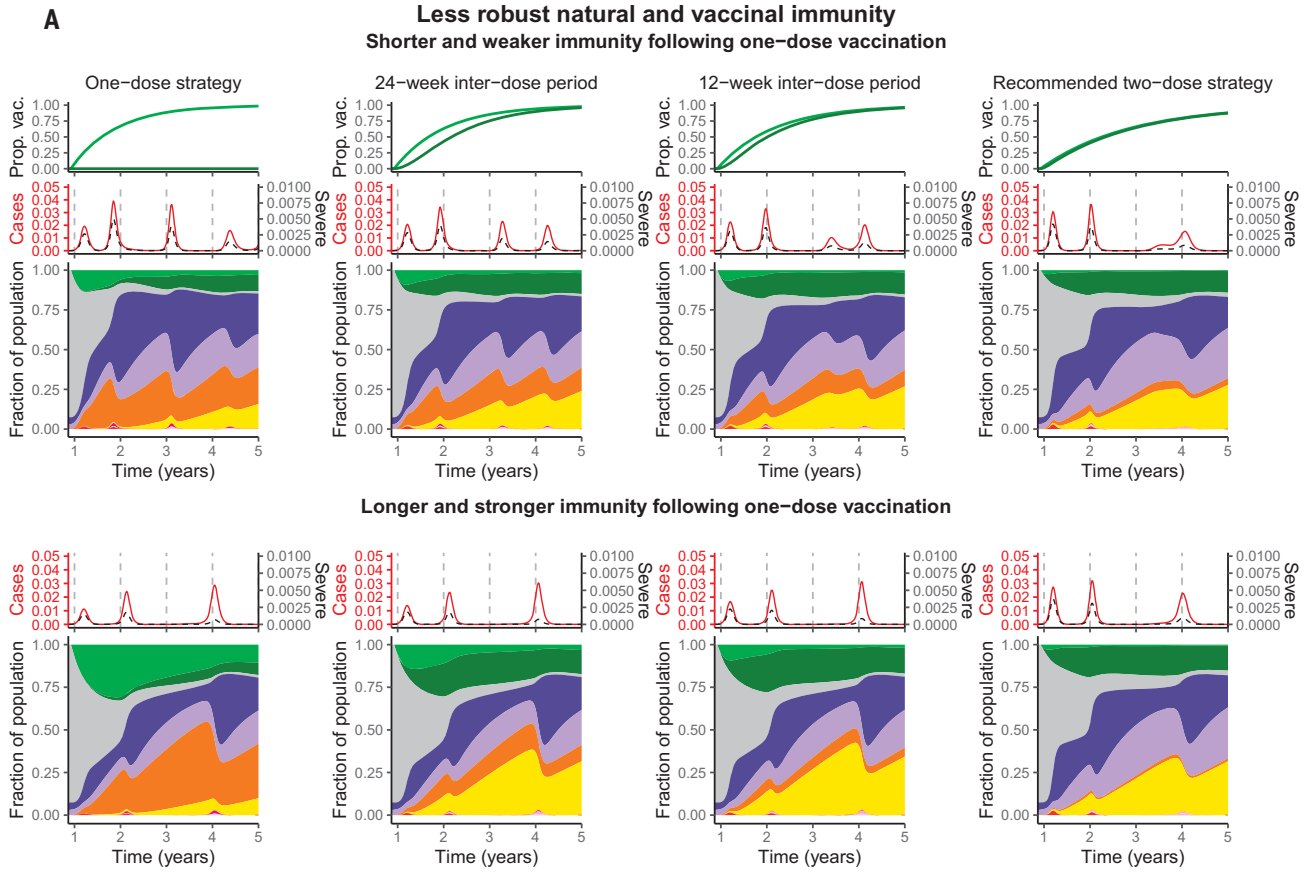
Vaccines will be central to efforts to attain community immunity (*22*) and thus prevent local spread due to case importation. We therefore analytically calculated the first vaccine dose administration rate for a given interdose spacing required for community immunity in our model (see supplementary materials). In the long term, however, in countries with adequate supplies, individuals whose one- or two-dose immunity has waned will likely be able to get vaccinated again before reinfection; we therefore incorporated revaccination of these individuals into the extended model and computed an analogous minimal vaccination rate, which we plot in Fig. 3B. We find that as the interdose period grows, this minimal rate depends increasingly on the degree of reduction in susceptibility after the waning of one-dose vaccinal immunity ϵ_1 (Fig. 3B and see fig. S13 for other parameter choices). Vaccine refusal (*23*) may also affect the attainment of community immunity through vaccinal immunity in the longer term (see supplementary materials).

Evolutionary impacts

The recent emergence of numerous SARS-CoV-2 variants in still relatively susceptible populations underlines the virus's evolutionary potential (*24*–*26*). We focus here on the longer-term potential for immune escape from natural or vaccinal immunity (*17*). For immune escape variants to spread within a population, they must first arise via mutation, and then there must be substantial selection pressure in their favor. We expect the greatest opportunity for variants to arise in (and spread from) hosts with the highest viral loads, likely those with the least immunity. On the other hand, we expect the greatest selection for escape where immunity is strongest. Previous research on the phylodynamic interaction between viral epidemiology and evolution (based on seasonal influenza) predicts that partially immune individuals (permitting intermediate levels of selection and transmission) could maximize levels of escape (*17*) (Fig. 4A). Under this model, we would project that different categories of secondarily infected people (after waning of natural immunity or immunity conferred by one or two doses of vaccine) would be key potential contributors to viral immune escape.

In Fig. 4, we consider three potential evolutionary scenarios, exploring different assumptions regarding viral abundance and within-host selection for the various immune classes. In all scenarios, we assume, for simplicity, that immunity elicited after two doses of the vaccine is equivalent to that elicited after natural infection. We also assume that transmission rises with viral abundance in hosts (*17*). In scenario I (black-bordered circles, top panel of Fig. 4A), we assume that infections of all classes of partially susceptible individuals lead to strong selective pressures and low viral abundance (a marker of low transmission) and thus low rates of adaptation, with only slightly reduced immune pressure for infections after a waned single vaccine dose relative to natural infection or two doses. Scenario II (blue-bordered circles, middle panel of Fig. 4A) considers a situation where natural and two-dose vaccinal immunity again lead to low viral abundance but where one-dose vaccinal immunity is associated with intermediate immune pressure that results in substantially higher rates of viral adaptation. Finally, in scenario III (purple-bordered circles, bottom panel of Fig. 4A), adaptive immune responses after waned natural, one-dose, and two-dose vaccinal immunity all lead to similar intermediate levels of immune pressure and high rates of viral adaptation. In all cases, we assume, for tractability, that viral immune escape is not correlated with clinical severity (*27*).

The relative potential viral adaptation rates [see (*17*) for more details] corresponding to each scenario are presented in the top rows of Fig. 4, B and C. This relative rate is estimated as the sum of the sizes of the infection classes after waned immunity (i.e., I_S after S_S, I_{S_1} after S_{S_1}, and I_{S_2} after S_{S_2}) weighted by the infection class–specific net viral adaptation rate assigned in each scenario. Therefore, this quantity reflects a weight-averaged potential rate for viral adaptation per individual per infection. The corresponding immune and susceptibility classes are plotted in the middle and bottom rows, respectively, according to the color scheme defined in Fig. 1A. The weaker immunity scenario of Fig. 2A is considered, with Fig. 4, B and C, corresponding, respectively, to the situations of a weaker and more robust single vaccine dose relative to two doses. The leftmost column corresponds to a one-dose strategy, an interdose period of $\frac{1}{\omega} = 24$ weeks is assumed in the middle column, and the rightmost column assumes a two-dose strategy with doses separated by the clinical trial window of $\frac{1}{\omega} = 4$ weeks.


A
Less robust natural and vaccinal immunity
Shorter and weaker immunity following one−dose vaccination
One−dose strategy
24−week inter−dose period
12−week inter−dose period
Recommended two−dose strategy
Prop. vac.
Cases
Severe
Fraction of population
Time (years)
Longer and stronger immunity following one−dose vaccination

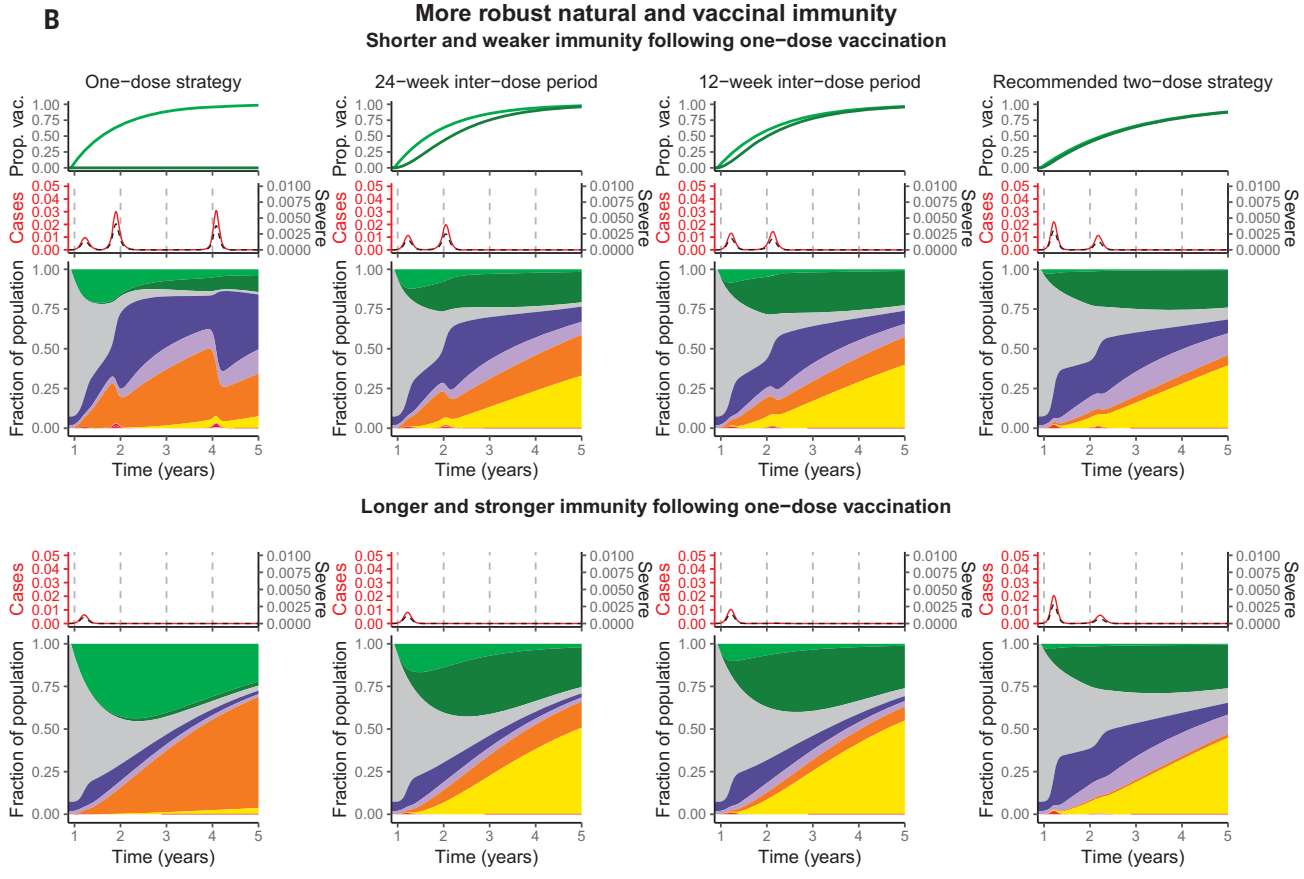

B
More robust natural and vaccinal immunity
Shorter and weaker immunity following one−dose vaccination
One−dose strategy
24−week inter−dose period
12−week inter−dose period
Recommended two−dose strategy
Prop. vac.
Cases
Severe
Fraction of population
Time (years)
Longer and stronger immunity following one−dose vaccination

Fig. 2. Synoptic medium-term immune landscapes and infection burden. The immune and infection class colors are the same as in Fig. 1A. Each panel shows the following: (Top) Illustrative time series of the fraction of the population vaccinated with one or two doses [see (*56*)]. (Middle) The fraction of total and severe infections [see (*57*)]. (Bottom) Area plots of the fraction of the population that makes up each immune class (S_P, R, S_S, V_1, V_2, S_{S_1}, S_{S_2}) or infection class (I_P, I_S, I_V, I_{S_1}, I_{S_2}) from just before the introduction of vaccination until 5 years after onset of the pandemic. In all plots, the maximum rate of administration of the first vaccine dose is taken to be ν_0 = 2%, and the vaccine is introduced at t_{vax} = 48 weeks. We take $\epsilon_{V_1} = 0.1$ and $\epsilon_{V_2} = 0.05$, in keeping with data from clinical trials (*3*). The fraction of severe cases for primary infections, secondary infections, infection after vaccination, and infection after waned two-dose immunity are taken to be $x_{sev,p} = 0.14$, $x_{sev,s} = 0.07$, $x_{sev,V} = 0.14$, and $x_{sev,2} = 0$, respectively. The transmission rates and periods of NPI adoption are defined in the materials and methods. The leftmost column corresponds to a one-dose vaccine strategy (ω = 0), followed by interdose spacings of 24 weeks, 12 weeks, and 4 weeks (rightmost column). (**A**) An overall more pessimistic natural and vaccinal immunity scenario, with $\epsilon = \epsilon_2 = 0.7$ and $1/\delta = 1/\rho_2 = 1$ year. For a less effective one-dose vaccine (top section), we take $\epsilon_1 = 0.9$ and $1/\rho_1 = 0.25$ years, and the fraction of severe cases associated with infection after waned one-dose immunity is $x_{sev,1} = 0.14$. For an effective one-dose vaccine (bottom section), we take $\epsilon_1 = 0.7$ and $1/\rho_1 = 1$ year, and the fraction of severe cases associated with infection after waned one-dose immunity is $x_{sev,1} = 0$. (**B**) An overall more optimistic natural and vaccinal immunity scenario, with $\epsilon = \epsilon_2 = 0.5$ and $1/\delta = 1/\rho_2 = 2$ years. For a less effective one-dose vaccine (top section), we take $\epsilon_1 = 0.9$ and $1/\rho_1 = 0.5$ years, and the fraction of severe cases associated with infection after waned one-dose immunity is $x_{sev,1} = 0.14$. For an effective one-dose vaccine (bottom section), we take $\epsilon_1 = 0.5$ and $1/\rho_1 = 2$ years, and the fraction of severe cases associated with infection after waned one-dose immunity is $x_{sev,1} = 0$.

Different assumptions regarding the strength and duration of adaptive immune responses to vaccines and natural infections alter projections for the proportions of individuals in the partially susceptible immune classes over time. When one-dose vaccinal immunity is poor, a one-dose strategy results in the rapid accumulation of partially susceptible S_{S_1} individuals (Fig. 4B, bottom row) and a greater infection burden. (Note, this S_{S_1} immune class is highlighted in orange for visibility in Figs. 1, 2, and 4.) When the assumed individual rates of evolutionary adaptation arising from these infection classes are high (scenarios II and III), we find that a one-dose strategy could lead to substantially higher relative rates of adaptation. This effect can be mitigated by implementing a two-dose strategy even with a longer interdose period than the recommended duration, echoing our epidemiological findings.

A single-dose strategy of a strongly immunizing vaccine reduces infection rates, resulting in lower relative rates of adaptation when a one-dose strategy is used; however the resulting large fraction of S_{S_1} individuals may still lead to evolutionary pressure, particularly when the potential viral adaptation rate associated with I_{S_1} infections is large. A two-dose strategy mitigates this effect, but the corresponding reduction in vaccinated individuals increases the infection burden from other classes. Thus, our results highlight the importance of rapid vaccine deployment to avoid these potentially pessimistic evolutionary outcomes. More broadly, our results further underline the importance of equitable, global vaccination (*28*, *29*): Immune escape anywhere will quickly spread.

Impact of increasing vaccination over time

In the supplementary materials (figs. S10 to S12), we explore the implications of ramping up vaccine deployment through two approaches. First, we examine a simple increase in the rate of administration of the first dose and unchanged dosing regimes (fig. S10). Qualitatively, these results are largely analogous to our previous results and reflect the benefits of increasing population immunity through an increase in vaccination deployment.

However, as vaccines become more widely available, policies on dosing regimes may change. The second approach we consider is a timely shift to a two-dose policy with recommended interdose spacing as vaccine deployment capacity increases (figs. S11 and S12). Initially delaying (or omitting) the second dose decreases the first epidemic peak after the initiation of vaccination. Such a reduction in first peak size would also reduce secondary infections and thus potentially immune escape in most cases (i.e., an evolutionary advantage). Subsequently, the switch to a manufacturer-timed vaccine dosage regime mitigates the potential medium-term disadvantages of delaying (or omitting) the second dose that may arise if immunity conferred from a single dose is relatively poor, including the accumulation of partially susceptible S_{S_1} individuals whose one-dose vaccinal immunity has waned. These contrasts highlight the importance of data-driven policies that undergo constant reevaluation as vaccination progresses.

Caveats

Our immuno-epidemiological model makes several assumptions. Although heterogeneities (superspreading, age, space, etc.) (*30*–*33*) are important for the quantitative prediction of SARS-CoV-2 dynamics, we previously found that these do not qualitatively affect our results (*13*). Nevertheless, we again briefly explore the epidemiological consequences of heterogeneities in transmission and vaccine coverage in the supplementary materials. We have also assumed that the robustness of immune responses after the second dose is independent of the interdose period, yet it is possible that delaying the second dose may actually enhance adaptive immune responses (*34*). Detailed clinical evaluation of adaptive immune responses after one and two vaccine doses with different interdose spacing is an important direction for future work.

Additionally, we have assumed highly simplified scenarios for NPIs. The chosen scenario was selected to qualitatively capture current estimates of SARS-CoV-2 prevalence and seropositivity in large cities. However, these values vary substantially between locations, a notable example being recent estimates of a large infection rate in Manaus, Brazil, during the first wave (*35*) or countries having almost no infections, owing to the successful implementation of NPIs (*36*–*38*). We have examined these scenarios in the supplementary materials (figs. S5 and S6). The qualitative projections of our model are sensitive to the composition of infection and immune classes at the onset of vaccination (including, therefore, the assumption of drastically higher seropositivity levels, i.e., the sum of the S_S and R classes). We further explore this in the supplementary materials through the initiation of vaccination at different times in the dynamic cycle (figs. S7 and S8). Thorough explorations of various NPIs, seasonal transmission rate patterns, vaccine deployment rates, dosing regimes, and clinical burdens can be investigated for broad ranges of epidemiological and immunological parameters with the online interactive application (*39*).

Finally, we have explored the simplest evolutionary model, which can only give a general indication of the potential for evolution under different scenarios. Including more-complex evolutionary models (*40*, *41*) into our framework is thus another important area for future work. Population heterogeneities likely have complex impacts on viral evolution. First, heterogeneities in immune responses and transmission [e.g., chronically infected hosts that shed virus for extended periods (*42*), or focused versus polyclonal responses] may have important impacts on the accumulation of genetic diversity and the strength of selection pressures and hence on evolutionary potential [e.g., for influenza, see (*43*)]. Second, there are

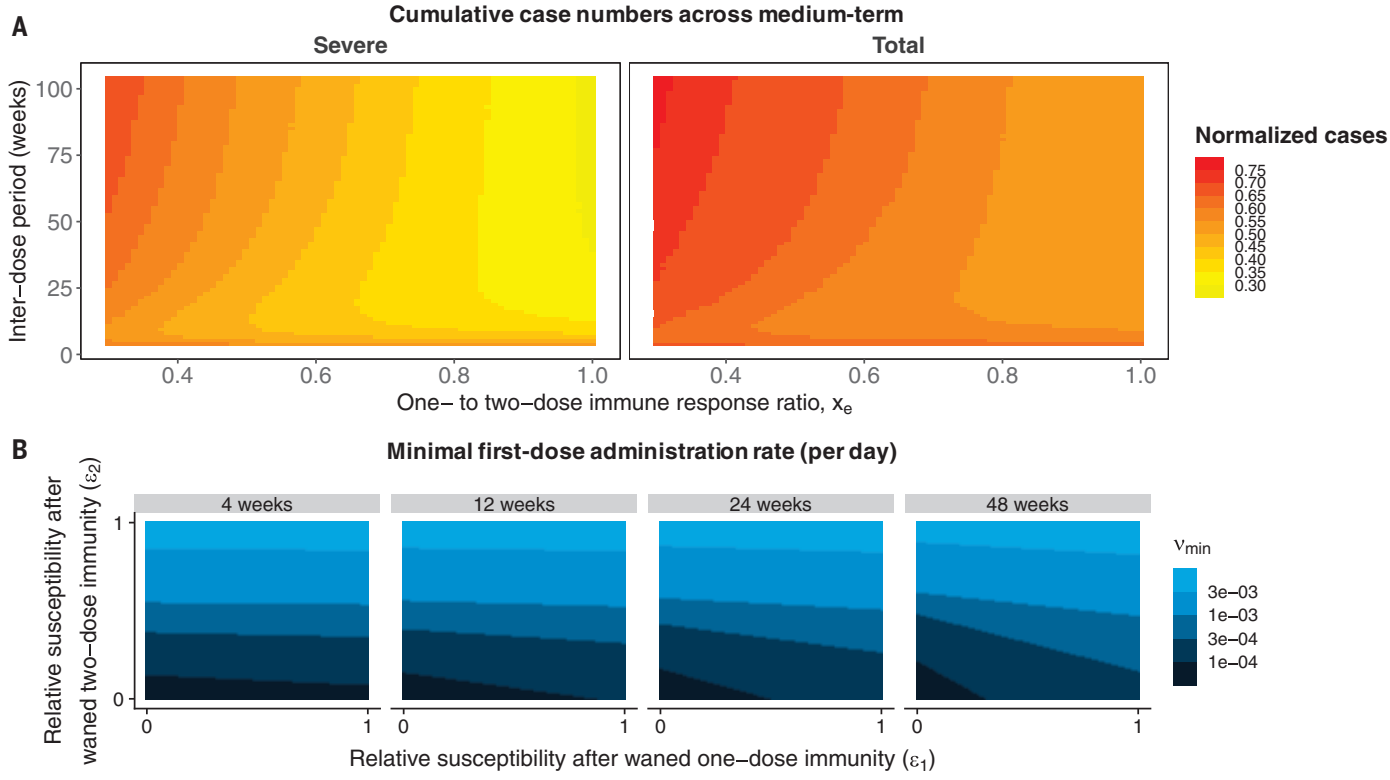


Fig. 3. Heatmaps depicting various epidemiological outcomes contingent on dosing regimes. (**A**) Cumulative severe (left) and total (right) case numbers relative to the scenario with no vaccine from the time of vaccine introduction through the end of the 5-year time period after onset of the pandemic as a function of the one-dose to two-dose immune response ratio x_e and the interdose period. Parameters correspond to the weak immunity scenario of Fig. 2A, but x_e sets the value of ϵ_1, ρ_1, and $x_{sev,1}$. Specifically, we take $\epsilon_1 = \epsilon_2 + (1 - x_e)(1 - \epsilon_2)$ such that the susceptibility to infection after a waned single dose interpolates linearly between the value after waned two doses (ϵ_2) when the one- and two-dose immune responses are equally strong ($x_e = 1$) and unity (full susceptibility) when a single dose offers no immune protection ($x_e = 0$). Similarly, we take $x_{sev,1} = x_{sev,2} + (1 - x_e)(x_{sev,V} - x_{sev,2})$, such that the fraction of severe cases for infections after a waned single dose interpolates linearly between the value after waned two doses ($x_{sev,2}$) when $x_e = 1$ and the value after a (failed) vaccination $x_{sev,V}$ when $x_e = 0$. Finally, ρ_1 is given by $\rho_1 = \rho_2/x_e$. (**B**) Values of ν_{min}, the minimal rate of first dose administration per day such that for any $\nu > \nu_{min}$ the basic reproduction $R_0[\nu] < 1$ and the disease cannot invade (see supplementary materials), as a function of the strength of immunity after one (ϵ_1) and two (ϵ_2) waned vaccines doses, for different interdose periods. We take the duration of one-dose and two-dose vaccinal immunity to be $1/\rho_1 = 0.5$ years and $1/\rho_2 = 1$ year, respectively, and set $\epsilon_{V_1} = 0.1$ and $\epsilon_{V_2} = 0.05$.

complex evolutionary implications of disease severity minimization by vaccination (*27*, *44*). Third, superspreading and contact structure could influence the rate of spread of novel variants through a population (*45*). Additionally, increases in viral avidity to the human angiotensin-converting enzyme 2 receptor might generate multiple benefits for the virus in terms of enhanced transmission and immune escape (*46*). Finally, genetic processes such as clonal interference, epistasis, and recombination also add substantial complexity to evolutionary dynamics [e.g., (*17*, *47*, *48*)]. Further model refinements should also include these details for increased accuracy. A full list of caveats is presented in the supplementary materials.

Conclusion

The deployment of SARS-CoV-2 vaccines in the coming months will strongly shape post-pandemic epidemiological trajectories and characteristics of accumulated population immunity. Dosing regimes should seek to navigate existing immunological and epidemiological trade-offs between individuals and populations. Using simple models, we have shown that different regimes may have crucial epidemiological and evolutionary impacts, resulting in a wide range of potential outcomes in the medium term. Our work also lays the foundation for a number of future considerations related to vaccine deployment during ongoing epidemics, especially preparing against future pandemics.

In line with intuition, spreading single doses in emergency settings (i.e., rising infections) is beneficial in the short term and reduces prevalence. Furthermore, we find that if immunity after a single dose is robust, then delaying the second dose is also optimal from an epidemiological perspective in the longer term. On the other hand, if one-dose vaccinal immunity is weak, the outcome could be more pessimistic; specifically, a vaccine strategy with a very long interdose period could lead to marginal short-term benefits (a decrease in the short-term burden) at the cost of a higher infection burden in the long term and substantially more potential for viral evolution. These negative longer-term effects may be alleviated by the eventual administration of a second dose, even if it is moderately delayed. With additional knowledge of the relative strength and duration of one-dose vaccinal immunity and corresponding clinically informed policies related to dosing regimes, pessimistic scenarios may be avoided. For context, at the time of writing, the UK, for example, has been particularly successful in rolling out vaccination to a large population with a wide spacing between doses (*49*). Our model illustrates that, ultimately, the long-term impacts of this strategy, especially in terms of transmission and

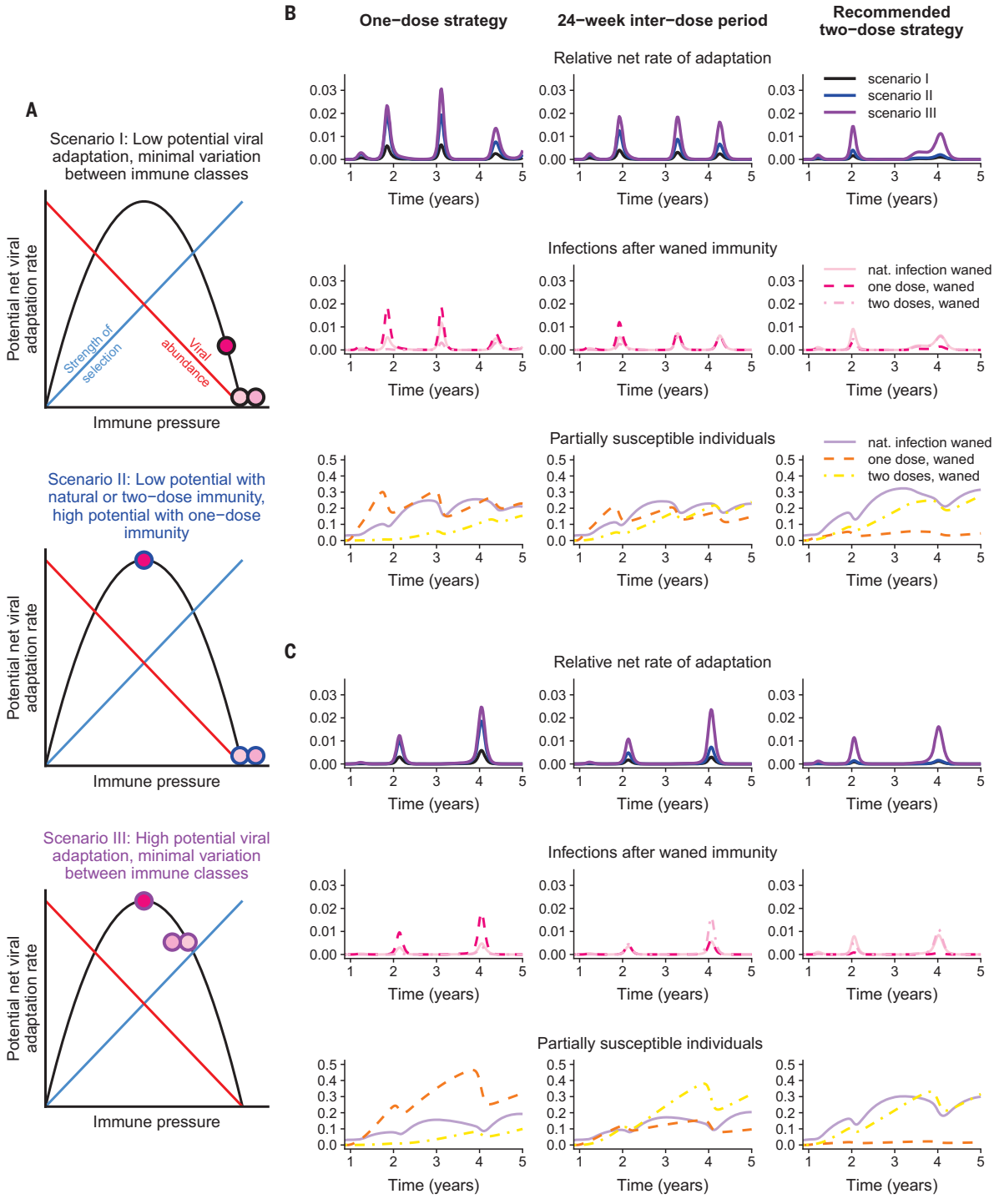


Fig. 4. Potential viral evolution scenarios under different vaccine regimes. (**A**) Schematic representations of the potential net viral adaptation rate associated with the I_S, I_{S_1}, and I_{S_2} infection classes under three different scenarios. These are illustrated by the filled circles, with the inside color denoting the infection class and corresponding to the legend in Fig. 1A. The circle borders correspond to the three scenarios considered (scenario I: black, top panel; scenario II: blue, middle panel; and scenario III: purple, bottom panel). The phylodynamic model for potential viral adaptation as a function of immune pressure is adapted from (*17*). (**B** and **C**) Relative net rates of adaptation [top rows; colors correspond to the scenarios in (A)] and composition of associated infection classes (middle rows; I_S, solid lines; I_{S_1}, dashed lines; I_{S_2}, dashed-dotted lines) and susceptible classes (bottom rows; S_S, solid lines; S_{S_1}, dashed lines; S_{S_2}, dashed-dotted lines). The colors in the middle and bottom rows correspond to the legend in Fig. 1A. The leftmost column corresponds to a one-dose strategy, an interdose period of $\frac{1}{\omega} = 24$ weeks is assumed in the middle column, and the rightmost column assumes a two-dose strategy with doses separated by the recommended window of $\frac{1}{\omega} = 4$ weeks. Both (B) and (C) correspond to a weak natural and vaccinal immunity scenario, with the same parameters as those in Fig. 2A. A weaker immune response after one vaccine dose is assumed in (B) (with parameters corresponding to those in the top section of Fig. 2A), and a stronger immune response after one vaccine dose is assumed in (C) (with parameters corresponding to those in the bottom section of Fig. 2A). The weights used to calculate the relative net rates of adaptation are $w_{IS,\mathrm{I}} = 0.05$, $w_{IS_1,\mathrm{I}} = 0.3$, and $w_{IS_2,\mathrm{I}} = 0.05$ in scenario I; $w_{IS,\mathrm{II}} = 0.05$, $w_{IS_1,\mathrm{II}} = 1$, and $w_{IS_2,\mathrm{II}} = 0.05$ in scenario II; and $w_{IS,\mathrm{III}} = 0.8$, $w_{IS_1,\mathrm{III}} = 1$, and $w_{IS_2,\mathrm{III}} = 0.8$ in scenario III.

immune escape, will depend on the duration and strength of one-dose vaccinal immunity. Recent evidence of weaker vaccinal immunity against the B.1.351 strain (*50*) underlines the importance of both detecting novel strains and titrating the strength of natural and vaccinal immunity against them.

Our results stress the negative epidemiological and evolutionary impacts that may emerge in places where vaccine deployment is delayed and vaccination rates are low. And because these consequences (e.g., the evolution of new variants) could emerge as global problems, this underlines the urgent need for global equity in vaccine distribution and deployment (*28*, *29*).

Current uncertainties surrounding the strength and duration of adaptive immunity in response to natural infection or vaccination lead to very broad ranges for the possible outcomes of various dosing regimes. Nevertheless, ongoing elevated COVID-19 case numbers stress the urgent need for effective mass vaccine deployment. Overall, our work emphasizes that the impact of vaccine dosing regimes is strongly dependent on the relative robustness of immunity conferred by a single dose. It is therefore imperative to determine the strength and duration of clinical protection and transmission-blocking immunity through careful clinical evaluations (including, for instance, randomized control trials of dose intervals and regular testing of viral loads in vaccinated individuals, their contacts, and those who have recovered from natural infections) in order to enforce sound public policies. More broadly, our results underscore the importance of exploring the phylodynamic interaction of pathogen dynamics and evolution, from within-host to global scales, for SARS-CoV-2, influenza, and other important pathogens (*40*, *41*, *47*, *48*, *51*, *52*).

REFERENCES AND NOTES

1. McGill COVID19 Vaccine Tracker Team, COVID19 Vaccine Tracker; https://covid19.trackvaccines.org/vaccines/.
2. L. R. Baden *et al*., *N. Engl. J. Med.* **384**, 403–416 (2021).
3. F. P. Polack *et al*., *N. Engl. J. Med.* **383**, 2603–2615 (2020).
4. M. Voysey *et al*., *Lancet* **397**, 99–111 (2021).
5. G. Iacobucci, E. Mahase, *BMJ* **372**, n18 (2021).
6. "Quebec opts to delay 2nd dose of vaccine in order to immunize health-care workers faster," CBC News, 5 January 2021; www.cbc.ca/news/canada/montreal/quebec-second-vaccine-dose-delays-1.5861194.
7. M. Rabson, "Provinces delaying or revisiting vaccine programs as Pfizer slows dose deliveries," CTV News, 18 January 2021; www.ctvnews.ca/health/coronavirus/provinces-delaying-or-revisiting-vaccine-programs-as-pfizer-slows-dose-deliveries-1.5271829.
8. G. Chodcik *et al*., The effectiveness of the first dose of BNT162b2 vaccine in reducing SARS-CoV-2 infection 13-24 days after immunization: real-world evidence. medRxiv 2021.01.27.21250612 [Preprint]. 29 January 2021. https://doi.org/10.1101/2021.01.27.21250612.
9. V. J. Hall *et al*., Effectiveness of BNT162b2 mRNA vaccine against infection and COVID-19 vaccine coverage in healthcare workers in England, multicentre prospective cohort study (the SIREN Study). SSRN 3790399 [Preprint]. 22 February 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3790399.
10. J. M. Dan *et al*., *Science* **371**, eabf4063 (2021).
11. S. F. Lumley *et al*., Antibodies to SARS-CoV-2 are associated with protection against reinfection. medRxiv 2020.11.18.20234369 [Preprint]. 19 November 2020. https://doi.org/10.1101/2020.11.18.20234369.
12. A. G. Letizia *et al*., *N. Engl. J. Med.* **383**, 2407–2416 (2020).
13. C. M. Saad-Roy *et al*., *Science* **370**, 811–818 (2020).
14. E. Volz *et al*., *Cell* **184**, 64–75.e11 (2021).
15. A. M. Gravagnuolo *et al*., Epidemiological investigation of new SARS-CoV-2 variant of concern 202012/01 in England. medRxiv 2021.01.14.21249386 [Preprint]. 15 January 2021. https://doi.org/10.1101/2021.01.14.21249386.
16. N. G. Davies *et al*., Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. medRxiv 2020.12.24.20248822 [Preprint]. 26 December 2020. https://doi.org/10.1101/2020.12.24.20248822.
17. B. T. Grenfell *et al*., *Science* **303**, 327–332 (2004).
18. R. Eguia *et al*., A human coronavirus evolves antigenically to escape antibody immunity. bioRxiv 2020.12.17.423313 [Preprint]. 18 December 2020. https://doi.org/10.1101/2020.12.17.423313.
19. S. R. Kadire, R. M. Wachter, N. Lurie, *N. Engl. J. Med.* **384**, e28 (2021).
20. S. E. Morris *et al*., *J. R. Soc. Interface* **12**, 20141245 (2015).
21. R. E. Baker, W. Yang, G. A. Vecchi, C. J. E. Metcalf, B. T. Grenfell, *Science* **369**, 315–319 (2020).
22. C. M. Saad-Roy, S. A. Levin, C. J. E. Metcalf, B. T. Grenfell, *J. R. Soc. Interface* **18**, 20200683 (2021).
23. C. E. Wagner *et al*., *Front. Public Health* **8**, 614113 (2020).
24. H. Tegally *et al*., Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv 2020.12.21.20248640 [Preprint]. 22 December 2020. https://doi.org/10.1101/2020.12.21.20248640.
25. N. R. Faria *et al*., "Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings," Virological.org (2021); https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.
26. A. Rambaut *et al*., "Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations," Virological.org (2020); https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.
27. I. F. Miller, C. J. E. Metcalf, No current evidence for risk of vaccine-driven virulence evolution in SARS-CoV-2. medRxiv 2020.12.01.20241836 [Preprint]. 3 December 2020. https://doi.org/10.1101/2020.12.01.20241836.
28. E. J. Emanuel *et al*., *Science* **369**, 1309–1312 (2020).
29. T. Burki, *Lancet Infect. Dis.* **21**, 33–34 (2021).
30. J. S. Lavine, O. N. Bjornstad, R. Antia, *Science* **371**, 741–745 (2021).
31. K. Sun *et al*., *Science* **371**, eabe2424 (2021).
32. R. Laxminarayan *et al*., *Science* **370**, 691–697 (2020).
33. D. C. Adam *et al*., *Nat. Med.* **26**, 1714–1719 (2020).
34. M. Voysey *et al*., Single dose administration, and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine. SSRN 3777268 [Preprint]. 1 February 2021. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3777268.
35. L. F. Buss *et al*., *Science* **371**, 288–292 (2021).
36. M. G. Baker, N. Wilson, A. Anglemyer, *N. Engl. J. Med.* **383**, e56 (2020).
37. D. J. Summers *et al*., *Lancet Reg. Health West. Pac.* **4**, 100044 (2020).
38. COVID-19 National Incident Room Surveillance Team, *Commun. Dis. Intell.* **44**, 10.33321/cdi.2020.44.92 (2020).
39. Epidemiological and evolutionary dynamics of SARS-CoV-2 interactive dashboard, http://grenfelllab.princeton.edu/sarscov2vaccine.
40. N. M. Ferguson, A. P. Galvani, R. M. Bush, *Nature* **422**, 428–433 (2003).
41. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, *Science* **314**, 1898–1903 (2006).
42. B. Choi *et al*., *N. Engl. J. Med.* **383**, 2291–2293 (2020).
43. D. H. Morris *et al*., *eLife* **9**, e62105 (2020).
44. J. R. Gog, E. M. Hill, L. Danon, R. Thompson, Vaccine escape in a heterogeneous population: insights for SARS-CoV-2 from a simple model. medRxiv 2021.03.14.21253544 [Preprint]. 28 March 2021. https://doi.org/10.1101/2021.03.14.21253544.
45. K. M. Pepin, I. Volkov, J. R. Banavar, C. O. Wilke, B. T. Grenfell, *J. Theor. Biol.* **265**, 501–510 (2010).
46. J. Zahradn *et al*., SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. bioRxiv 2021.01.06.425392 [Preprint]. 6 January 2021. https://doi.org/10.1101/2021.01.06.425392.
47. E. M. Volz, K. Koelle, T. Bedford, *PLOS Comput. Biol.* **9**, e1002947 (2013).
48. R. Biek, O. G. Pybus, J. O. Lloyd-Smith, X. Didelot, *Trends Ecol. Evol.* **30**, 306–313 (2015).
49. C. Baraniuk, *BMJ* **372**, n421 (2021).
50. S. A. Madhi *et al*., Safety and efficacy of the ChAdOx1 nCoV-19 (AZD1222) Covid-19 vaccine against the B.1.351 variant in South Africa. medRxiv 2021.02.10.21251247 [Preprint]. 12 February 2021. https://doi.org/10.1101/2021.02.10.21251247.
51. C. M. Saad-Roy, A. B. McDermott, B. T. Grenfell, *J. Infect. Dis.* **219** (suppl. 1), S46–S56 (2019).
52. E. M. Volz, S. L. Kosakovsky Pond, M. J. Ward, A. J. Leigh Brown, S. D. W. Frost, *Genetics* **183**, 1421–1430 (2009).
53. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, *Science* **368**, 860–868 (2020).
54. J. Zhang *et al*., *Lancet Infect. Dis.* **20**, 793–802 (2020).
55. R. E. Baker *et al*., *Nat. Commun.* **10**, 5512 (2019).

56. We qualitatively describe the fraction of the population that has received vaccines by assuming that vaccination occurs at random in the population. Specifically, suppose vaccination occurs at time $\tau = 0$. The rate of change of the fraction not vaccinated follows $\frac{dN}{d\tau} = -\nu N$, $N(0) = 1$, giving $N(\tau) = e^{-\nu\tau}$ so that the fraction vaccinated with one dose $Y(\tau) = 1 - N(\tau) = 1 - e^{-\nu\tau}$. The fractions of those vaccinated with one but not two doses follows $\frac{dW_1}{d\tau} = \nu N - \omega W_1$, giving $W_1(\tau) = \frac{\nu(e^{-\nu\tau} - e^{-\omega\tau})}{\omega - \nu}$. Then, the fraction vaccinated with two doses is $W_2(\tau) = Y(\tau) - W_1(\tau) = 1 - \frac{\omega}{\omega-\nu}e^{-\nu\tau} + \frac{\nu}{\omega-\nu}e^{-\omega\tau}$.
57. We calculate the total number of cases at any time point as $I_T = I_P + I_S + I_V + I_{S_1} + I_{S_2}$. Similarly, the number of severe cases is given by $I_{T,sev} = x_{sev,P}I_P + x_{sev,S}I_S + x_{sev,V}I_V + x_{sev,1}I_{S_1} + x_{sev,2}I_{S_2}$. Cumulative case numbers for a given time period are calculated through $\gamma\sum I_T$ (total cases) and $\gamma\sum I_{T,sev}$ (severe cases), where the summation occurs over all time steps.

58. C. M. Saad-Roy *et al*., Code for Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes, Version 1, Zenodo (2021); https://doi.org/10.5281/zenodo.4556111.

ACKNOWLEDGMENTS

Funding: This work was funded in part by Open Philanthropy; the Natural Sciences and Engineering Research Council of Canada through a post-graduate doctoral scholarship (C.M.S.-R.); the Cooperative Institute for Modeling the Earth System (CIMES) (R.E.B.); the James S. McDonnell Foundation 21st Century Science Initiative Collaborative Award in Understanding Dynamic and Multi-scale Systems (C.M.S.-R., S.A.L.); the C3.ai Digital Transformation Institute and Microsoft Corporation (S.A.L.); a gift from Google, LLC (S.A.L.); the National Science Foundation (CNS-2027908, CCF1917819) (S.A.L.); the U.S. CDC (B.T.G.); and Flu Lab (B.T.G.).
Author contributions: C.M.S.-R., C.E.W., and B.T.G. designed the study. C.M.S.-R. and C.E.W. performed the simulations and analyses and wrote the manuscript. S.E.M. developed the Shiny application. All authors contributed to interpreting the results and editing the manuscript. **Competing interests:** The authors have no competing interests. **Data and materials availability:** This manuscript contains no new data, and all referenced datasets are publicly available. Code for all analyses is available at Zenodo (*58*).

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

Materials and methods and references cited therein can be found in the main article online.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/363/suppl/DC1
Supplementary Text
Figs. S1 to S14
References (*59*–*64*)
MDAR Reproducibility Checklist

1 February 2021; accepted 4 March 2021
Published online 9 March 2021
10.1126/science.abg8663

EPIGENOME

Replication timing maintains the global epigenetic state in human cells

Kyle N. Klein[1]†, Peiyao A. Zhao[1]†, Xiaowen Lyu[2,3]†, Takayo Sasaki[1,4], Daniel A. Bartlett[1], Amar M. Singh[5], Ipek Tasan[6], Meng Zhang[6], Lotte P. Watts[7], Shin-ichiro Hiraga[7], Toyoaki Natsume[8,9], Xuemeng Zhou[10], Timour Baslan[11], Danny Leung[10], Masato T. Kanemaki[8,9], Anne D. Donaldson[7], Huimin Zhao[6], Stephen Dalton[5], Victor G. Corces[2], David M. Gilbert[1,4]*

The temporal order of DNA replication [replication timing (RT)] is correlated with chromatin modifications and three-dimensional genome architecture; however, causal links have not been established, largely because of an inability to manipulate the global RT program. We show that loss of RIF1 causes near-complete elimination of the RT program by increasing heterogeneity between individual cells. RT changes are coupled with widespread alterations in chromatin modifications and genome compartmentalization. Conditional depletion of RIF1 causes replication-dependent disruption of histone modifications and alterations in genome architecture. These effects were magnified with successive cycles of altered RT. These results support models in which the timing of chromatin replication and thus assembly plays a key role in maintaining the global epigenetic state.

DNA is replicated during S phase of the cell cycle in a temporal order known as the replication timing (RT) program. RT is conserved among eukaryotes, is developmentally controlled, and correlates with many important epigenomic features (*1*). Early-replicating chromatin generally contains active histone modifications, is located in the nuclear interior, and correlates with the A compartment defined by high-throughput chromosome conformation capture (Hi-C) (*2*), whereas late-replicating chromatin is associated with transcriptionally repressive histone modifications, localization at the nuclear periphery, and the B compartment. Histone modifications are both recycled from parental chromatin and added de novo after passage of the replication fork, with different chromatin states showing differing dynamics of reassembly (*2*, *3*). Despite these close correlations, the mechanistic link between RT and the accurate maintenance of chromatin through cell cycles remains elusive. It has long been hypothesized that RT influences chromatin maintenance. Indeed, microinjection of plasmids into mammalian nuclei revealed that plasmids replicated in early S phase were decorated with acetylated histones, whereas those replicated later in S phase were devoid of acetylated histones (*4*, *5*). However, there is no direct evidence implicating RT in epigenetic state maintenance, largely because of the inability to manipulate genome-wide RT. The conserved protein RIF1 has been shown to affect RT in many eukaryotes; however, because the effects have been partial or localized, RIF1 disruption has not been exploited to study the effects of RT abrogation (*6*, *7*).

[1]Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA. [2]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA. [3]State Key Laboratory of Cellular Stress Biology, Fujian Provincial Key Laboratory of Reproductive Health Research, School of Medicine, Xiamen University, Xiamen, China. [4]San Diego Biomedical Research Institute, La Jolla, CA 92121, USA. [5]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA. [6]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [7]Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. [8]Department of Chromosome Science, National Institute of Genetics, Research Organization of Information and Systems (ROIS), Yata 1111, Mishima, Shizuoka 411-8540, Japan. [9]Department of Genetics, The Graduate University for Advanced Studies (SOKENDAI), Yata 1111, Mishima, Shizuoka 411-8540, Japan. [10]Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. [11]Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.
***Corresponding author. Email: gilbert@sdbri.org**
†These authors contributed equally to this work.

RIF1 knockout causes heterogeneous RT

To gain insight into the role of genome-wide RT in shaping the epigenome, we knocked out RIF1 (RIF1 KO) in H9 human embryonic stem cell (hESC), HCT116 (fig. S1, A to C), and HAP1 cell lines (supplementary methods). As previously reported (*8*), all three RIF1 KO cell lines proceeded through the cell cycle with nearly wild-type (WT) kinetics (fig. S1, D and E) and exhibited genome-wide aberrations in RT, albeit with varying degrees of severity (Fig. 1A). Similar to prior reports in mammalian cells (*6*), domains changed RT from either early to late (EtL) or late to early (LtE) in log2(E/L) RT profiles of HCT116 and HAP1 cells. RIF1 KO caused 43% of the genome to change RT in HCT116 cells (23% EtL and 20% LtE) and 39% of the genome to change in HAP1 cells (20% EtL and 19% LtE) (fig. S1F). However, in H9 hESCs (Fig. 1A), nearly the entire genome acquired a log2(E/L) close to zero (fig. S1F). RIF1 control of RT in hESCs was dosage dependent, because partial knockdown (KD) of RIF1 in H9 hESCs (fig. S2A) resulted in a partial effect on the RT program (fig. S2, B and C). Replication foci (*9*) in RIF1 KO cells showed a "blending" of early and middle spatial patterns (fig. S3). These results demonstrate a considerably more extensive role for RIF1 in RT control in hESCs than in other cell types.

Previous reports have interpreted RT changes in RIF1 KO cells as distinct RT switches (*6*), but the severity of the RT phenotype in RIF1 KO hESCs suggested an alternative mechanism. To address this, we performed high-resolution Repli-seq (*10*) (fig. S4A) in RIF1 KO cells, which identifies peaks of replication initiation termed initiation zones (IZs) and valleys of late replication containing broadly distributed, low-efficiency initiation events (*10*). We detected dramatic diffusion of RT patterns and loss of defined IZs in both HCT116 and H9 hESCs (Fig. 1B and fig. S4B), indicating major RT variation within the cell population (fig. S4, C and D). Most notably, there was a nearly complete genome-wide abrogation of the RT program in RIF1 KO H9 hESCs (Fig. 1B and fig. S4B). In RIF1 KO HCT116 cells, EtL and LtE regions called using E/L Repli-seq showed substantial loss of temporal control (fig. S4, E and F). Even early regions that were not called as EtL switches in E/L Repli-seq lost defined patterns (compare Fig. 1, A and B, with fig. S4, E and F), indicating that the entirety of the early-replicating genome lost RT control upon RIF1 KO in both cell lines. By contrast, many late-replicating regions in HCT116 retained late replication (fig. S4, E and F), indicating a RIF1-independent mechanism controlling RT for these regions. We next divided each IZ called in WT high-resolution Repli-seq into four S-phase timing classes (*10*) and plotted the cumulative percentage of DNA replicated through S phase (Fig. 1C). WT cells showed typical segregation of IZs according to the temporal order (*10*), whereas RIF1 KO cells showed major overlap of IZ classes and flatter sigmoidal-like curves (Fig. 1C). We subsequently calculated the genome-wide heterogeneity parameter T_{width}, which is positively correlated with heterogeneity. We found that genome-wide T_{width} was greatly increased in both cell types upon RIF1 KO (fig. S4G), indicating a significant increase in RT heterogeneity.

To directly validate increased cell-to-cell RT heterogeneity in RIF1 KO, we performed single-cell Repli-seq (*11*) on WT and RIF1 KO HAP1 cells. RIF1 KO cells showed co-replication of early and late domains (fig. S5) and had a notably larger T_{width} compared with WT cells (Fig. 1D), directly confirming that RIF1 KO disrupts RT by substantially increasing RT heterogeneity rather than causing discrete RT shifts in all cells.

RIF1 KO alters distribution of heterochromatic histone modifications

To map RIF1 binding, we performed Cut&Run (cleavage under targets and release using nuclease) against green fluorescent protein (GFP)

on GFP-tagged RIF1 in HCT116 and H9 hESCs (fig. S6A). RIF1 was enriched in the late-replicating portion of the genome in both cell lines (fig. S6, B and C) and bound chromatin in broad domains (Fig. 2A). In both cell lines, RIF1 binding was enriched at regions that lost RT control upon RIF1 KO (affected regions) (Fig. 2, A, B, and E; and fig. S6, D to F), whereas those regions that maintained their late RT (unaffected regions) had low RIF1 enrichment (Fig. 2, C, D, and E; and fig. S6, D and E). However, in RIF1 KO H9 hESCs, the regions defined as "unaffected regions" are rare and still display RT dysregulation but significantly less than "affected regions" (Fig. 2, compare D with B).

We next performed chromatin immunoprecipitation combined with sequencing (ChIP-seq) (spike-in) on the late replication–associated histone mark H3K9me3 (histone 3 lysine 9 trimethylation) in RIF1 KO cells (fig. S7). In WT cells, affected regions were enriched for smaller H3K9me3 peaks, whereas unaffected regions contained large H3K9me3 domains (Fig. 2, A, C, and E; and fig. S8, A and B). In HCT116, small H3K9me3 peaks at affected regions (n = 390) were lost (Fig. 2, A and E), and the large H3K9me3 domains at unaffected regions (n = 209) were strengthened (Fig. 2, C and E). H3K9me3 was globally depleted in H9 hESCs (fig. S8, C and D). The rare unaffected domains in RIF1 KO H9 hESCs (n = 49) were large H3K9me3 domains that were only slightly diminished for H3K9me3 compared with affected domains (Fig. 2E). Unaffected H3K9me3 domains did not overlap greatly between cell types (fig. S8E). H3K9me3 depletion in HCT116 RIF1 KO cells (fig. S8F) caused partial earlier replication of unaffected regions (Fig. 2F and fig. S8, G and H), indicating a role for H3K9me3 in their late replication but implicating additional unknown mechanisms [possibly association with the nuclear lamina (*8*)] in maintaining late RT at unaffected regions. By contrast, RIF1 KO caused cell type–specific changes to H3K27me3 (histone 3 lysine 27 trimethylation), with H9 hESCs primarily exhibiting down-regulation and HCT116 exhibiting overall up-regulation (fig. S9). These results reveal that multiple mechanisms, including RIF1 and large H3K9me3 domains, orchestrate late RT to different extents in different cell types.

Affected and unaffected late-replicating regions form separate interaction hubs

The strong association between late replication and B compartmentalization (*12*) compelled us to investigate the genomic compartmentalization of RIF1 KO cells by Hi-C. Notably, we found that affected and unaffected regions had distinct PC1 eigenvector associations despite being similarly late-replicating in WT cells, and this distinction was increased upon RIF1 KO (fig. S10A). WT interactions between unaffected regions were significantly stronger than interactions between affected regions or those between unaffected regions and affected regions (fig. S10B). Upon RIF1 KO, unaffected regions formed new interactions and strengthened existing interactions (Fig. 2G and fig. S10C). Strengthened interactions were correlated with up-regulated H3K9me3 domains (fig. S10D). We then sorted affected and unaffected regions according to the extent of H3K9me3 changes [negative (down-regulated) or positive (up-regulated)] and calculated their interaction frequencies. In WT cells, up-regulated and down-regulated H3K9me3 peaks formed separate interaction hubs (Fig. 2H, left

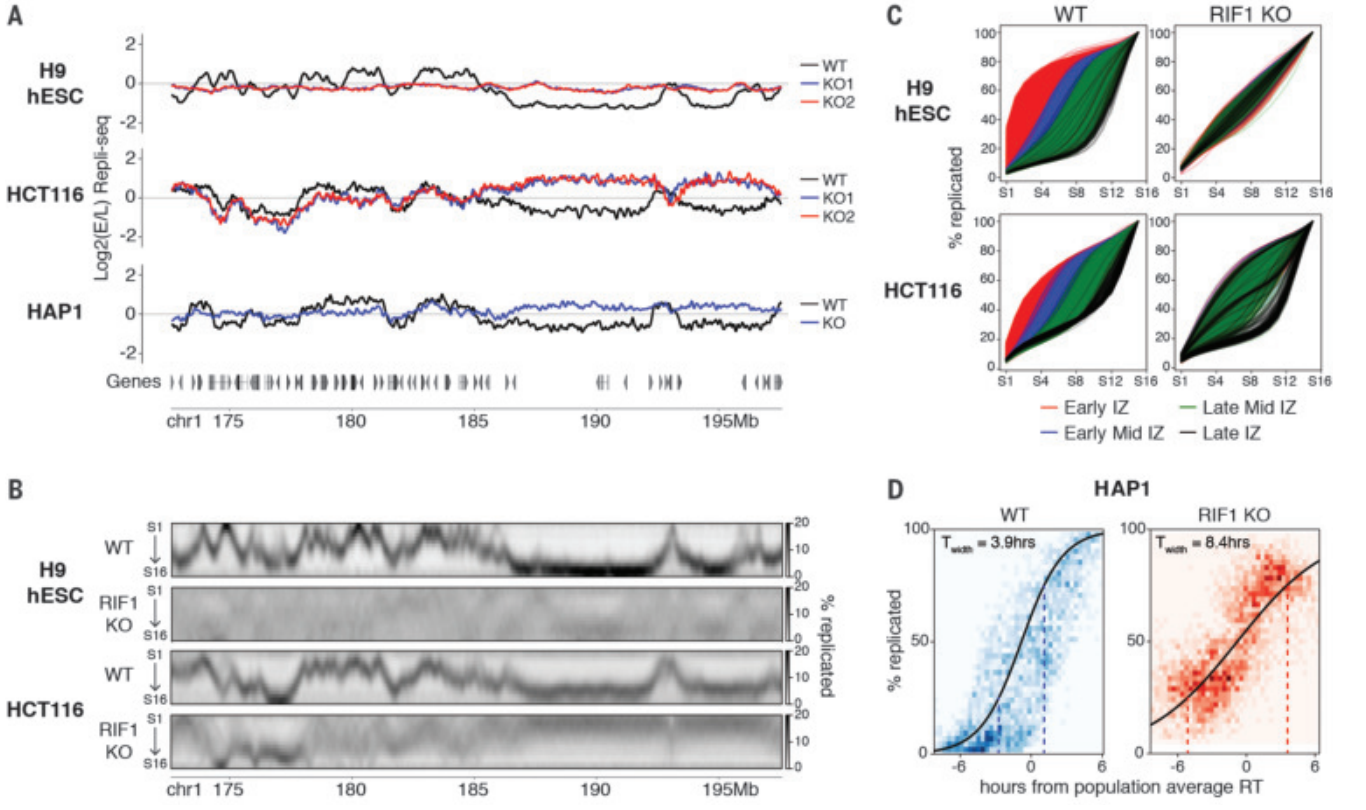


Fig. 1. RIF1 controls RT by reducing cell-to-cell variation in replication timing. (**A**) Log2(E/L) Repli-seq plots of chromosome 1 (Chr1), 172.6 to 197.6 Mb, in WT (black) and two RIF1 KO clones (blue and red) in H9 hESC (top), HCT116 (middle), and HAP1 (bottom) cell lines. (**B**) High-resolution Repli-seq plots of Chr1 172.6 to 197.6 Mb in WT and RIF1 KO in H9 hESCs (top two) and HCT116 (bottom two); same locus as in (A). (**C**) Cumulative percentage of replicated plots for each IZ called in WT cells versus S-phase fraction of 16-fraction Repli-seq, color-coded by their timing (red, early; blue, early middle; green, late middle; black, late). (**D**) Sigmoidal fitting of the percentage of cells having replicated in each bin (y axis) against time in hours from population average RT (x axis) for HAP1 WT (left) and RIF1 KO (right). The heatmaps (blue, WT; red, KO) represent the data spread for all 50-kb bins genome-wide in all single cells (51 WT and 69 KO). Dotted lines at 25% of cells replicated and 75% of cells replicated indicate the span of T_{width}.

column) that were strengthened and weakened, respectively, upon RIF1 KO (Fig. 2H, right column), accentuating the separation between the two. In individual cases, the juxtaposition of affected and unaffected domains created new compartment and topologically associating domain (TAD) boundaries (fig. S10, E and F). These results demonstrate that affected and unaffected H3K9me3 regions are two intrinsically different classes of late-replicating chromatin with distinct interaction preferences; unaffected domains form strong interactions with one another to form a compartment that maintains late RT without RIF1, whereas affected domains form a separate hub of

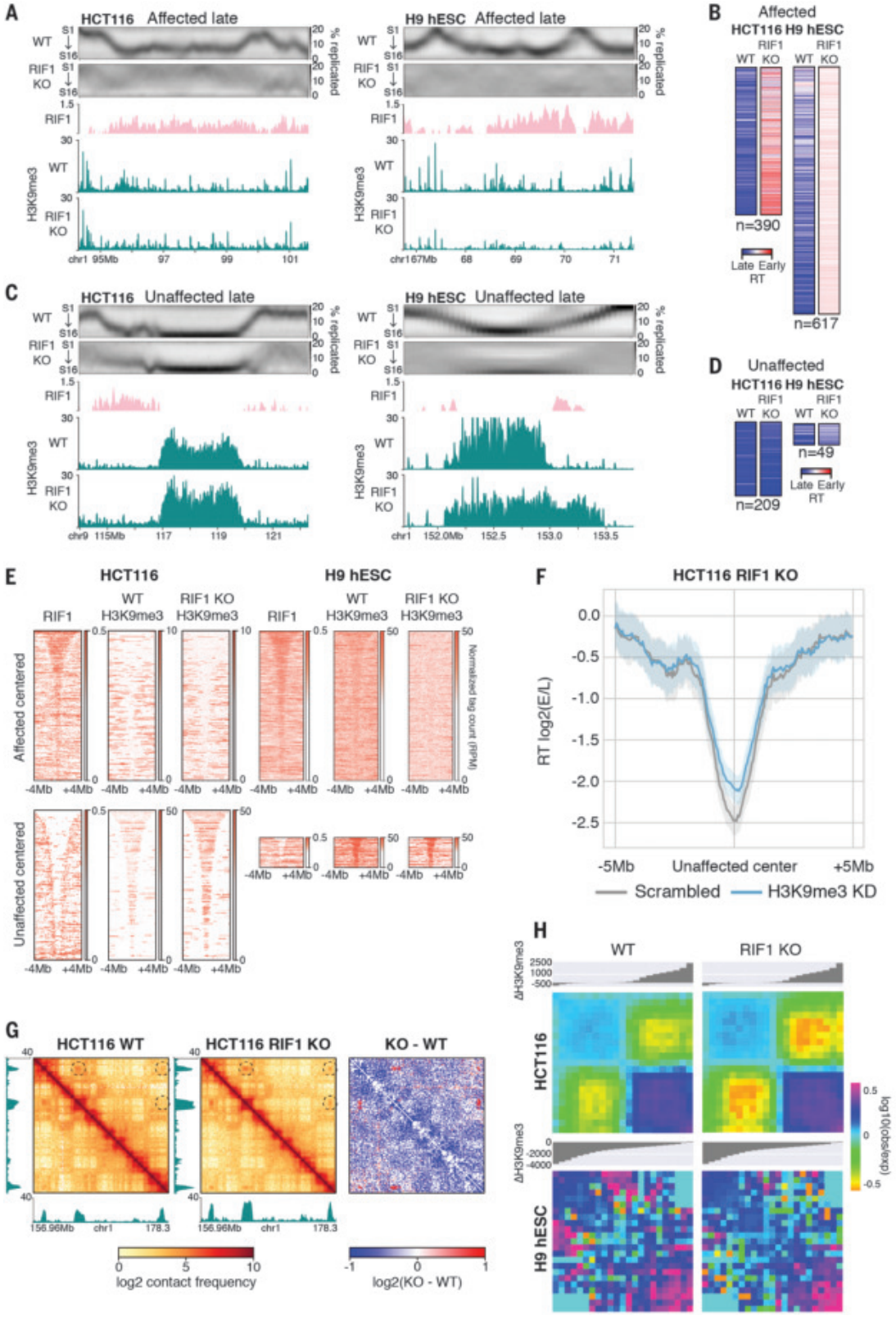


Fig. 2. RT affected and unaffected late regions are distinct classes of chromatin. (**A**) Affected late region at Chr1 94.3 to 101.6 Mb in HCT116 (left) and Chr1 66.75 to 71.4 Mb in H9 hESCs (right) showing, from top to bottom, high-resolution Repli-seq in WT and RIF1 KO cells, RIF1 fold enrichment in WT cells, and H3K9me3 ChIP-seq tag counts in WT and RIF1 KO cells. (**B**) Heatmaps of RT indices for affected late regions in WT and RIF1 KO of HCT116 and H9 hESCs. (**C**) Unaffected late region at Chr9 113.95 to 122.3 Mb in HCT116 (left) and Chr1 151.7 to 153.75 Mb in H9 hESCs (right), displayed as in (A). (**D**) Heatmaps of RT indices for unaffected late regions in WT and RIF1 KO of HCT116 and H9 hESCs. (**E**) RIF1 and H3K9me3 binding in HCT116 (left) and H9 hESC (right) WT or RIF1 KO cells centered on affected late regions (top) or unaffected late regions (bottom) ±4 Mb and sorted by size. (**F**) Mean log2(E/L) Repli-seq scores centered at all unaffected late regions ±5 Mb in HCT116 RIF1 KO transfected with scrambled control (gray) or H3K9me3 KD (blue) small interfering RNA. Shadow represents 95% confidence interval. (**G**) H3K9me3 tag count tracks beside ICE (iterative correction and eigenvector decomposition)–normalized Hi-C and subtraction contact maps of HCT116 WT and RIF1 KO at Chr1 156.95 to 178.3 Mb. Dotted circles denote regions of increased interaction. (**H**) Log10(observed/expected) aggregate interactions between late regions in WT and RIF1 KO HCT116 (top) and H9 hESCs (bottom). The interactions were binned into 11 equal segments, which were ranked by increasing ΔH3K9me3, where negative and positive values indicate, respectively, decreased and increased H3K9me3 in RIF1 KO compared with WT.

interactions and require RIF1 to enforce late replication.

RIF1 KO causes depleted active histone modifications and attenuated A compartment interactions

ChIP-seq of H3K27ac (histone 3 lysine 27 acetylation) (spike-in) and H3K4me3 (histone 3 lysine 4 trimethylation) revealed that both marks became depleted in both RIF1 KO cells (figs. S7 and S11, A to C). Loss of active marks from the A compartment was concurrent with changes in compartmentalization and Hi-C interactions in both cell lines. In H9 hESCs, the majority of compartment changes involved the disappearance of A compartment chromatin into neighboring B compartments (Fig. 3, A and B; and fig. S11D), whereas HCT116 exhibited discrete shifts in compartmentalization in both directions (Fig. 3A and fig. S11, D and E). Loss of A compartment interactions in Hi-C contact maps corresponded to reduced H3K27ac peaks in both cell lines [Fig. 3C inset (arrows) and fig. S11F (arrows)]. Within A-to-B compartment switches, the levels of H3K27ac and H3K4me3 were significantly depleted (Fig. 3D), and the interactions between H3K27ac peaks were reduced

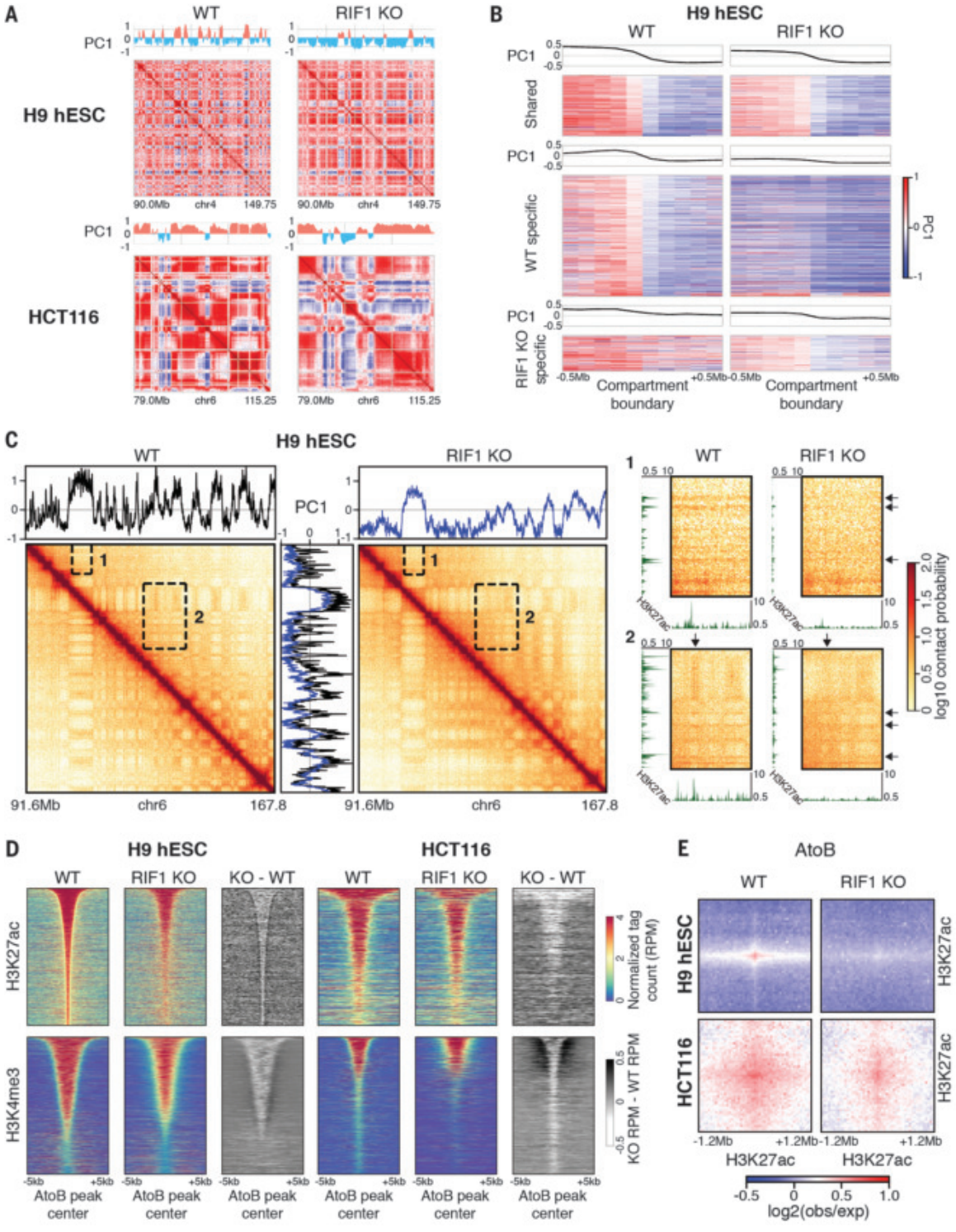


Fig. 3. RIF1 KO causes global alterations of compartments and epigenetic state. (**A**) Correlation matrices and PC1 eigenvector of Chr4 90.0 to 149.75 Mb in WT and RIF1 KO H9 hESCs (top) and Chr6 79.0 to 115.25 Mb in WT and RIF1 KO HCT116 (bottom). (**B**) Heatmaps of PC1 values centered on shared (top), WT-specific (middle), and RIF1 KO–specific (bottom) compartment boundaries ±0.5 Mb in WT and RIF1 KO H9 hESCs. (**C**) ICE-normalized Hi-C contact map of Chr6 91.6 to 167.8 Mb in H9 hESC WT and RIF1 KO cells with accompanying PC1 eigenvector plots (black, WT; blue, KO). To the right are expanded views of insets 1 and 2 with accompanying H3K27ac ChIP-seq plots. Arrows indicate compartments and ChIP-seq peaks that are lost upon RIF1 KO. (**D**) Normalized tag counts [reads per million (RPM)] of signal and subtraction plots for H3K27ac and H3K4me3 centered on peaks within A-B compartment-switching regions ±5 kb in WT and RIF1 KO cells sorted by peak size. (**E**) Aggregate Hi-C log2(observed/expected) interactions between H3K27ac peaks within A-B compartment-switching regions ±1.2 Mb in WT and RIF1 KO cells.

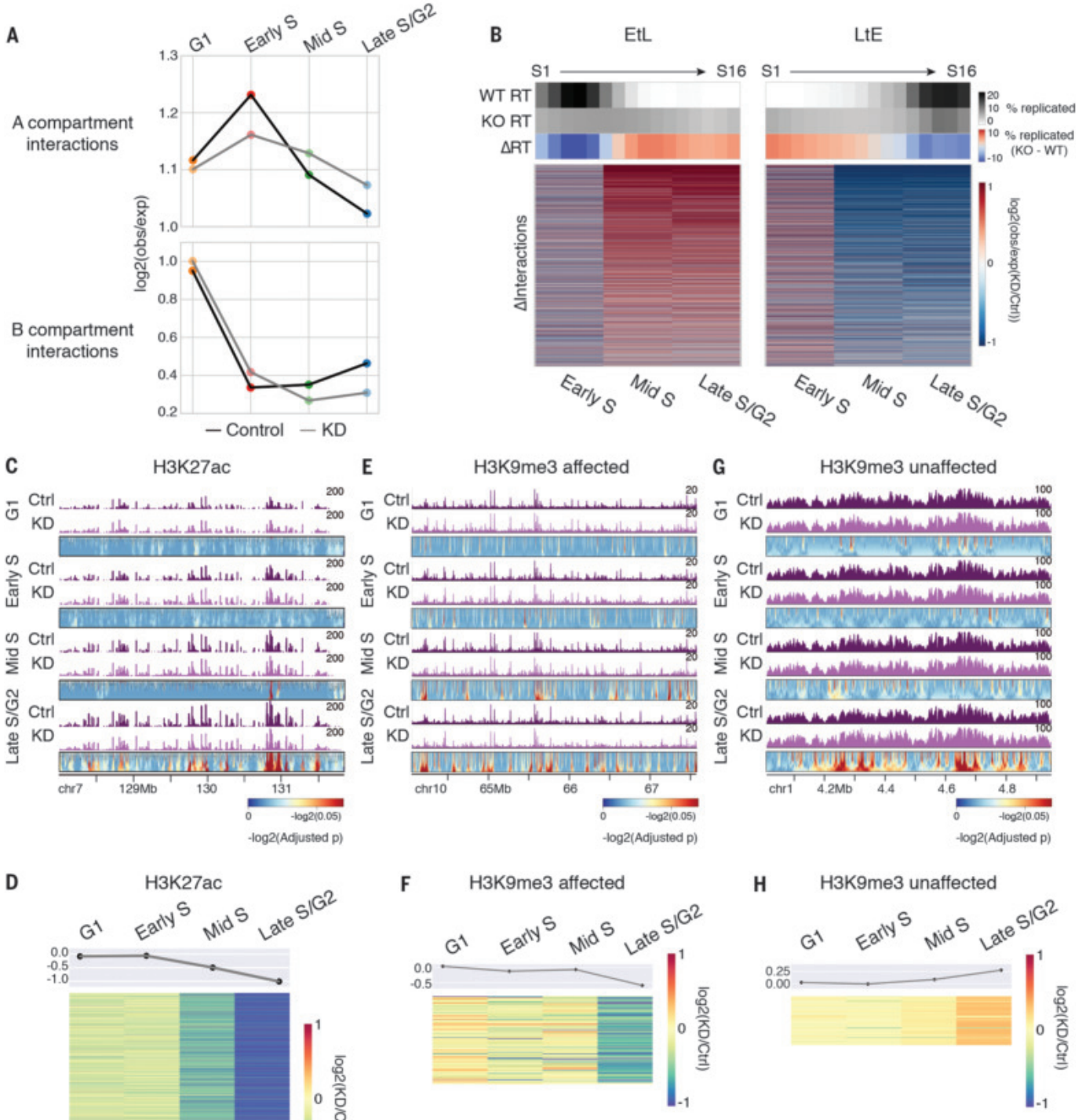


Fig. 4. Epigenome effects require DNA replication and occur during the first S phase after RIF1 degradation. (**A**) Average log2(observed/expected) within the top 10% A (top) and B (bottom) compartments in control (black line) and RIF1 KD (gray line) cells at G_1, early S, middle S, and late S/G_2 time points. (**B**) Interaction frequency differences between control and RIF1 KD at top 10% EtL and LtE regions at G_1, early S, middle S, and late S/G_2 time points whose RT is indicated by average high-resolution Repli-seq heatmaps. (**C**) H3K27ac ChIP-seq of Chr7 128 to 131.9 Mb in control (top), RIF1 KD (middle), and domainogram (bottom), indicating −log2 (Benjamini-Hochberg–adjusted *P* values) calculated for the differences of control subtracted from KD tracks (supplementary methods). (**D**) Heatmap showing log2(KD/control) of H3K27ac peaks. (**E**) H3K9me3 ChIP-seq of Chr10 64.05 to 67.6 Mb in control (top), RIF1 KD (middle), and domainogram (bottom). (**F**) Heatmap showing log2(KD/control) of H3K9me3 peaks at affected regions. (**G**) H3K9me3 ChIP-seq of Chr1 4.05 to 5.0 Mb in control (top), RIF1 KD (middle), and domainogram (bottom). (**H**) Heatmap showing log2(KD/control) of H3K9me3 domains at unaffected regions.

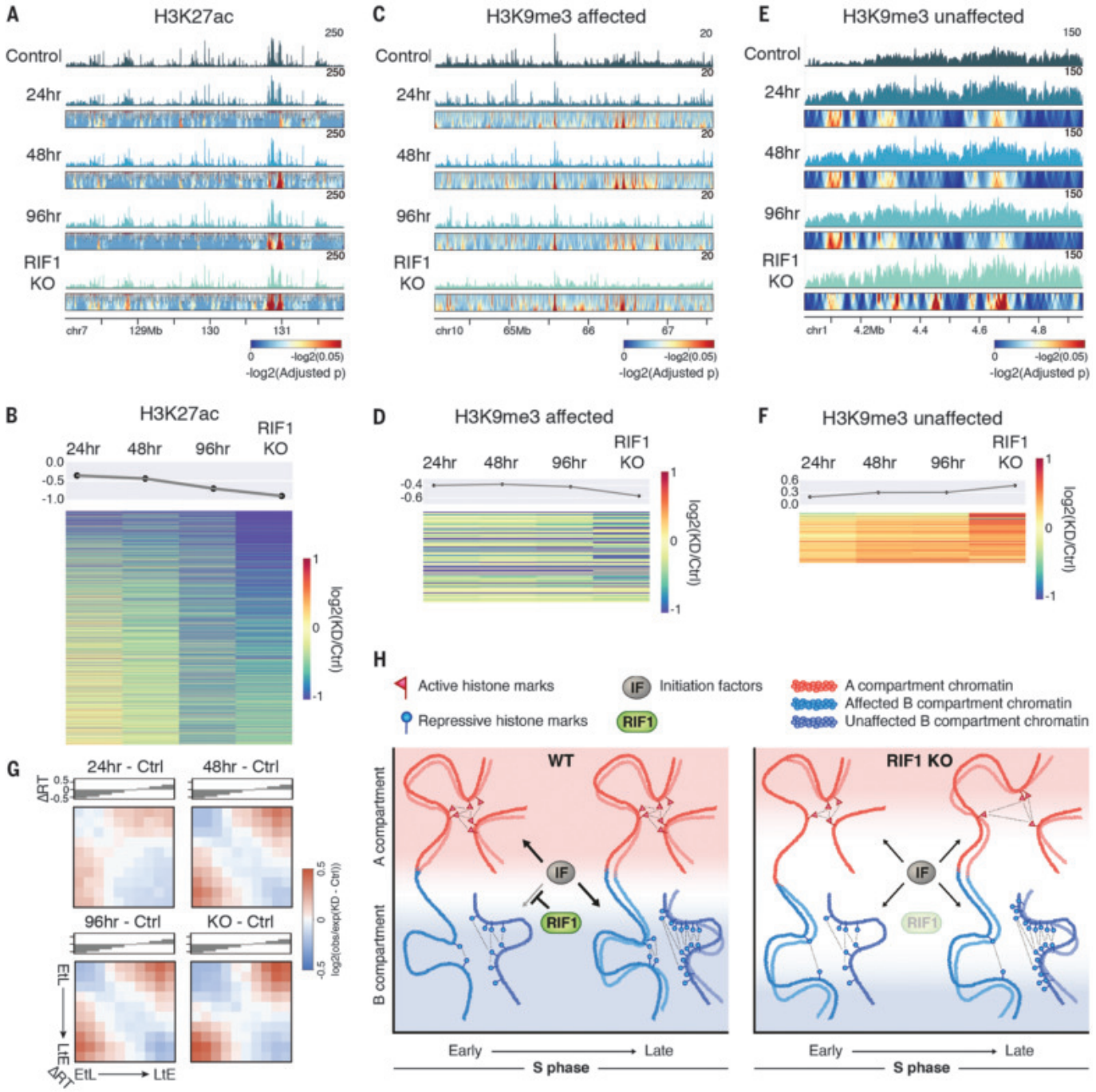


Fig. 5. Epigenome effects are exacerbated as cells cycle without RT control. (**A**) H3K27ac ChIP-seq of Chr7 128 to 131.9 Mb in control (top), RIF1 KD (middle), and domainograms (bottom) in controls, after 24, 48, or 96 hours of degradation, and RIF1 KO. (**B**) Heatmap showing log2(KD/control) of H3K27ac tags. (**C**) H3K9me3 ChIP-seq of Chr10 64.05 to 67.6 Mb in control (top), RIF1 KD (middle), and domainograms (bottom). (**D**) Heatmap showing log2(KD/control) of H3K9me3 tags at affected regions. (**E**) H3K9me3 ChIP-seq of Chr1 4.05 to 5.0 Mb in control (top), RIF1 KD (middle), and domainograms (bottom). (**F**) Heatmap showing log2(KD/control) of H3K9me3 tags at unaffected regions. (**G**) Subtraction saddle plots of cis-log2(observed/expected) contacts of control sample from 24, 48, and 96 hours of degradation, and RIF1 KO sorted by ΔRT (KD RT − control RT). (**H**) Working model illustrating the role of RT in epigenome maintenance. In WT cells, RIF1 prevents replication initiation factors from activating replication within repressive chromatin, allowing early replication of active chromatin, which is then assembled with active histone marks (red flags) that self-interact to form the A compartment. Repressive chromatin is replicated in late S phase and is assembled with repressive histone marks (blue circles), which self-interact to form the B compartment. The B compartment is divided into regions that depend on RIF1 for late replication (affected) and those that do not (unaffected). RIF1 KO allows limiting replication initiation factors to associate heterogeneously with both active and repressive chromatin, delaying replication of the former and advancing replication of the latter. Delayed replication of active chromatin causes depletion of active histone marks and weakens interactions within the A compartment. Advanced replication of affected repressive chromatin causes depletion of repressive histone marks and weakens interactions within the affected B compartment hub. Unaffected B compartment hubs associated with repressive domains maintain their late replication in RIF1 KO and become enriched for repressive histone marks.

(Fig. 3E). These data show that depletion of active histone modifications, particularly H3K27ac, strongly correlates with loss of A compartment interactions in RIF1 KO cells.

Both cell lines exhibited genome-wide weakening of A-A compartment interactions that was accompanied by strengthening of B-B compartment interactions (fig. S12A). This was further confirmed when we called statistically significant differential interactions using differential Hi-C. Strengthened interactions were seen predominantly within the B compartment and weakened interactions were concentrated in the A compartment in both cell lines (fig. S12, B and C). This is likely the combined effect of strengthened interactions between up-regulated H3K9me3 domains within the B compartment (fig. S12D) and weakened interactions between down-regulated H3K27ac peaks in the A compartment (fig. S12E) and suggests a redefining of compartment identities in RIF1 KO cells. EtL and LtE regions showed occasional loss and gain of TAD boundaries, respectively (fig. S13A). However, globally, the positioning and number of TAD boundaries was not significantly affected by RIF1 KO in either cell line (fig. S13B), whereas the strength of TAD boundaries was increased (fig. S13, C and D), independent of RAD21 binding (fig. S14). Together, these data indicate a substantial reorganization of genomic compartments that correlates with epigenome changes in RIF1 KO cells.

DNA replication is required for epigenomic changes

To directly address the primary role of RIF1, we used an auxin-inducible degron (AID) system (*13*) in HCT116 cells to rapidly degrade RIF1 (fig. S15A). We first confirmed that RT defects observed in RIF1 KO occur during the first S phase after RIF1 loss (fig. S15, B and C) by degrading RIF1 in G_1-synchronized cells and performing E/L Repli-seq upon release into S phase. We next synchronized cells in G_1, degraded RIF1, either held them in G_1 or released them into early, middle, or late S phase/G_2 (fig. S15, D and E), and performed Hi-C and ChIP-seq. We found no notable difference in genome organization between RIF1 degraded and control cells held in G_1 (Fig. 4A and fig. S16, A to F). We observed extensive architectural changes at both A and B compartment chromatin as control cells transitioned from G_1 into and through S phase (Fig. 4A and fig. S16, A to F), controlled for copy number (fig. S16G), and this was characterized by an initial increase and subsequent decrease in interaction frequencies within A compartment and the converse within B compartment regions (Fig. 4A, control lines). RIF1 KD cells showed diminished dynamics with decreased intra-A interactions and increased intra-B interactions during early S phase and the reverse during middle and late S phase compared with control cells (KD lines in Fig. 4A and fig. S16, C to F). These changes were coincidental with RT disruptions at corresponding loci (Fig. 4B and fig. S16H). These results demonstrate that RIF1 is required for extensive architectural changes during S phase, suggesting that these changes are linked to the normal temporal order of replication.

H3K27ac and H3K9me3 ChIP-seq (spike-in) also revealed no significant difference between RIF1 degraded and control cells held in G_1 (G_1 data in Fig. 4, C to H; and fig. S17, A and B), indicating that histone mark changes are also not a direct result of RIF1 loss. H3K27ac peaks were not depleted until cells reached middle S phase with the extent of depletion greatest at late S phase [Fig. 4, C and D; and fig. S17, A (blue dots) and C], suggesting that disruption of early-enriched histone marks takes place only when they are replicated late. H3K9me3 peaks at affected regions became depleted in late S/G_2 [Fig. 4, E and F; and fig. S17, B (blue dots) and D], whereas H3K9me3 domains at unaffected regions became more enriched for H3K9me3 at late S phase/G_2 [Fig. 4, G and H; and fig. S17, B (red dots) and E]. Together, these results indicate that disruption of genomic structure and chromatin identity in RIF1 degraded cells is a result of a dysregulated DNA replication program.

Epigenomic changes are quantitatively correlated with lengths of RT dysregulation

We further hypothesized that extended degradation of RIF1 and RT disruption through multiple rounds of DNA replication would exacerbate the observed epigenomic phenotypes. To test this, we degraded RIF1 for 24, 48, and 96 hours (fig. S18A), constituting ~1, 2, and 4 cell cycles respectively. After 24 hours of RIF1 degradation, the RIF1 KO RT phenotype was fully recapitulated and remained so throughout the time course (fig. S17, B and C). H3K27ac peaks that were depleted in RIF1 KO cells, and the interactions between them, progressively diminished toward RIF1 KO levels with time (Fig. 5, A and B; and fig. S18, D to F). H3K9me3 peaks at affected late-replicating regions became significantly depleted after 24 hours of RIF1 degradation and remained depleted throughout the time course [Fig. 5, C and D; and fig. S18, G and H (blue dots)]. Unaffected regions progressively accumulated H3K9me3 throughout the time course [Fig. 5, E and F; and fig. S18, H (red dots) and I]; however, the strengthening of interactions between H3K9me3 domains seen in RIF1 KO was not observed even after 96 hours of RIF1 degradation (fig. S18J). Sorting chromatin interactions by the change in RT (ΔRT) revealed that regions that changed RT (both EtL and LtE) gradually lost interactions with regions of normally similar RT (upper left and lower right corners of Fig. 5G) and gained interactions with regions of normally different RT (upper-right and lower-left corners of Fig. 5G), indicating a strong correlation between RT loss and disruption of chromatin contacts. Together, these data show that as cells progress through multiple rounds of DNA replication with dysregulated RT, they gradually accumulate aberrant epigenomic and structural signatures, supporting the model that stochastic RT gradually affects newly synthesized chromatin with each round of mistimed chromatin assembly.

RIF1 KO causes limited gene expression changes

RIF1 KO caused expression changes to 2284 genes in H9 hESCs and 1737 genes in HCT116 (fig. S19, A and B) as well as to some endogenous retrovirus (ERV) elements (fig. S19C). Expression of pluripotency factors or histone modification writers was not affected (fig. S19D). Genes affected in HCT116 RIF1 KO gradually changed their expression during the RIF1-AID degradation time course toward RIF1 KO levels (fig. S19E); however, only six significantly changed genes were shared between all RNA sequencing (RNA-seq) samples (fig. S19F and table S1). Single-cell RNA-seq revealed an increase in global cell-to-cell heterogeneity in gene expression in RIF1 KO H9 hESCs and HAP1 cells but not HCT116 (fig. S19, G to I).

RT and chromatin compartmentalization are correlated with gene expression (*1*); however, neither changes in RT nor compartment switches were able to predict gene expression changes in RIF1 KO cells (fig. S20, A to D). Differentially expressed genes showed cell type–specific changes in the distribution of specific histone modifications around their transcription start sites that correlated with expression changes (fig. S20, E and F). Altogether, these results suggest that gene expression changes are indirect effects of RIF1 depletion.

Discussion

We show that deletion of RIF1 results in a dramatic increase in cell-to-cell RT heterogeneity rather than discrete population-level RT switches. Although RIF1 is bound to late-replicating chromatin, the entire early-replicating genome as well as most of the late-replicating genome becomes disrupted for RT control without RIF1 (Fig. 1). This is reminiscent of observations in budding yeast, where derepression of late-replicating origins increased competition for limiting replication initiation factors and delayed the RT of early origins (*14*), and it suggests a similar mechanism of RT disruption in RIF1 KO cells. The late-replicating genome is composed of two types of domains whose delayed replication is enforced by different mechanisms and that form separate chromatin hubs. In both hESCs and HCT116 cells, RIF1 KO leads to widespread aberrant

histone modification patterns that correlate with distinct genome-wide changes in three-dimensional genome architecture. We further show that disruptions to epigenome and chromatin structure require DNA replication and that continuous rounds of replication with heterogeneous RT quantitatively exacerbate these changes. We propose that RT changes due to RIF1 KO result in aberrant reestablishment of epigenetic marks that cause profound changes in the epigenetic landscape that then alter genome architecture (Fig. 5H). Although mouse RIF1 has been shown to interact with histone methytransferases (*15*), our cell cycle–synchronized data (Fig. 4) do not support a direct role for RIF1 in heterochromatin maintenance. This work provides mechanistic evidence linking the RT program with maintenance of the global epigenetic state and genome compartmentalization and establishes RIF1 as a key regulator of epigenome maintenance through its role in RT control.

REFERENCES AND NOTES

1. N. Rhind, D. M. Gilbert, *Cold Spring Harb. Perspect. Biol.* **5**, a010132 (2013).
2. N. Reverón-Gómez *et al.*, *Mol. Cell* **72**, 239–249.e5 (2018).
3. T. M. Escobar *et al.*, *Cell* **179**, 953–963.e11 (2019).
4. J. Zhang, F. Xu, T. Hashimshony, I. Keshet, H. Cedar, *Nature* **420**, 198–202 (2002).
5. L. Lande-Diner, J. Zhang, H. Cedar, *Mol. Cell* **34**, 767–774 (2009).
6. S. B. C. Buonomo, in *DNA Replication*, H. Masai, M., Foiani, Eds. (Advances in Experimental Medicine and Biology Book Series, vol. 1024, Springer, 2017), pp. 259–272.
7. C. A. Seller, P. H. O'Farrell, *PLOS Biol.* **16**, e2005687 (2018).
8. R. Foti *et al.*, *Mol. Cell* **61**, 260–273 (2016).
9. R. T. O'Keefe, S. C. Henderson, D. L. Spector, *J. Cell Biol.* **116**, 1095–1110 (1992).
10. P. A. Zhao, T. Sasaki, D. M. Gilbert, *Genome Biol.* **21**, 76 (2020).
11. V. Dileep, D. M. Gilbert, *Nat. Commun.* **9**, 427 (2018).
12. T. Ryba *et al.*, *Genome Res.* **20**, 761–770 (2010).
13. K. Nishimura, T. Fukagawa, H. Takisawa, T. Kakimoto, M. Kanemaki, *Nat. Methods* **6**, 917–922 (2009).
14. K. Yoshida *et al.*, *Mol. Cell* **54**, 691–697 (2014).
15. P. Li *et al.*, *Nucleic Acids Res.* **45**, 12723–12738 (2017).
16. P. A. Zhao, Code for: Replication timing maintains the global epigenetic state in human cells, version 1, Zenodo (2021); https://doi.org/10.5281.zenodo.4606710.

ACKNOWLEDGMENTS

We thank R. Didier and B. Alexander of the FSU Flow Cytometry and Confocal Microscopy Facilities for their help with flow cytometry and fluorescence-activated cell sorting for this project. Thanks to A. Brown of the FSU Biological Science Core Labs and to Y. Yang and C. Vied of the FSU Translational Labs. Thanks to S. R. Westermann of SCIGRAPHIX for generating the model figure. Thanks to B. van Steensel, J. Phillips-Cremins, and P. Fraser for critical reading of the manuscript. **Funding:** This work was supported by NIH grant GM083337 to D.M.G., GM035463 to V.G.C., and GM085354 to D.M.G., S.D., and V.G.C. D.L. is supported by the Hong Kong Research Grant Council (ECS 26104216). T.B. is supported by the William C. and Joyce C. O'Neil Charitable Trust, Memorial Sloan Kettering Single Cell Sequencing Initiative. **Author contributions:** Conceptualization: K.N.K., P.A.Z., V.G.C., and D.M.G; Methodology: K.N.K., A.M.S., I.T., M.Z., L.P.W., S.H., T.N., and T.B.; Formal analysis: P.A.Z.; Investigation: K.N.K., X.L., T.S., D.A.B., and X.Z.; Writing – original draft: K.N.K., P.A.Z., and D.M.G.; Writing – review and editing: K.N.K., P.A.Z., X.L., V.G.C., and D.M.G.; Visualization: K.N.K. and P.A.Z.; Supervision: D.L., M.T.K., A.D.D., H.Z., S.D., V.G.C., and D.M.G.; Funding acquisition: D.M.G., S.D., and V.G.C. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available on the Gene Expression Omnibus (GEO) #GSE160563. Code for data processing is available on Zenodo (*16*) and at https://github.com/oliviacamel/RIF1_KO_analysis.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/371/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S5
References (*17*–*35*)
MDAR Reproducibility Checklist
Data S1 to S3

30 December 2019; resubmitted 1 November 2020
Accepted 19 March 2021
10.1126/science.aba5545

CLIMATE ECONOMICS

Assessing China's efforts to pursue the 1.5°C warming limit

Hongbo Duan[1]*, Sheng Zhou[2]*, Kejun Jiang[3], Christoph Bertram[4], Mathijs Harmsen[5,6], Elmar Kriegler[4,7], Detlef P. van Vuuren[5,6]*, Shouyang Wang[1,8]*, Shinichiro Fujimori[9,10,11], Massimo Tavoni[12,13], Xi Ming[1], Kimon Keramidas[14], Gokul Iyer[15], James Edmonds[15]

Given the increasing interest in keeping global warming below 1.5°C, a key question is what this would mean for China's emission pathway, energy restructuring, and decarbonization. By conducting a multimodel study, we find that the 1.5°C-consistent goal would require China to reduce its carbon emissions and energy consumption by more than 90 and 39%, respectively, compared with the "no policy" case. Negative emission technologies play an important role in achieving near-zero emissions, with captured carbon accounting on average for 20% of the total reductions in 2050. Our multimodel comparisons reveal large differences in necessary emission reductions across sectors, whereas what is consistent is that the power sector is required to achieve full decarbonization by 2050. The cross-model averages indicate that China's accumulated policy costs may amount to 2.8 to 5.7% of its gross domestic product by 2050, given the 1.5°C warming limit.

As part of the Paris Agreement, nearly all countries have agreed to take steps to limit the global surface average temperature increase to less than 2° or 1.5°C compared with preindustrial levels (*1*). Without stringent climate policy, global mean warming is likely to result in temperature increases greater than 2.5°C, which could decrease the world's per capita output by 15 to 40% (*2*), although those economic impacts are highly uncertain. The climate impacts in different regions are likely to show large differences (*3*, *4*), with some countries suffering disproportionately serious consequences without the adaptation of early and stringent climate change mitigation policies (*5*, *6*). For some less-developed states that are endowed with rich natural resources, half of their existing species may be in danger of extinction (*7*), and many coastal regions may be exposed to multiple risks of diseases resulting from climate change (*8*). Even worse, the imbalanced distribution of loss and damage will exacerbate the inequal trends of economic development across countries (*9*).

Although both the 2° and 1.5°C goals have been discussed frequently since the Copenhagen Accord in 2009, most research efforts have focused on the 2°C target (*10*–*12*). The 1.5°C goal began to receive considerable attention only after it was formally adopted in the Paris Agreement, and a minority of studies since, particularly at the country level, have been conducted using it (*9*, *13*). Still, research has now shown that there are important differences between a 2° and 1.5°C warming (*14*, *15*); for example, the economic cost to reach the 1.5°C goal may be at least threefold that of the 2°C goal (*16*). The Intergovernmental Panel on Climate Change (IPCC) released a special report in 2019 on the impacts of global warming of 1.5°C above preindustrial levels and the

[1]School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China. [2]Institute of Energy, Environment and Economy, Tsinghua University, Beijing 100084, China. [3]Energy Research Institute, Chinese Academy of Macroeconomic Research, Beijing 100038, China. [4]Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, 14473 Potsdam, Germany. [5]Netherlands Environmental Assessment Agency (PBL), Postbus 30314, The Hague, Netherlands. [6]Copernicus Institute of Sustainable Development, Utrecht University, P.O. Box 80125 3508 TC Utrecht, Netherlands. [7]Faculty of Economics and Social Sciences, University of Potsdam, 14482 Potsdam, Germany. [8]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. [9]Department of Environmental Engineering, Kyoto University, Kyoto 615-8540, Japan. [10]Center for Social and Environmental Systems Research, National Institute for Environmental Studies, Tsukuba 305-8506, Japan. [11]International Institute for Applied Systems Analysis, 2361 Laxenburg, Austria. [12]Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan 20123, Italy. [13]RFF-CMCC European Institute on Economics and the Environment, Fondazione Centro Euromediterraneo sui Cambiamenti Climatici, Milan 20123, Italy. [14]European Commission, Joint Research Centre (JRC), Seville E-41092, Spain. [15]Joint Global Change Research Institute, University of Maryland and the Pacific Northwest National Laboratory, College Park, MD 20740, USA.
***Corresponding author. Email: hbduan@ucas.ac.cn (H.D.); zhshinet@tsinghua.edu.cn (S.Z.); detlef.vanvuuren@pbl.nl (D.P.v.V.); sywang@amss.ac.cn (S.W.)**

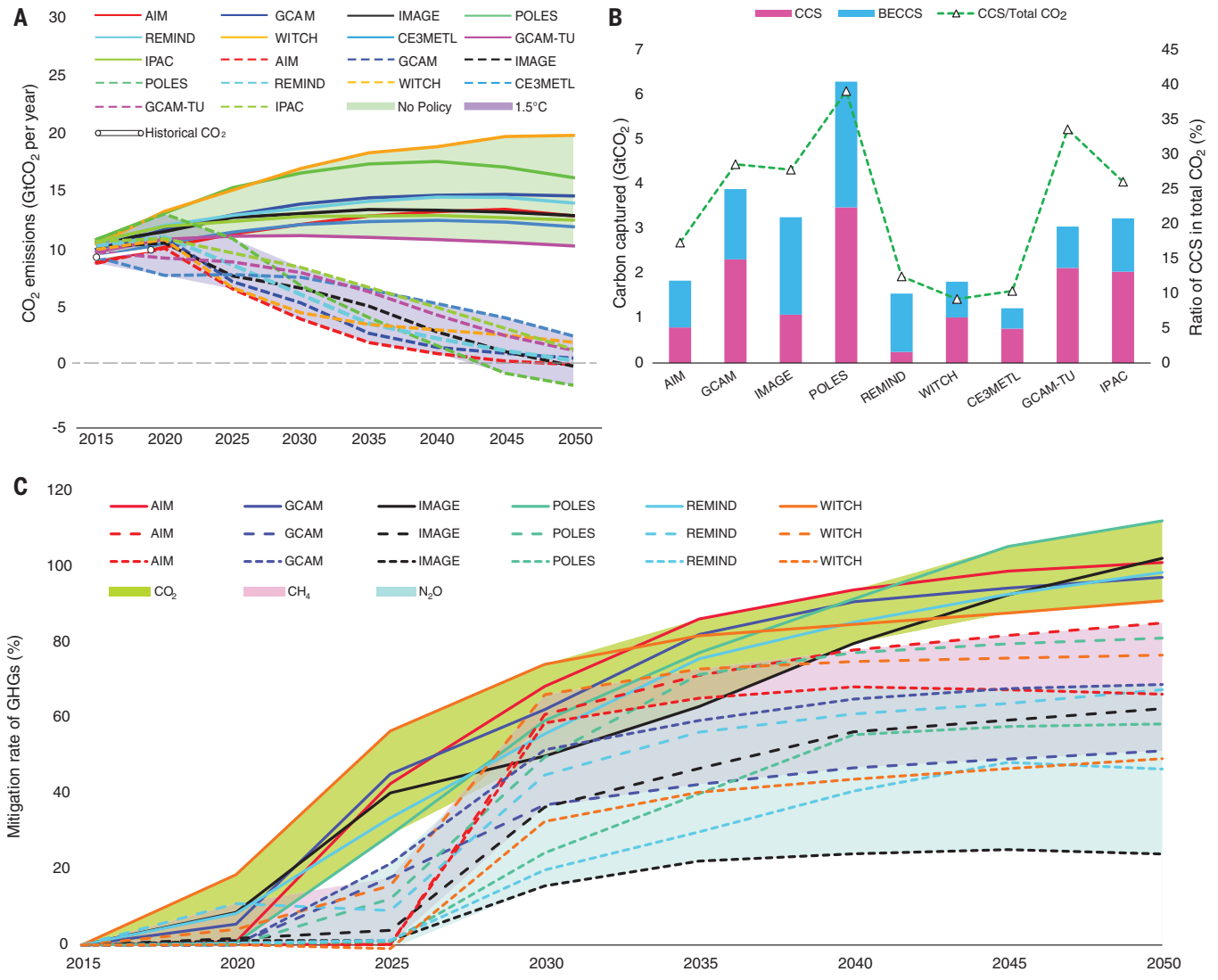


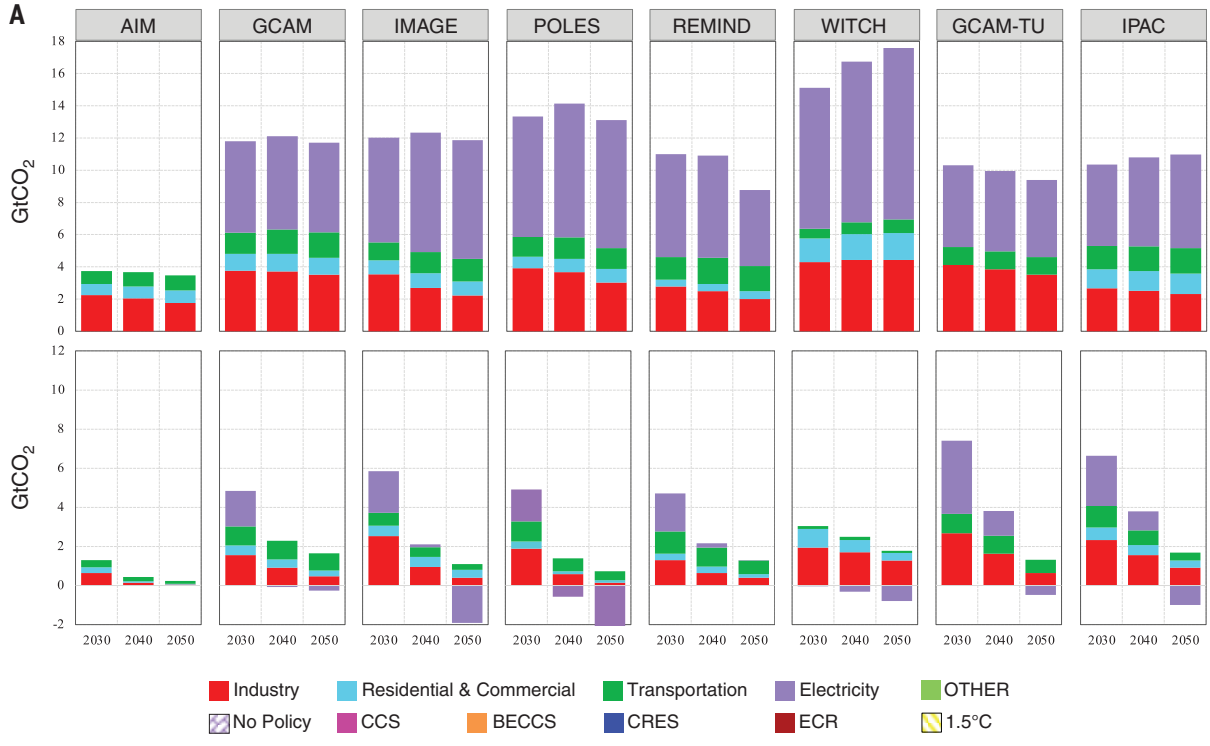
Fig. 1. Cross-model emission pathways of China. (**A**) Changes in total carbon emissions under both the no policy (solid lines colored with light green) and warming-limit scenarios (dashed lines colored with light purple). (**B**) The roles of the CCS for fossil fuels and biomass in attaining the 1.5°C goal by 2050. (**C**) Reduction ratios of carbon and noncarbon emissions across the models under the 1.5°C-consistent scenario (relative to the no policy case). GHGs, greenhouse gases.

related greenhouse gas emission pathways (*15*). Nevertheless, current studies on the 1.5°C goal are far from adequate to be enough for the sixth assessment report (AR 6) on climate change that is due to be released in 2022 (*17*, *18*).

At the same time, there is clearly no consensus on the attractiveness of a 1.5°C target, particularly given the uncertainty in associated mitigation costs. For instance, little attention has been paid to what this strict warming-stabilizing target means for China, particularly from the perspective of a multimodel comparison framework (*19*). The latest report released by the China Coal-Control Project provides a preliminary discussion of China's mitigation scenarios and technology pathways under the 1.5°C goal, emphasizing the importance of immediate mitigation actions followed by an annual decrease of nearly 400 million tonnes (Mt) in CO_2 emissions after 2020 and the first realization of negative emissions of the power sector in 2050 (*5*, *20*). It suggests that as a high population–exposed country, China may suffer substantially and more than other countries from the imbalanced distribution of adverse climate impacts. Therefore, we conducted this multimodel study to explore the consequences for China of doing its part to reach the 1.5°C warming target, based on the existing scenario from global models and new national modeling. The aim is to understand what results are model-robust, given different models that are widely used in both global and domestic climate policy research, and what results are most uncertain across models. Our findings are informative and insightful for China's long-term mitigation and carbon neutrality, as well as its contribution to global climate goals, and for identifying the need for future research.

Pathways of carbon and noncarbon emissions

Figure 1 reports the difference in emission trajectories across participating integrated assessment models (IAMs). For example, the range of the carbon emissions in 2050 under the "no policy" case is 10.2 to 19.8 billion tonnes of CO_2 ($GtCO_2$), and this is 1.9 to 2.3 $GtCO_2$ for the 1.5°C-consistent scenario. Except for the WITCH (World Induced Technical Change Hybrid) model, all the other models agree on China's carbon peaking in the no policy case, despite great uncertainty on peak values (from 10.9 to 17.5 $GtCO_2$) and timing (from 2035 to 2045). Based on the weak growth in coal consumption and the declining trend of CO_2 emissions in recent years, an emission peak during the



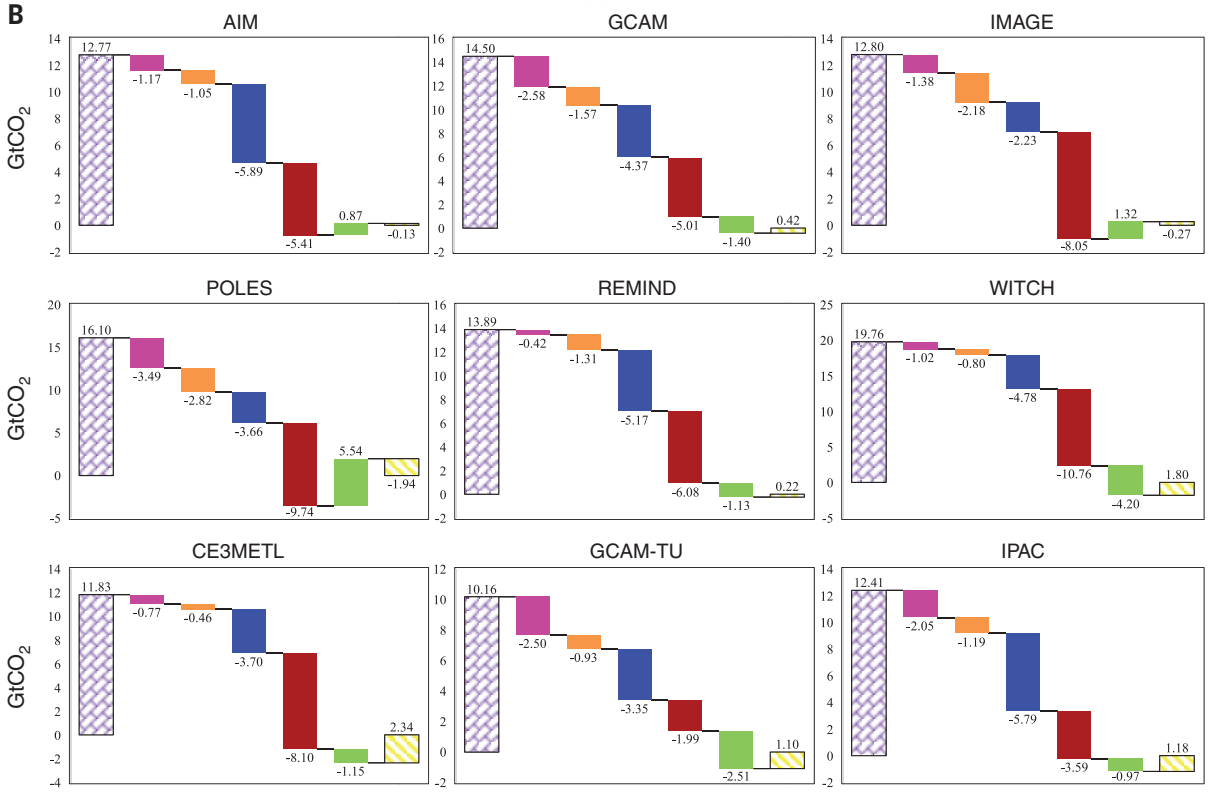


Fig. 2. Cross-model comparisons of China's emission reductions. (**A**) Sectoral carbon emissions from both the energy supply and demand sides across models. The top row shows the no policy baseline versus the 1.5°C-consistent scenario shown in the bottom row. No CO_2 data in the power sector were reported for the AIM model; emissions of heating were also absent for many models, and we therefore exclude it in this analysis. (**B**) Cross-model analysis of distributions of mitigation contributions under the warming limits in 2050. CRES denotes carbon emission reductions resulting from energy substitution (nonfossil energy for fossil fuels); ECR measures the emission abatement that is directly from energy consumption reduction. Other forms of change in carbon emissions (such as changes in land use) are included in the category OTHER.

period of 2035 to 2040 is certainly feasible, even in the absence of additional mitigation efforts (*21*). Despite limited convergence emission trajectories across models under the 1.5°C scenario, one highly consistent finding is that carbon emissions decrease steeply beginning in 2020. The majority of the IAMs will achieve near-zero or negative emissions by around 2050 (from −1.94 to 2 $GtCO_2$), a result in line with results from a global-scale analysis (*17*). We find that an early emissions peak followed by steep reductions thereafter reduces dependence on negative emission technologies (NETs) for realizing the 1.5°C target. As a consequence, an important trade-off exists between substantial early mitigation actions and reliance on NETs with uncertain performance, when developing pathways to realize the 1.5°C warming goal.

Carbon capture and storage (CCS) technologies, including conventional fossil fuel CCS and bioenergy with CCS (BECCS) technologies, play important roles in limiting global average temperature change in the year 2100 to 1.5°C (Fig. 1B), and this is consistent across various models (*22*). As seen from the results for 2050, the lowest emission appears in the POLES (Prospective Outlook on Long-term Energy Systems) model, which has the greatest reliance on CCS and BECCS, with a share of total emissions reaching 39.2% (Fig. 1B). Similarly, the lesser use of CCS in the WITCH and CE3METL (Chinese Energy-Economy-Environmental Model with Endogenous Technological change by employing Logistic curves) models (less than 15%) largely explains their failure to reach net-zero or negative emissions by 2050. It is worth noting that the carbon emission for the AIM (Asia-Pacific Integrated Model) becomes negative in 2050, with a relative low ratio of CCS capture to total emissions, implying the substantial potential of carbon reduction associated with a low-carbon energy transition. Noncarbon emissions also need to be substantially curbed to ensure that the increase in the temperatures stays below 1.5°C (Fig. 1C). The multimodel comparisons reveal that the amount of noncarbon emissions, including primary CH_4 and N_2O, is rather small compared with that of the carbon emissions (12.8 to 16.1 Gt), with magnitudes of 70.7 to 98.9 Mt and 1.8 to 3.3 Mt, respectively. However, the reduction in noncarbon emissions still proves to be indispensable, particularly for attaining strict climate goals (*23*). Actually, carbon emissions must be reduced by more than 90% to hit the 1.5°C limit, and CH_4 and N_2O emissions are expected to decrease, on average, by 70.6 and 52.2%, respectively, in comparison to the no policy case (Fig. 1C).

Industrial CO_2 emissions and mitigation contributions

The mitigation challenges of the 1.5°C limit differ considerably across sectors in terms of both demand and supply levels (Fig. 2). From a demand perspective, industry is the main emission contributor, and this is a model-consistent finding. In the absence of additional mitigation policies, China's industrial emissions account for 48.7 to 75.2% of the total emissions of energy demand in 2030. As seen in the time scale (Fig. 2A), most of the models show declining trends in industrial emissions, and this decrease could be up to 37% from 2030 to 2050. When moving to the emission structure, we observe another model-consistent finding: The share of China's industrial emissions to total emissions of energy demand gradually declines to about 50% in 2050. This result is driven by three forces operating in China's industrial sector, i.e., structure adjustments, low-carbon transition of energy use, and substantial enhancement of energy efficiency. Indeed, China might be expected to go through an energy-use transition even absent greenhouse gas emissions mitigation. Such energy transitions accompany economic development and its associated economic restructuring, air pollutant controls, and water and soil contamination treatment. In terms of energy supply, electricity production is the primary source of carbon emissions in China. As shown in Fig. 2A, emission pathways for power supply are not consistently decreasing across models. The emission structure of the supply side is quite stable when it is compared with that of the demand side, which implies the difficulty in carbon abatement arising from supply-side energy restructuring.

The 1.5°C warming limit directly leads to a pronounced decrease in sectoral emissions, which is consistent on both the supply and demand sides. On average, carbon emissions in the power sector will decline by 65.6% across models in 2030, and near-zero or negative emission status could be attained as deep decarbonization proceeds over time. The WITCH model appears to be the first to achieve negative emissions (in 2030), whereas the POLES and IMAGE (Integrated Model to Assess the Global Environment) models gain the highest values for negative emissions, i.e., 2.6 and 1.9 $GtCO_2$ in 2050. The cross-model analysis on the distributions of carbon emission mitigation shows that CCS plays a formidable role in achieving the 1.5°C goal (Fig. 2B), but it cannot be the dominant contributor to emission reduction. The largest proportion of the reduction comes from a substantial energy

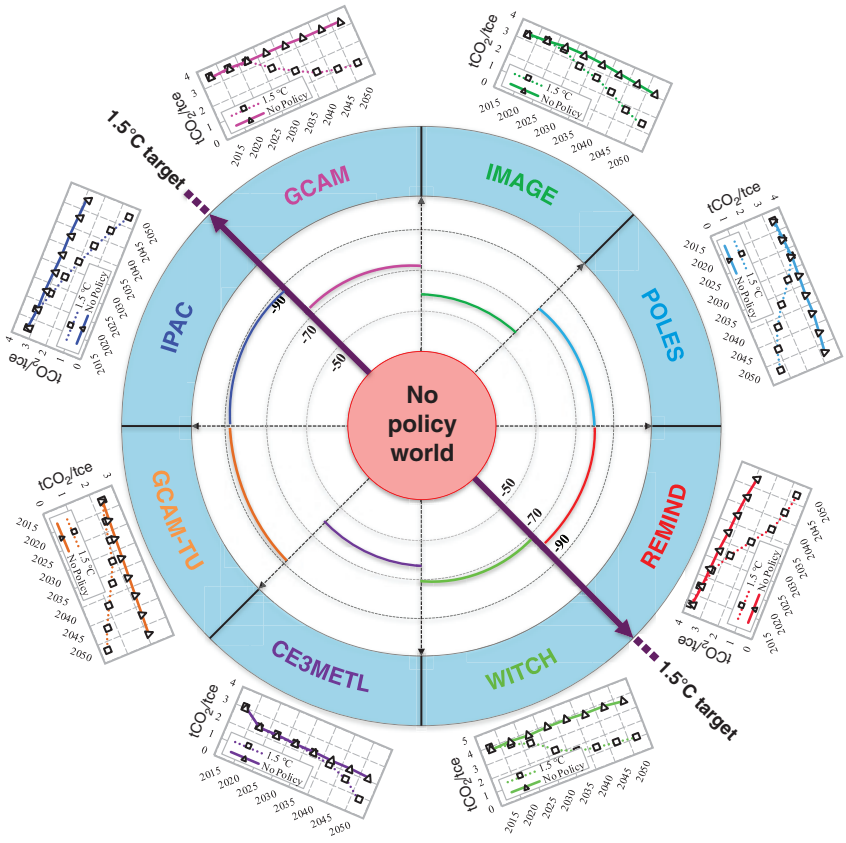


Fig. 3. Challenges regarding carbon intensity given the 1.5°C warming limit. The main chart depicts the percentages of carbon intensity improvement (relative to the no policy case) across the target IAMs in 2050, and the subfigures show the paths of the carbon intensity of end-use energy under a 1.5°C-consistent scenario.

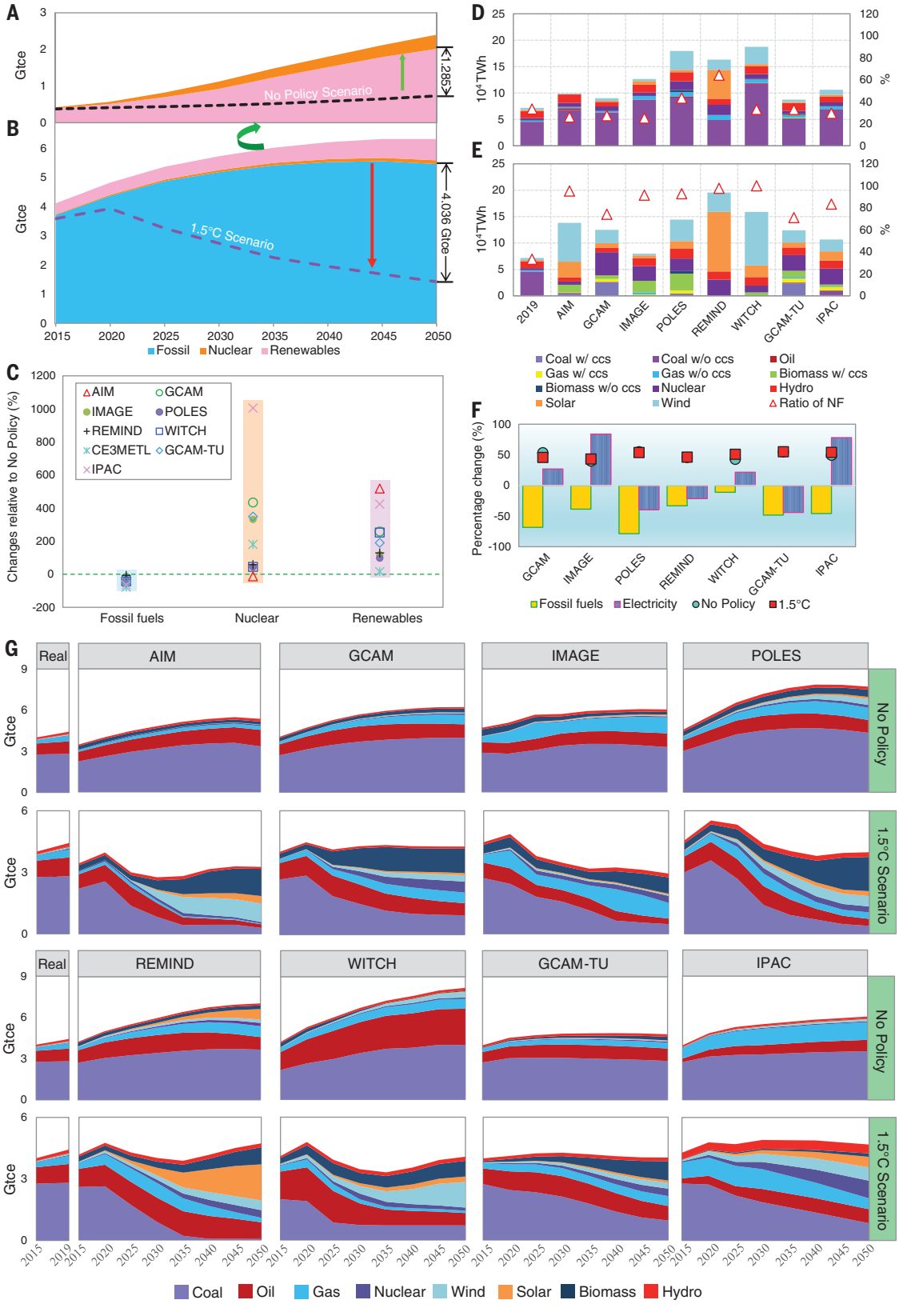


Fig. 4. Changes in the TPEC across models under the 1.5°C warming limit. (**A**) The growth in the magnitude of renewables under the 1.5°C scenario. The pink area represents the growth of renewables under the 1.5°C scenario, and the dashed black line shows the results under the no policy scenario; the green arrow indicates the magnitude of growth. (**B**) The decline in fossil energy demand necessary to hit the 1.5°C goal. The blue area represents the evolution of fossil fuels under the no policy scenario, and the dashed purple line shows the results under the 1.5°C scenario; the red arrow shows the magnitude of fossil fuel decrease under the 1.5°C scenario. The curved green arrow highlights the growth of renewables from the no policy scenario to the 1.5°C scenario. (**C**) Changes in fossil fuels, nuclear, and renewables across the models relative to the no policy levels. (**D** and **E**) The evolution of the secondary energy structure in 2050 under the no policy case (D) and the corresponding changes in the power structure under the 1.5°C case (E). The triangles in both subfigures denote the ratios of nonfossil power (NF) to total secondary power consumption. (**F**) Changes in industrial energy demand (relative to the no policy case). The bars in this figure show changes in industrial fossil fuels and electricity consumption in 2050 by comparing the 1.5°C case to the no policy case; circles and squares represent shares of the industrial final energy demand to the total final energy under the two scenarios, respectively. (**G**) Energy restructuring across the target models. The titles at the top and to the right give considered models and policy scenarios, respectively. Different energy consumption is uniformly measured by Gtce. On the left, "Real" provides the historical energy consumption in 2015 and 2019 from China's National Bureau of Statistics.

consumption decline, followed by the substitution of clean energy for fossil fuels, which is a highly consistent finding across all models. This implies that stringent energy demand control and substantial clean energy development are critical to hit the ambitious climate goal.

Challenges on the carbon intensity of end-use energy

Carbon intensity is an important indicator to weight the effect of policy interventions. Carbon intensity is defined as carbon emissions per unit consumption of end-use energy (tCO_2/tce). As depicted in Fig. 3, the 1.5°C limit calls for a substantial decrease in carbon intensity, with a cross-model consistent reduction of 60% relative to the no policy case. The declines in the GCAM-TU (Global Change Assessment Model–Tsinghua University), IPAC (Integrated Policy Assessment model for China), POLES, and REMIND (REgional Model of INvestment and Development) models could even be as high as 80%, leading to a high average decrease of 75.6% across participating models. As indicated by the absolute values, carbon intensity in most of the models is less than 1 tCO_2/tce (except IMAGE) in 2050, and it is the lowest in GCAM-TU, i.e., 0.3 tCO_2/tce. We could uncover two driving factors behind the decrease in carbon intensity: The first is both endogenous and exogenous energy efficiency enhancement for single-energy technology, and the second is the structure optimization of end-use energy consumption, including the electrification of the end-use energy system and transitions from carbon-based power to cleaner natural gas or renewables. Carbon intensity improvement may not be a direct objective of climate policy, which acts more on energy consumption control and clean technology development, but it is an important pathway through which energy policy operates.

Energy transformation and decarbonization

Attainment of the 1.5°C target largely depends on the substantial replacement of renewables for fossil fuels (*24*). Figure 4 shows that the 1.5°C goal requires a steep decrease in China's fossil energy consumption, and this decrease could be 4.04 Gt standard coal equivalent (Gtce) on average in 2050, 73.9% lower than the no policy case. In this context, China's nationally determined contribution (NDC) goal of non–fossil energy deployment can be realized early in 2025, and the share of non–fossil fuels in the total primary energy consumption (TPEC) could be further expanded to 30% in 2030 and 62.8% in 2050. By 2050, renewables will dominate China's primary energy supply, with a 1.3 Gtce increase in their amount and a 175% growth on average with respect to the no policy case. The leapfrog development of nuclear and renewables under the 1.5°C warming limit could also be consistently observed from independent

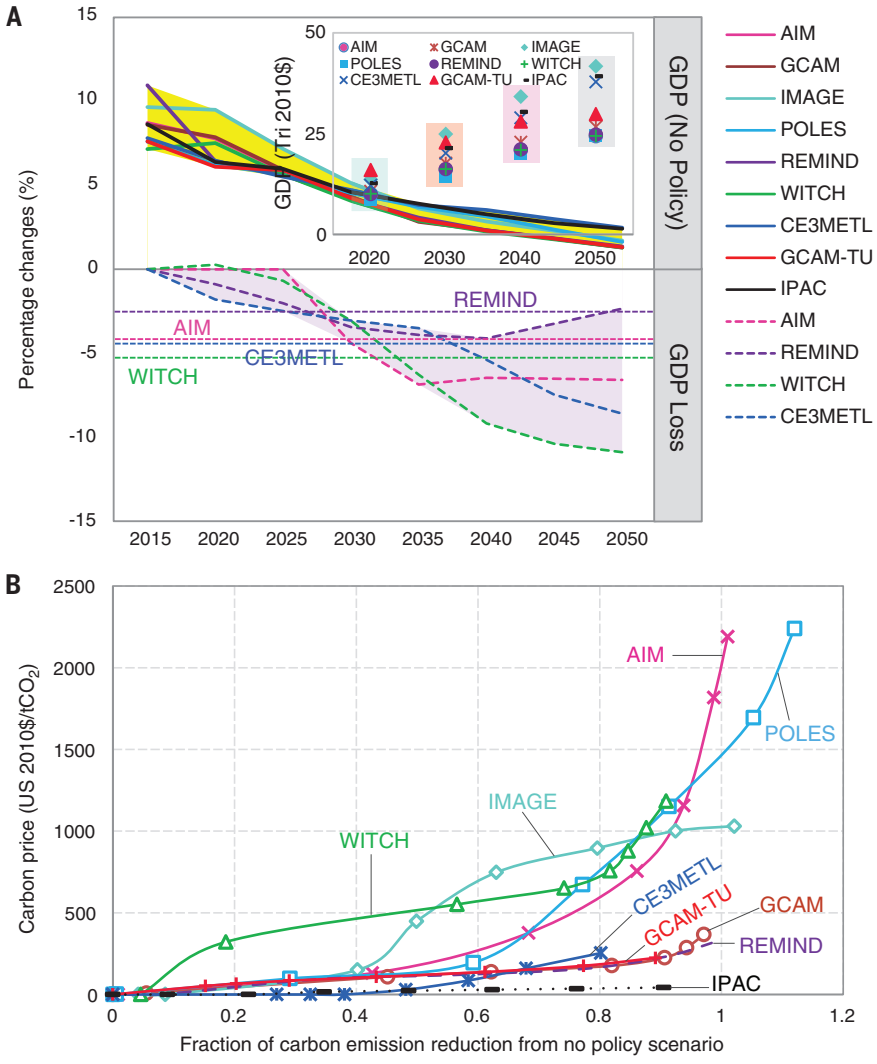


Fig. 5. Impacts of the 1.5°C goal on China's economy. (**A**) The top panel shows GDP growth rates under the no policy scenario (5-year averages from 2010–2015 to 2045–2050), with the magnitude of the GDP across all models embedded, and the bottom panel shows the yearly GDP losses (dashed lines with purple colored band) and cumulative GDP loss (the horizonal dotted lines, from 2020 to 2050 with a 5% discount rate) given the 1.5°C goal (percentage change relative to the no policy results). (**B**) Carbon prices required for China given the 1.5°C warming limit. The horizontal axis gives emission control rates relative to the no policy case.

model results (Fig. 4C). The aggregated impacts of the warming limit on TPEC could be demonstrated in detail by energy restructurings across various models (Fig. 4G). Our model comparison demonstrates the variety of possible energy transitions consistent with the strict 1.5°C temperature control. Some model-robust findings include a steep model-consistent reduction in coal consumption and a substantial development of biomass, wind, and nuclear power technologies. These results hold up well against changes in the cost of renewables (fig. S6).

As a main contributor to emission reduction, the electricity sector faces severe challenges in terms of restructuring and deep decarbonization in the presence of a strict warming limit. Although changes in energy demand across models are highly varied in the no policy case, fossil fuels consistently play a dominant role in the power sector. With the exception of the REMIND model, which is optimistic about the deployment of solar power, the share of fossil energy in sectoral TPEC for all the other models is more than 67% in the no policy scenarios (Fig. 4D). Figure 4, D and E, reports inconsistent cross-model impacts of warming control on energy consumption in the power sector. In effect, there are two typical factors driving the changes in the energy demand in the power sector. First, mitigation drives higher carbon prices, which increase end-use energy prices. Second, because the power sector has many ways to decarbonize, its price falls relative to fossil fuel prices in end use, which leads to a substitution of power for the

direct use of fossil fuels. The latter is the electrification effect (*20*). The final changes in energy consumption in the electricity sector are contingent on the trade-offs of these two effects (see supplementary materials for further discussion). Two robust results emerge from this model intercomparison. First, China's power sector will achieve zero-coal status in 2050, though technology mixes vary widely across models. Second, of all the low-carbon technologies, wind and solar have observed the fastest increases in generation over recent years.

Industry is a primary sector of end-use energy consumption, and substantial changes in energy use must occur to reach deep decarbonization of the entire economy and realization of the given climate goals. Because the power sector has many low-carbon and no-emission technology options available and their future cost and performance are uncertain, models reflect a range of possible decarbonization pathways for this sector (Fig. 4F). What is interesting is that the ratio of industrial energy demand to total end-use energy demand is less affected and relatively stable, slightly fluctuating around the average level of 49% (42 to 55%) across all models under both scenarios. This largely reflects high internal consistency on the dynamic adjustment of China's industrial structure [e.g., share of industrial gross domestic product (GDP) in the economy] and energy efficiency enhancement in the industrial sector.

Economic analysis

We understand the economic implications of climate policy by considering three aspects, i.e., climate damage, the economic impacts of climate policy, and the social cost of carbon (SCC) for various climate targets. The first two indicators are usually measured by the proportion of GDP, and the SCC could be largely equivalent to the optimal carbon tax given exogenous warming and emission controls (*25*). Because most considered models do not report climate-related damage, we conduct our cross-model economic analysis by mainly considering policy costs and the SCC, as depicted in Fig. 5, A and B.

The models do not tell a consistent story regarding the policy costs under the 1.5°C scenario (Fig. 5A). For example, the GDP losses associated with the 1.5°C warming limit could be as high as 10.9% and as low as 2.3% in 2050. This story is still true when moving to the accumulated perspective, with the GDP loss over the time frame ranging from 2.8 to 5.7% (with a 5% discount rate), and they are observably larger than the global levels (fig. S1). However, as seen in the model averages, we find a similar sized accumulated impact of the 1.5°C warming limit on China and the world, i.e., 3.3% (relative to the no policy level). Given that the no policy scenario is mostly counterfactual, the GDP loss could be understood as an upper-bound cost to hit the 1.5°C goal. Generally, increased GDP growth implies an increase in energy consumption. Although energy systems have long been dominated by fossil fuels in China, in this context, the attainment of the stricter climate target relies on a pronounced decrease in energy consumption, which inevitably plays a negative role in the economy (*26*). By contrast, substantial technological change not only effectively drives economic growth but also greatly helps to reduce the policy costs of climate governance, and this can be seen in the results of the CE3METL model in Fig. 5A.

All the models in this study report pathways of the SCC under the 1.5°C scenario, as informed in Fig. 5B, and we find a great cross-model uncertainty on China's SCC (from $315 to $2240 per tCO_2 in 2050), which repeats the result that has been clearly examined at the global scale (*19*). The notable difference in carbon prices is not surprising because there are many primary factors that lead to this difference, including the basic setting on costs and mitigation performance for various energy technologies, the substitutability of fossil fuels and alternatives, cross-regional trade modes, and the transmission mechanism of policy costs (see supplementary materials for more discussion).

Model-intercomparison findings

Based on the IAM-based multimodel analysis, we found that China's carbon-neutral announcement in 2060 is largely consistent with the 1.5°C warming limit, despite a greater challenge for the latter (fig. S7). Our model intercomparison produced several consistent insights. First, the global 1.5°C warming limit is associated with emission reductions exceeding 90% (90 to 112%) in China's total CO_2 emissions and an average decrease of 70.6% (51.3 to 85%) and 52.2% (24.1 to 68.8%) in CH_4 and N_2O emissions, respectively, as compared with the no policy level; the carbon intensity of end-use energy in 2050 needs to be reduced by at least 60% (60 to 87.1%). Second, our model comparison confirms the formidable role of CCS technologies, particularly fossil-based CCS and NETs, in keeping temperature from exceeding critical thresholds. Last, China's fossil fuels and TPEC need to be dramatically reduced by more than 73 and 39%, respectively, to hit the 1.5°C warming limit, and most models report that China's coal demand will be declining to near zero around 2050. The attainment of the 1.5°C goal calls for large-scale emission reductions resulting from the accelerated deployment of renewables and their substantial replacement of carbon-based fuels.

Uncertain implications

The emission pathways are sensitive to variation in model structure and technology assumptions that lead to variation in the date and magnitude of peak emissions, a finding that is largely in agreement with the global-scale findings (*27*, *28*). Although the 1.5°C target drives early mitigation action with steep reductions thereafter, the magnitude of the decrease greatly differs across participating models. The models in this study displayed a wide range of mitigation pathways, including differences in the role of the sectors. By 2030, the power sector is required to cut down its CO_2 emissions by about 66% and achieve full decarbonization by 2050. Nonetheless, models reported a wide variety of patterns in sectoral energy demand and structure. Similar variation exists in the final energy consumption of industry, though the ratio in the total final energy demand seems to be less affected. The model comparison also produced a wide range of policy cost estimates for compliance with the 1.5°C target, a feature reflected in the associated estimates of SCC, with a cross-model difference of more than 10-fold.

The model-consistent findings are mainly around changes in TPEC, emission reductions, and roles of NETs, given the 1.5°C temperature control, with variation in results at the sector level for emission reductions and energy restructuring, as well as for the economic cost of climate policy. Hence, to make the 1.5°C-consistent or carbon-neutral target attainable, explicit emission mitigation and clean-energy development goals at the industrial level in both the coming 14th Five-Year Plan and mid-term strategy are needed. Policy costs should also be paid central attention. This research suggests the need for an in-depth study of China's multimodel comparisons. Models need to improve their representation of technology, including updating the technical details (e.g., model harmonization based on the rapid decline of renewables' cost) and putting in place the mechanisms to sufficiently take the latest energy and climate policies into account (e.g., policy adjustments associated with China's carbon neutrality). In addition, models need to make a firm commitment to quantify the SCC by monetarizing the marginal damage of carbon emissions, which could reduce the uncertainty of policy cost evaluation (*29*, *30*).

REFERENCE AND NOTES

1. Intergovernmental Panel on Climate Change (IPCC), "Climate change 2013: The physical science basis. Working group I contribution to the fifth assessment report," T. F. Stocker *et al*., Eds. (IPCC, 2014).
2. N. S. Diffenbaugh, M. Burke, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9808–9813 (2019).
3. W. Pizer *et al*., *Science* **346**, 1189–1190 (2014).
4. Z. Vrontisi *et al*., *Environ. Res. Lett.* **13**, 044039 (2018).
5. Climate Vulnerable Forum, "Statement of the CVF chair, Philippines. UNFCCC COP21 ministerial dialogue on the long-term goal" (2015); www.thecvf.org/wp-content/uploads/2015/12/Statement-of-PH-CVF-Chair-Ministerial-Dialogue-LTG-08-Dec-15-COP21.pdf.
6. C. Schleussner *et al*., *Nat. Clim. Chang.* **6**, 827–835 (2016).

7. R. Warren, J. Price, J. VanDerWal, S. Cornelius, H. Sohl, *Clim. Change* **147**, 395–409 (2018).
8. C. Mora *et al*., *Nat. Clim. Chang.* **8**, 1062–1071 (2018).
9. F. Pretis, M. Schwarz, K. Tang, K. Haustein, M. R. Allen, *Philos. Trans. R. Soc. London Ser. A* **376**, 20160460 (2018).
10. W. L. Hare, W. Cramer, M. Schaeffer, A. Battaglini, C. C. Jaeger, *Reg. Environ. Change* **11**, 1–13 (2011).
11. A. Jordan *et al*., *Clim. Policy* **13**, 751–769 (2013).
12. J. Nieto, O. Carpintero, L. J. Miguel, *Ecol. Econ.* **146**, 69–84 (2018).
13. G. Peters, *Nat. Clim. Chang.* **6**, 646–649 (2016).
14. B. M. Sanderson *et al*., *Earth Syst. Dynam.* **8**, 827–847 (2017).
15. IPCC, "Special report on global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty," V. Masson-Delmotte *et al*., Eds. (IPCC, 2018).
16. A. F. Hof *et al*., *Environ. Sci. Policy* **71**, 30–40 (2017).
17. J. Rogelj *et al*., *Nat. Clim. Chang.* **5**, 519–527 (2015).
18. O. Hoegh-Guldberg *et al*., *Science* **365**, eaaw6974 (2019).
19. H. B. Duan, G. P. Zhang, S. Y. Wang, Y. Fan, *Environ. Res. Lett.* **14**, 033001 (2019).
20. China Coal-Control Project (CCCP), "Energy scenario analysis and feasibility to achieve the 1.5°C warming-rise target" (China Coal Consumption Cap Plan and Policy Research Project, 2018)
21. H. B. Duan, J. L. Mo, Y. Fan, S. Y. Wang, *Energy Econ.* **70**, 45–60 (2018).
22. D. P. van Vuuren *et al*., *Nat. Clim. Chang.* **8**, 391–397 (2018).
23. A. Bows-Larkin *et al*., *Carbon Manag.* **5**, 193–210 (2014).
24. A. Méjean, C. Guivarch, J. Lefèvre, M. Hamdi-Cherif, *Energy Effic.* **12**, 441–462 (2019).
25. W. D. Nordhaus, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1518–1523 (2017).
26. S. J. Davis *et al*., *Science* **360**, eaas9793 (2018).
27. F. E. L. Otto, D. J. Frame, A. Otto, M. R. Allen, *Nat. Clim. Chang.* **5**, 917–920 (2015).
28. K. Calvin *et al*., *Energy Econ.* **34**, S251–S260 (2012).
29. K. Gillingham *et al*., *J. Assoc. Environ. Resour. Econ.* **5**, 791–826 (2018).
30. M. Sugiyama *et al*., *Energy* **167**, 1120–1131 (2019).

ACKNOWLEDGMENTS

We thank the China Energy Modeling Forum (CEMF) for organizing the second multimodel comparison exercise; the experts who substantially commented on the draft are also acknowledged. **Funding:** H.D., S.Z., and S.W. acknowledge financial support from the National Natural Science Foundation of China (71874177, 72022019, 71874096, and 71988101) and the National Key Research and Development Program of China (2020YFA0608603). S.F. acknowledges funding from The Environment Research and Technology Development Fund (JPMEERF20202002) of the Environmental Restoration and Conservation Agency of Japan. The results and conclusions in this research are those of the authors and do not necessarily represent the official position of the granting organizations. **Author contributions:** H.D. designed and produced an initial draft of this research, with substantial input from all authors. K.J. and S.W. coordinated and co-designed the research. S.Z., M.T., C.B., E.K., M.H., D.P.v.V., S.F., M.T., K.K., G.I., and J.E. contributed to the modeling results, comparison analysis, and manuscript preparation. X.M. and K.K. designed the figures and added supplementary analysis, respectively. All the authors were involved in interpreting the results and revising the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Scenario data from global integrated assessment models were collected in the ADVANCE project, which received funding from the European Union's Seventh Programme FP7/2007-2013 under grant agreement no. 308329 (ADVANCE) and is accessible at https://db1.ene.iiasa.ac.at/ADVANCEDB/dsd?Action=htmlpage&page=welcome. The license under which the ADVANCE Synthesis scenarios are made available is adapted from the Creative Commons Attribution 4.0 International Public License, which keeps the licensed material always up-to-date and avoids the circulation of obsolescent data constituting substantial portions of the licensed material. Data that are not reported in this database are provided by the specific models, and all the model results, particularly from the models that are not participants of the ADVANCE project, are available at https://github.com/xmxming/China-1.5-degree-research. Codes and details for all the participating models can be provided upon request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/378/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*31*–*54*)

13 January 2020; accepted 22 March 2021
10.1126/science.aba8767

NEURODEVELOPMENT

Conserved genetic signatures parcellate cardinal spinal neuron classes into local and projection subsets

Peter J. Osseward II[1,2]†, Neal D. Amin[1]‡, Jeffrey D. Moore[3], Benjamin A. Temple[1,2], Bianca K. Barriga[1,4], Lukas C. Bachmann[1], Fernando Beltran Jr.[1], Miriam Gullo[1], Robert C. Clark[1], Shawn P. Driscoll[1], Samuel L. Pfaff[1]*, Marito Hayashi[1]†§

Motor and sensory functions of the spinal cord are mediated by populations of cardinal neurons arising from separate progenitor lineages. However, each cardinal class is composed of multiple neuronal types with distinct molecular, anatomical, and physiological features, and there is not a unifying logic that systematically accounts for this diversity. We reasoned that the expansion of new neuronal types occurred in a stepwise manner analogous to animal speciation, and we explored this by defining transcriptomic relationships using a top-down approach. We uncovered orderly genetic tiers that sequentially divide groups of neurons by their motor-sensory, local-long range, and excitatory-inhibitory features. The genetic signatures defining neuronal projections were tied to neuronal birth date and conserved across cardinal classes. Thus, the intersection of cardinal class with projection markers provides a unifying taxonomic solution for systematically identifying distinct functional subsets.

From a functional perspective, spinal neurons can be divided along several axes, including motor-sensory, excitatory-inhibitory, and locally connected neurons for intracord processing versus projection neurons for communication with the brain (*1*–*6*). Neuron heterogeneity is often characterized on the basis of landmarks such as their neurotransmitter type, connectivity, cytoarchitecture, morphology, physiology, developmental origin, and genetic profile (*7*). In the spinal cord, separate cardinal neuron classes arise from molecularly distinct progenitor cells arrayed along the dorsoventral axis of the neural tube (*2*, *8*). Neurons within each cardinal class share properties such as the same neurotransmitter identity and have been targeted in functional studies (*1*, *2*). However, it is increasingly apparent that each cardinal class is itself composed of heterogeneous populations of neuron types, impeding a detailed understanding how spinal circuits function (*9*–*13*). We sought to determine whether there was a coherent logic for spinal neuron diversification in mice that could be used to systematically describe the heterogeneity within all the cardinal classes.

One perspective of neuronal diversity is that it arose through evolution in order to expand neural functions (*14*–*16*). In this view, primitive neuron types served as the precursors for more specialized subtypes, leading to the prediction that individual neuronal attributes may have emerged in a stepwise fashion with ordered hierarchical relationships (*14*, *16*, *17*). This analogy to animal speciation prompted us to investigate whether molecular and cellular correlates for this type of stepwise diversification could be detected among large heterogeneous populations of spinal neurons. Thus, rather than focusing on fine-grained molecular differences that may or may not correspond to functional cellular features, we used a top-down approach to identify the transcriptomic signatures linked to the emergence of different neuronal types. At each branch point in our established hierarchy, we defined the

[1]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [2]Neurosciences Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92037, USA. [3]Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, MA, USA. [4]Biological Sciences Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92037, USA.
†These authors contributed equally to this work.
‡Present address: Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA. §Present address: Howard Hughes Medical Institute, Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.
***Corresponding author. Email: pfaff@salk.edu**

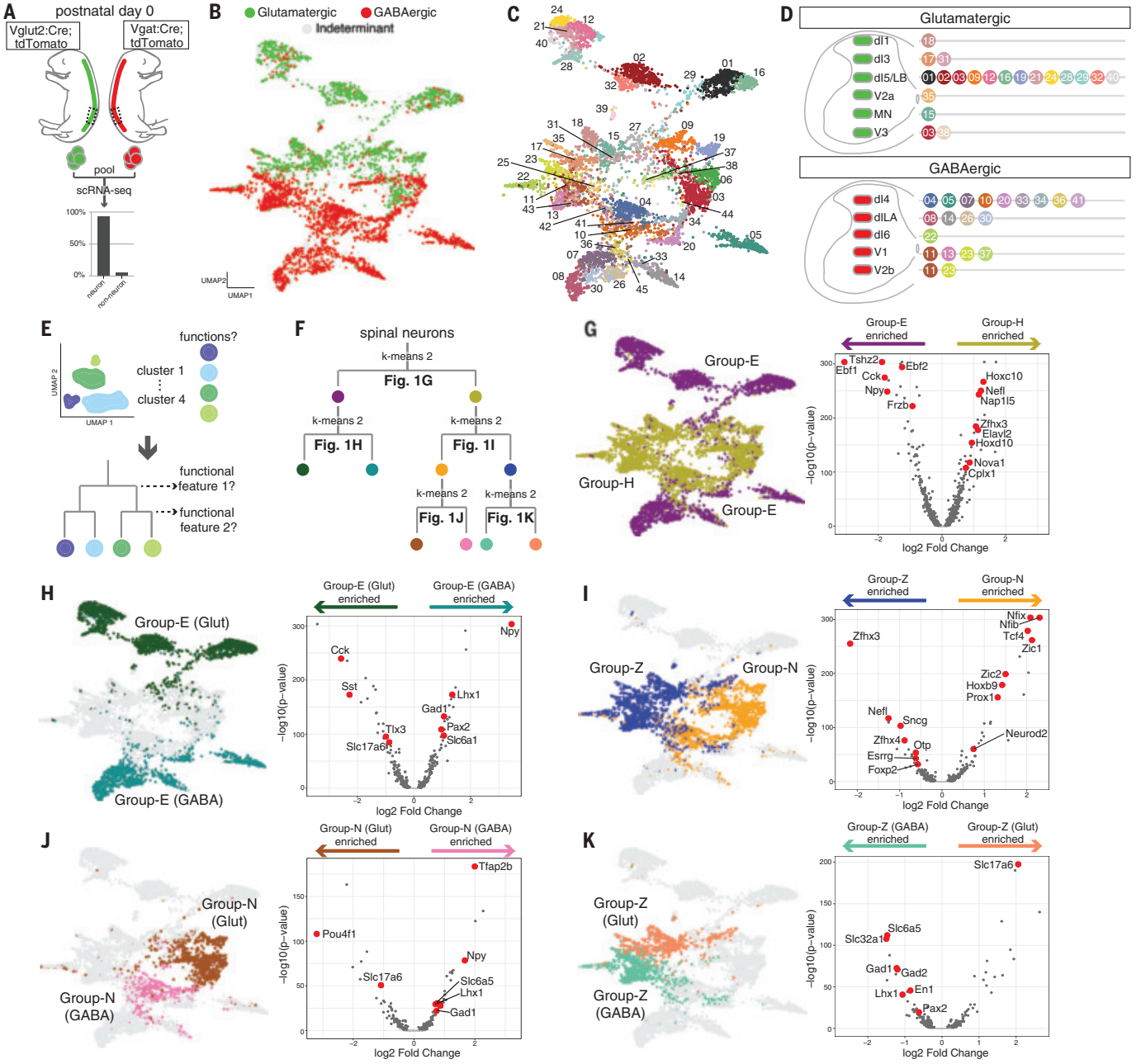


Fig. 1. scRNA-seq identifies relationships among groups of spinal neurons. (**A**) Lumbar segments from P0 Vglut2:Cre; Ai14 and Vgat:Cre; Ai14 neonates (two females and one male from each litter) were microdissected. tdTomato$^+$ cells were sorted and pooled as one sample before scRNA-seq. Using marker analyses and quality control measures, 6743 cells were identified as neurons (94.4% of total cells detected). These neurons had 1012 median genes per cell and 1606 median unique molecular identifiers (UMIs) per cell. (**B**) UMAP visualization of neurons identifies that glutamatergic neurons and GABAergic neurons largely segregate in UMAP space. (**C**) Forty-five identified clusters are color coded in UMAP space. (**D**) Colored dots with numbers correspond to clusters from Fig. 1C. As expected, clusters corresponded to known cardinal neuron classes, defined by their marker expression. V1 and V2b markers were found in the same clusters (*12*). Several clusters are not shown because they were not enriched for obvious cardinal class marker genes (table S1). (**E**) Schematic outlining methods to identify neuronal subtypes. (Top) Dimensionality reduction tools can be used to define the maximum number of genetically distinct clusters. Functions must be independently defined for each cluster. (Bottom) Hierarchical comparisons of neuronal populations can be used to define branch points. The functional differences at the branch points can be used to identify increasingly more specific neuronal attributes linked to function. (**F**) Schematic of analysis used to investigate the higher-order relationships across spinal neuron populations conducted with iterative *k*-means 2 divisive clustering. Data for each color-coded division are indicated for (G) to (K). (**G** to **K**) (Left) UMAP color coded according to identified division. (Right) Volcano plot depicting differentially expressed genes across each *k*-mean division. [(G) and (I)] Further analysis of highlighted genes are available in Fig. 2 and fig. S2. [(H), (J), and (K)] Glutamatergic- and GABAergic-related genes are highlighted. (G) Group-E: 3360 cells, Jaccard similarity [across bootstraps (supplementary materials, materials and methods)] = 0.960; group-H: 3383 cells, Jaccard similarity = 0.959. (H) Group-E (Glut): 1851 cells, Jaccard similarity = 0.960; group-E (GABA): 1509 cells, Jaccard similarity = 0.922. (I) Group-Z: 1707 cells, Jaccard similarity = 0.867; group-N: 1676 cells, Jaccard similarity = 0.809. (J) Group-N (Glut): 1040 cells, Jaccard similarity = 0.791; group-N (GABA): 636 cells, Jaccard similarity = 0.613. (K) Group-Z (Glut): 768 cells, Jaccard similarity = 0.769; group-Z (GABA): 939 cells, Jaccard similarity = 0.760.

molecular, developmental, cytoarchitectural, neurotransmitter, and connectivity properties of the neurons. Our analysis uncovered molecular markers for local and projection neurons regardless of cardinal class or neurotransmitter. By combining markers for cardinal classes with those for conserved genetic signatures for projection status, we established a simple combinatorial matrix that systematically identified discrete subsets of spinal neurons.

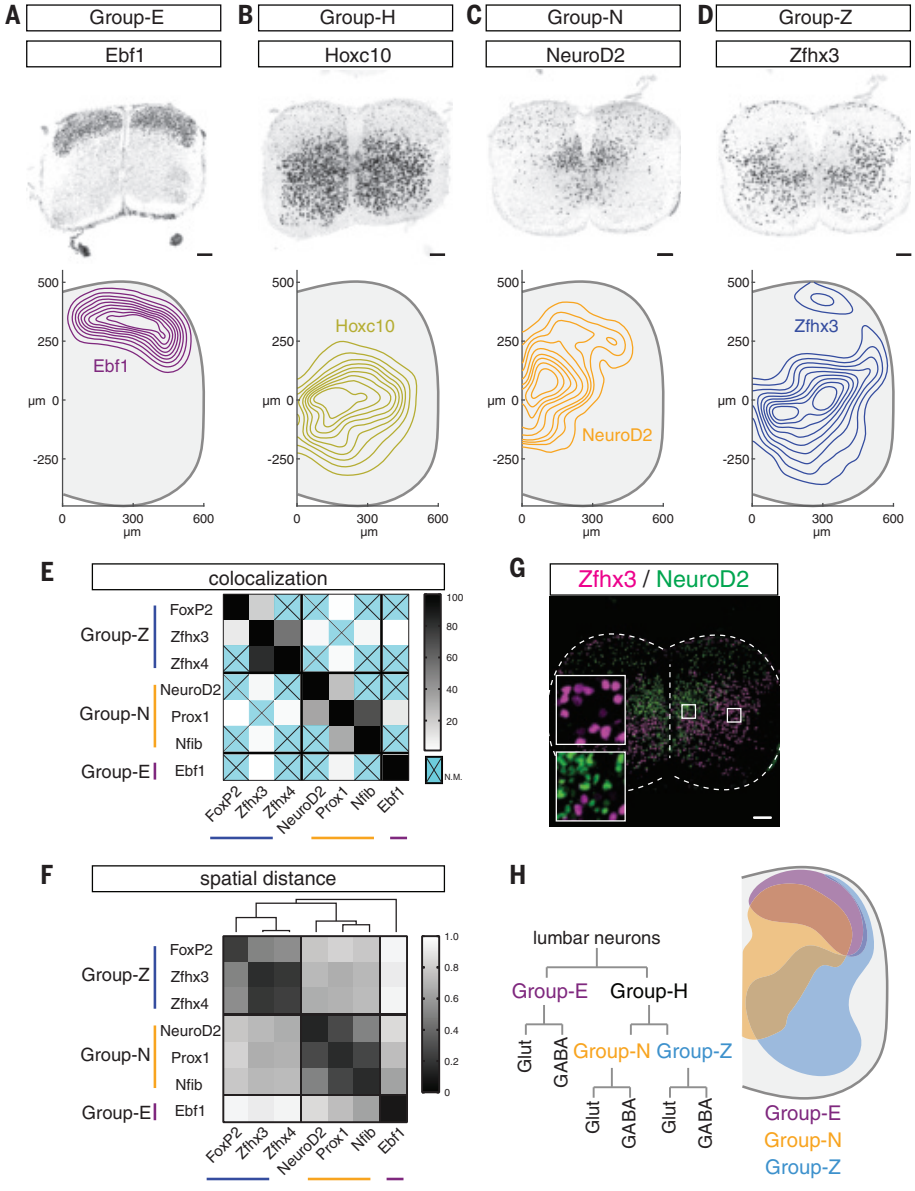


Fig. 2. Branch points in the hierarchical comparisons define spinal neuron groups with discrete locations. (**A** to **D**) (Top) Immunostaining against representative markers for group-E, -H, -N, and -Z neurons in lumbar segments at P0. All images are 20-μm cryosections. Scale bars, 100 μm. (Bottom) Spatial analysis of representative markers displayed on right half of lumbar spinal cord. Contours indicate density of representative markers at the 10th to 90th percentiles. *y* axis = 0 represents the top end of the central canal, and *x* axis = 0 represents the midline. (**E**) Matrix of pairwise mean colocalization rates between group-Z, -N, and -E markers. "N.M." (blue boxes with X) indicates not measured because of antibody incompatibility (figs. S4 and S5). NeuN$^+$ Nfib$^+$ neurons were quantified to exclude Nfib$^+$ glial cells. (**F**) Matrix of pairwise weighted Jaccard distance between group-Z, -N, and -E markers. 1.0, no overlap; 0, complete overlap. Hierarchical clustering dendrogram reveals closer spatial associations among marker genes within each group than between separate groups. (**G**) Immunostaining of Zfhx3 and NeuroD2 in lumbar spinal cord at P0 reveals low incidences of colocalization. Cryosection, 20 μm. Scale bar, 100 μm. This image is a composite of Fig. 2, C and D. (**H**) (Left) Transcriptomic hierarchy of lumbar spinal neurons. (Right) Schematic of spatial enrichments of group-Z, -N, and -E neurons.

Group-E and -H neurons divide the spinal motor-sensory circuitry

To enrich for neurons while minimizing technical and biological variability, we genetically labeled glutamatergic and γ-aminobutyric acid (GABA)–ergic neurons in postnatal day 0 (P0) mice, sorted lumbar spinal neurons separately, pooled them, and performed single-cell RNA-sequencing (scRNA-seq) on the combined sample (Fig. 1A and fig. S1). At P0, spinal neurons have been generated and the basic functional features of adult spinal circuitry have been established, but expression persists for many developmental markers of neuronal subtype diversity (*2*, *9*, *10*, *18*). Using marker analyses, 6743 cells passed quality control and were identified as neurons (Fig. 1A). Glutamatergic and GABAergic neurons largely segregated in the uniform manifold approximation and projection (UMAP) space, and a graph-based comparison using the highly variable genes conservatively identified 45 distinct clusters (Fig. 1, B and C). As expected, many of these clusters were enriched for genes that have been previously identified as marker genes for cardinal classes (Fig. 1D and table S1) (*1*, *2*).

To explore the conserved relationships that exist among lumbar spinal neurons, we developed a divisive clustering pipeline using the *k*-means algorithm to split spinal neurons into pairwise groups iteratively on the basis of their overall transcriptional composition (Fig. 1, E and F). We focused on transcription factors because they regulate neuronal identity and function (*2*, *14*, *15*). We anticipated that this approach would first reveal the glutamatergic and GABAergic neurotransmitter division in the cell population because it is such a prominent phenotypic difference, and the two cell types were isolated from separate litters. However, the neuronal populations split along an axis unrelated to neurotransmitter identity (Fig. 1, B and G) (*11*, *19*). Differential gene expression of the first division indicated that transcription factor Ebf1 is highly expressed in a cell group we defined as group-E neurons, whereas transcription factor Hoxc10 is enriched in a different group we termed group-H neurons (Fig. 1G). Histological analysis of Ebf1$^+$ and Hoxc10$^+$ neurons identified a spatial division of these groups (Fig. 2, A and B, and fig. S2). Ebf1$^+$ neurons were located in laminae I to III of the superficial dorsal horn, corresponding to the site where many exteroceptive sensory inputs terminate. By contrast, Hoxc10$^+$ neurons were predominantly located in laminae IV to X, areas involved in proprioception and motor control (*3*, *4*). Although group-E and group-H neurons broadly differ in motor-sensory functions associated with their positions, both types are composed of a mixture of glutamatergic and GABAergic neurons (Fig. 1, B and G, and fig. S3) (*3*, *4*).

Group-N and -Z neurons are arrayed along the mediolateral axis of the spinal cord

Using the hierarchical-divisive strategy outlined above (Fig. 1F), we further split the group-E neurons into two subgroups and found that this division identifies glutamatergic-related and GABAergic-related genes and their corresponding neurons (Fig. 1, B and H) (*2*, *4*). Division of the group-H population produced a different result. Rather than splitting group-H neurons into glutamatergic and GABAergic cell populations, we found a subgroup of neurons enriched for Nfib, NeuroD2, and Prox1 (designated group-N) and a separate subgroup enriched for Zfhx3, Zfhx4, and FoxP2 (group-Z)

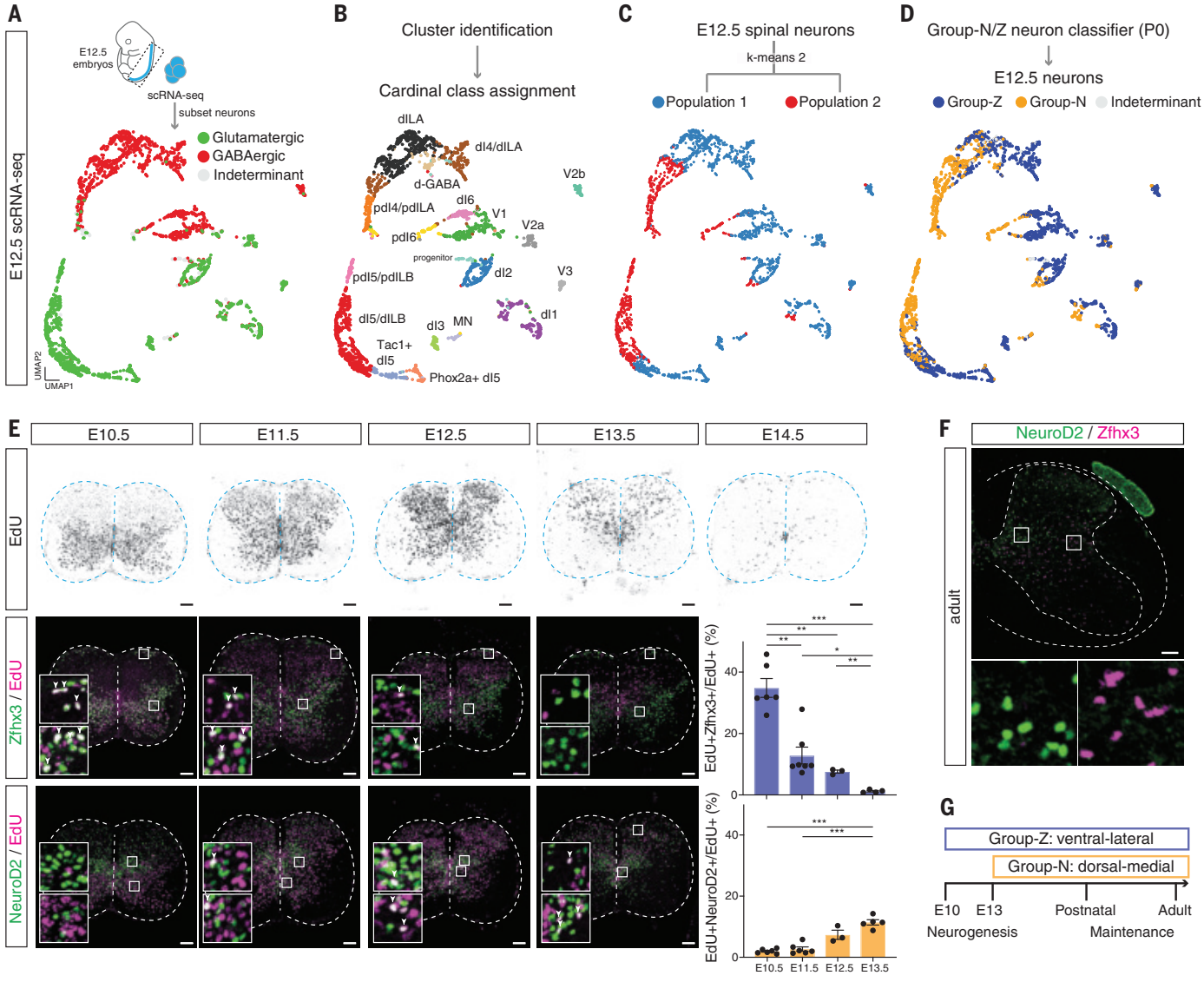


Fig. 3. The group-N and -Z division is linked to neuronal birth date. (**A**) Cells along the entire spinal cord were isolated from two E12.5 embryos from two litters and were subjected to scRNA-seq as separate samples. These datasets were merged post hoc. We bioinformatically identified 2945 neurons, which were selected for further analysis (fig. S8). These neurons had 3566 median genes per cell and 12,189 median UMIs per cell. Glutamatergic and GABAergic neurons are labeled on the UMAP. (**B**) Cardinal classes segregate in UMAP space. Twenty-seven clusters were identified (fig. S8). On the basis of the top 20 differentially expressed genes for each cluster, cardinal classes were identified. Clusters from the same cardinal class were merged (table S3). (**C**) Higher-order groups of spinal neurons were identified with *k*-means 2 divisive clustering (Fig. 1F). Embryonic neurons are color coded according to the first *k*-means 2 division. Population 1: 1971 cells, Jaccard similarity = 0.978; population 2: 974 cells, Jaccard similarity = 0.955. Further divisions are provided in fig. S9. (**D**) The first *k*-means 2 division of embryonic neurons largely corresponded to group-N and group-Z identity established from P0: 79% of population 1 [(C), blue cells] neurons expressed group-Z markers, and 84% of population 2 [(C), red cells] neurons expressed group-N markers. Thus, the *k*-means 2 division identifies group-N and group-Z neurons at an early embryonic stage. (**E**) EdU was injected into multiple pregnant females (E10.5 to E14.5), and lumbar segments were collected from their pups at P0. Representative markers Zfhx3 and NeuroD2 were used to birth date group-Z and group-N neurons, respectively. The peak birth date for group-Z neurons was E10.5, whereas the peak for group-N neurons was E13.5. Each dot represents the colocalization rate for a single animal. Welch one-way analysis of variance (ANOVA) comparing every time point to every other time point, with Dunnett T3 correction for multiple comparisons. $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. Cryosections, 20 μm. Scale bars, 100 μm. (**F**) Immunostaining of Zfhx3 and NeuroD2 in lumbar spinal cord at P70. The mediolateral location and marker expression patterns in adult P70 animals were similar to P0. Cryosection, 25 μm. Scale bar, 100 μm. (**G**) Summary. Group-Z neurons are born before group-N neurons. Molecular markers for both groups are expressed embryonically, postnatally, and into adulthood. Markers for group-Z and group-N neurons are not coexpressed.

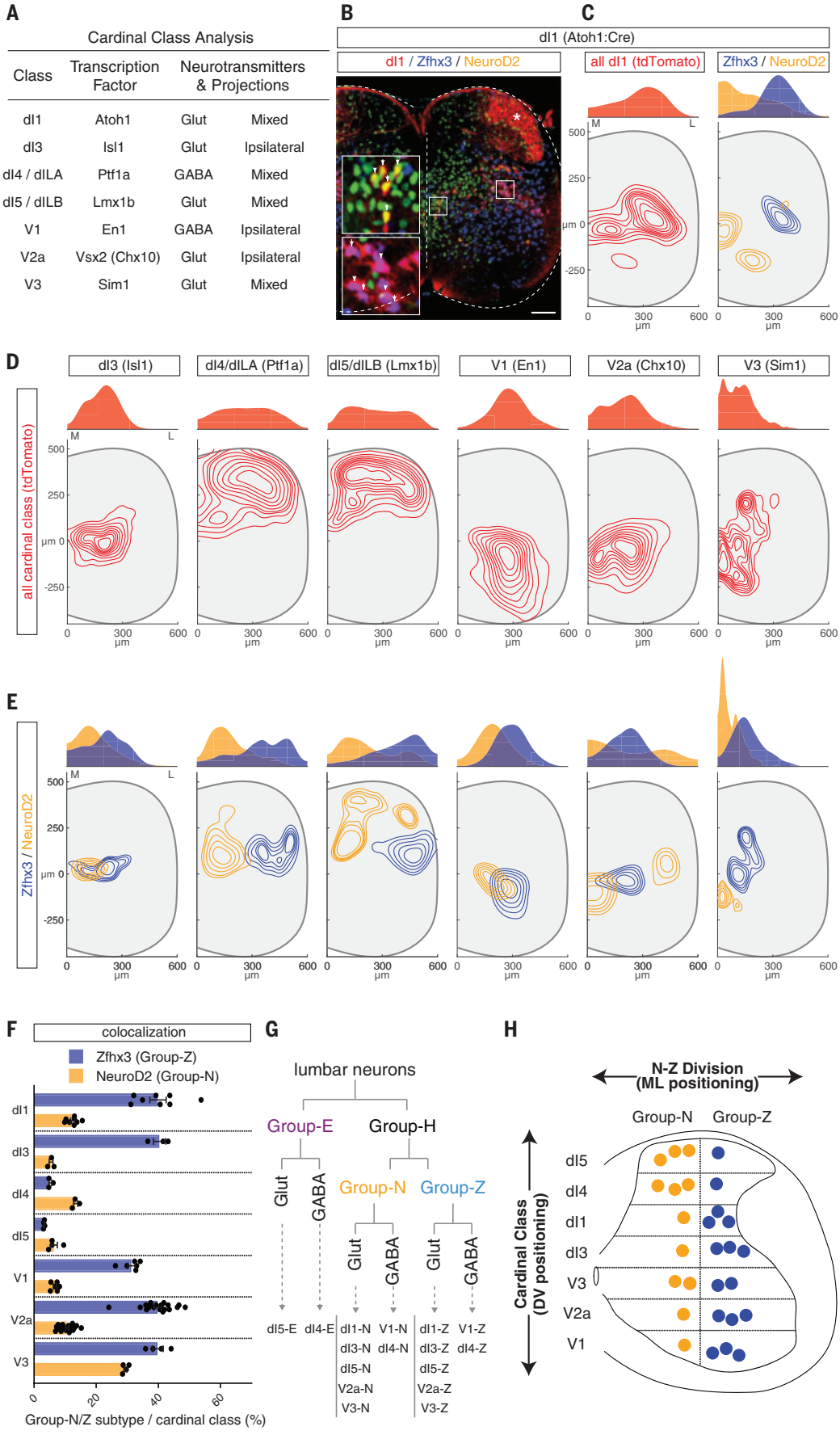


Fig. 4. Each cardinal class is a mixture of both group-N and -Z neurons. (**A**) Summary of mouse lines used to label cardinal classes. Each cardinal class has a characteristic transcription factor expression, neurotransmitter, and axon laterality profile. (**B**) Atoh1:Cre was crossed to a tdTomato reporter to indelibly label dI1 neurons. Immunostaining for NeuroD2 (group-N) and Zfhx3 (group-Z) reveals discrete dI1 populations marked by the two genes. Asterisk in the dorsal horn indicates ectopic sensory fibers from the DRG (*29*). Cryosection, 20 μm. Scale bar, 100 μm. (**C**) (Left) Contours indicate density of dI1 neurons at the 10th to 90th percentiles. Mediolateral density displayed above a two-dimensional contour plot. (Right) Contours indicate densities of group-N and group-Z dI1 neurons at 50th to 90th percentiles. Group-N dI1 neurons were generally located medially, whereas group-Z dI1 neurons were located laterally. (**D**) Contours and mediolateral densities of indicated cardinal classes. Contours indicate density of cardinal class neurons at the 10th to 90th percentiles. Dorsoventral densities are provided in fig. S14. (**E**) Distributions of group-N and group-Z neurons within each cardinal class. Contours indicate density of interneuron subpopulations at the 50th to 90th percentiles. Each cardinal class was split into medial group-N and lateral group-Z neurons. (**F**) Quantification of group-Z and group-N neurons within each cardinal class. Each dot indicates one animal. The ratio of group-N to -Z neurons varies across cardinal classes. (**G**) Summary of the cardinal class types within the hierarchical comparisons of neurons. (**H**) Stylized overview of cell spatial relationships defined by coexpression of cardinal and N-Z markers. Cardinal classes have stereotyped dorsoventral settling positions. Neurons composing each cardinal class are split into medial group-N and lateral group-Z cells.

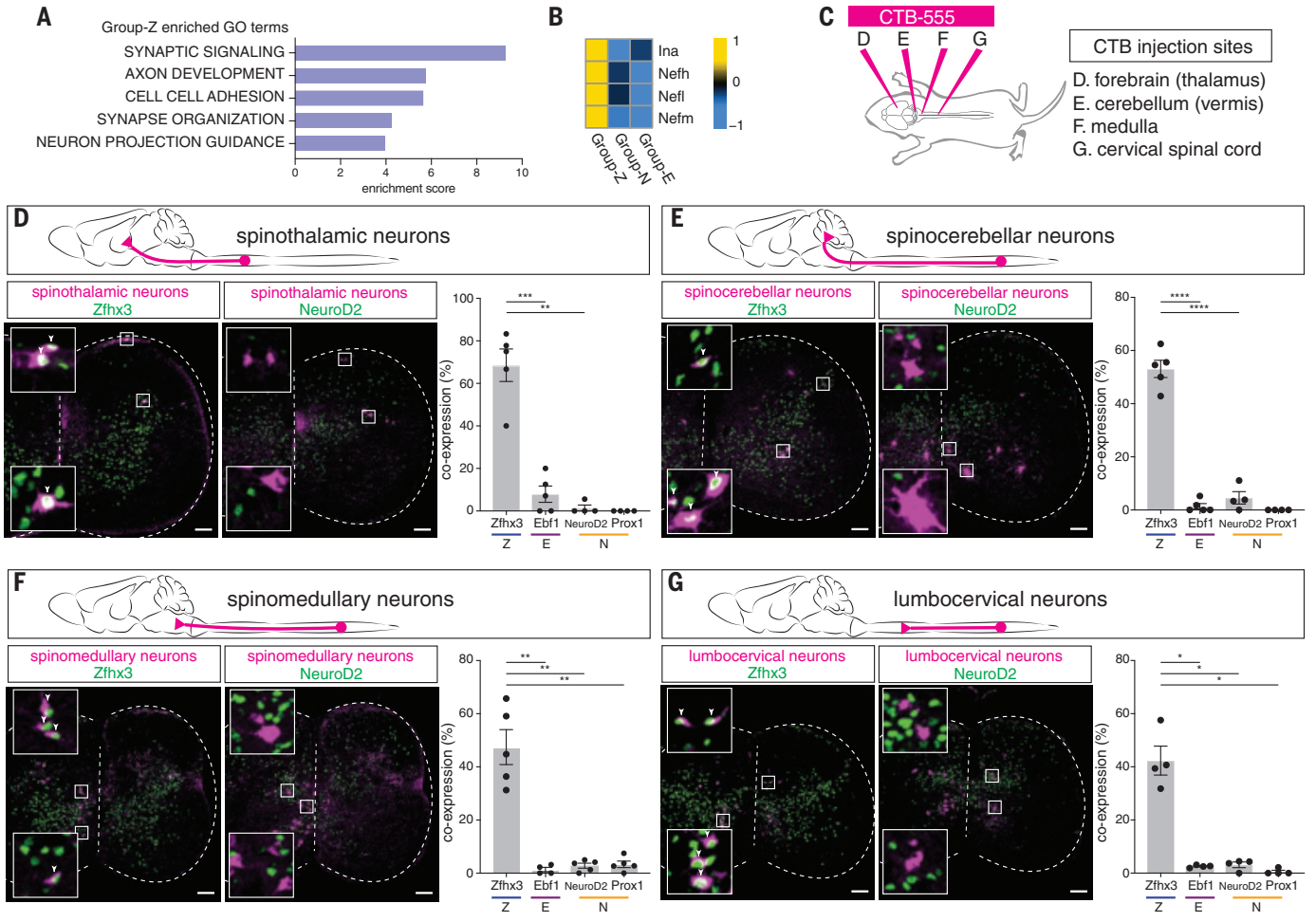


Fig. 5. Projection and long-range neurons express group-Z markers. (**A**) Neuronal morphology-related GO terms are enriched in group-Z compared with group-N neurons (table S4). (**B**) Heatmap of neurofilament genes expressed in spinal neurons showing enrichment in group-Z neurons (log2 scale). (**C**) Schematic of retrograde tracing analysis. CTB was injected into the forebrain (thalamus), cerebellum (vermis), medulla, and cervical spinal cord, respectively, to retrogradely label long-range projection neurons within the spinal cord. Injections were done in neonates, and tissue was collected 4 to 5 days later. (**D** to **G**) Immunostaining against group-Z (Zfhx3), -E (Ebf1), and -N (NeuroD2 and Prox1) markers combined with indicated CTB labeling. White arrowheads indicate group-Z long-range neurons. Each dot on the graph indicates the colocalization rate of one animal. Cryosections, 20 μm. Scale bars, 100 μm. Welch one-way ANOVA comparing Zfhx3 to every other marker, with Dunnett T3 correction for multiple comparisons. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$. (D) Spinothalamic neurons within cervical segments (region of greatest abundance) expressed group-Z markers. Spinothalamic neurons within lumbar segments are shown in fig. S17. (E) Spinocerebellar neurons within lumbar segments expressed group-Z markers. Additional markers are provided in fig. S18. (F) Spinomedullary neurons within lumbar segments expressed group-Z markers. Additional markers are provided in fig. S19. (G) Cervicolumbar neurons expressed group-Z markers. Additional markers are provided in fig. S20.

(Fig. 1I and table S2). Histological analysis found that these two sets of genes spatially divided the spinal cord into large wedge patterns: group-N cells located dorsomedially and group-Z cells positioned ventrolaterally (Fig. 2, C and D, and fig. S4). Group-N and group-Z neurons occasionally intermingle (Fig. 2, C to H, and fig. S5), but few cells coexpressed markers for both group-N and group-Z neurons (Fig. 2, E and G, and fig. S5). Although group-N and group-Z neurons are defined by gene sets, we found that NeuroD2 and Zfhx3 were representative markers of group-N and -Z neurons, respectively (Figs. 1I and 2, C and D; and figs. S2 to S5). Our sequencing was conducted on cells from the lumbar spinal cord, but we found that NeuroD2 and Zfhx3 labeled neurons in the same mediolateral pattern along the entire rostrocaudal axis of the cord (figs. S6 and S7). Further *k*-means splitting of group-N and -Z neurons resolved the glutamatergic and GABAergic subsets in each population (Fig. 1, J and K, and fig. S3) (*2*, *4*). Our findings unmask a stepwise hierarchy of genetic divisions among spinal neurons with layers that supersede neurotransmitter physiology (Fig. 2H).

Group-N and -Z neurons have distinct birth dates

Although the group-E–group-H division appeared to correspond to neurons involved in exteroception and motor control, respectively, we hypothesized that the group-N and -Z subsets derived from the group-H population might segregate along alternative features. To investigate a possible developmental distinction between group-N and group-Z neurons, we performed scRNA-seq of embryonic day 12.5 (E12.5) spinal cord cells composed of the embryonically defined cardinal cell classes (Fig. 3, A and B; fig. S8; and table S3). The E-H division was not evident at E12.5; however, the group-N and group-Z neuronal populations were detected (Fig. 3, C and D, and figs. S8 and S9). We birth-dated the neurons marked by NeuroD2 and Zfhx3 at P0 using EdU pulses embryonically. The peak of Zfhx3$^+$ neuron birth was at E10.5, whereas the peak of NeuroD2$^+$

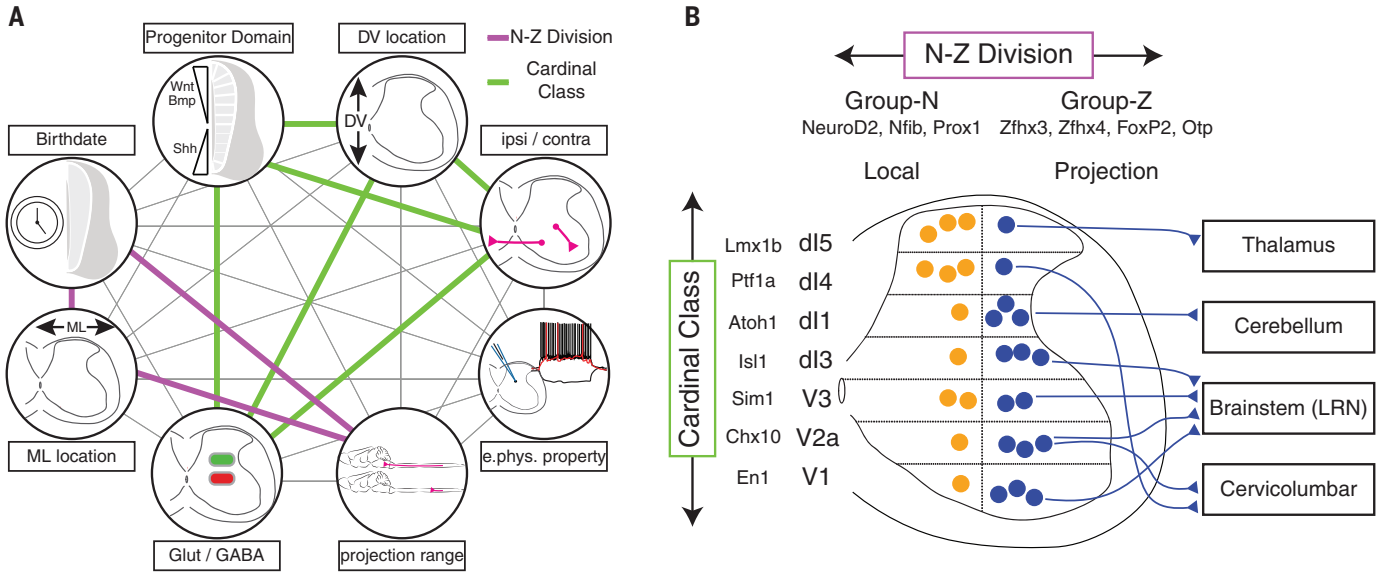


Fig. 6. A unified taxonomic index of spinal neurons. (**A**) Summary of relationships of neuronal attributes. Neuronal subtypes can be viewed as a composite of distinct properties that fall into a range of categories (circles). Neuronal subtypes arise by joining different combinations of these attributes. Cardinal class and N-Z markers simplify the task of defining neuronal subtypes because both correspond to multiple neuronal features (magenta and green lines). (**B**) Summary of neuronal subsets defined by coexpression of cardinal class markers and N-Z labels. This matrix provides a simple labeling index to define neuronal subsets. The postsynaptic targets of group-Z neurons within each cardinal population are predicted from projection neuron–tracing studies (*26*, *29*, *31*, *37*, *38*). Local connections and collaterals are not shown (*26*, *27*).

neuron generation was later at E13.5 (Fig. 3E and fig. S10). To examine whether group-N and group-Z identity is dependent on neuronal connectivity, we genetically ablated proprioceptive sensory neurons or motor neurons. Despite the absence of sensory and motor neurons, Zfhx3 and NeuroD2 expression patterns were unaltered (fig. S11). Furthermore, we found that Zfhx3 and NeuroD2 expression was maintained in adult spinal neurons (P70), approximating the labeling pattern observed at P0 (Fig. 3F). Taken together, our findings indicate that the group-N and -Z populations have distinct timings of neurogenesis and are present embryonically through adulthood (Fig. 3G).

Group-N and -Z neurons divide each cardinal neuron class into subsets

We set out to understand how group-E, group-N, and group-Z neurons identified here spatially and molecularly intersect with the previously characterized cardinal classes. Cre lines labeling the cardinal classes dI1, dI3, dI4/dILA, dI5/dILB, V1, V2a, and V3 were immunostained for Ebf1 (group-E), NeuroD2 (group-N), and Zfhx3 (group-Z) (Fig. 4A). Group-E neurons were found to be restricted to the dI4/dILA and dI5/dILB classes (figs. S12 and S13), whereas all tested cardinal classes were composed of a mixture of both group-N and group-Z neurons derived from the group-H population (Fig. 4, B to F, and figs. S12 and S13). Analysis of E12.5 scRNA-seq data confirmed that the dI2, dI6, and V2b classes also were a mixture of both group-N and group-Z neurons (fig. S13) (*13*). Moreover, the group-N neurons from each cardinal class were always medial to the group-Z cells (Fig. 4, C to E and H, and fig. S14) (*9*, *10*, *20*). Although each cardinal class contained both group-N and group-Z neurons, the ratio of these types varied within each class (Fig. 4, F and H, and figs. S13 and S15) (*10*, *20*). Group-N and group-Z neurons do not obey cardinal class divisions; rather, all cardinal classes are a mix of group-N and group-Z cells with discrete mediolateral positions. Thus, the group-N–group-Z division is an orthogonal molecular and spatial axis that intersects with cardinal class identity (Fig. 4H).

Local versus long-range projection neuron types underlie the N-Z division

To further understand the cellular features of group-N and group-Z neurons, we explored the connectivity patterns of spinal neurons. We first tested whether the group-N and group-Z neurons represented premotor interneurons, using monosynaptic rabies virus tracing. We found that both the group-N and group-Z populations contained ipsi- and contralateral premotor neurons (fig. S16A) (*21*–*23*). Next, using immunostaining and mouse genetics to visualize presynaptic terminals, we found that both group-N and group-Z neurons receive a variety of inputs (for example, dorsal root ganglia sensory, descending serotonergic, dI1, dI4-6, V1, and V3) (fig. S16, B and C) (*24*). Gene Ontology (GO) term analysis revealed that group-Z neurons were enriched for genes involved in axon and synapse development, including high neurofilament (Nef-l, -m, and -h) expression (Fig. 5, A and B). Consequently, we speculated that group-Z neurons may contain one or more populations of long-range projection neurons (*25*). Subsets of neurons from multiple cardinal classes have been shown to share projection targets, further raising the possibility that group-Z neurons represent long-range projections across neuronal classes (*7*, *9*, *26*, *27*). We labeled spinothalamic, spinocerebellar, spinomedullary, and lumbocervical projection neurons with retrograde tracer cholera toxin subunit β (CTB) in combination with group-E, -N, and -Z marker immunostaining (Fig. 5C). Although group-N and group-Z neurons are approximately equal in number (fig. S6), the long-range projection neurons in both the lumbar and cervical cord were preferentially enriched for Zfhx3^{+} neurons (Fig. 5, D to G, and figs. S17 to S20). Because the single marker Zfhx3 does not label all group-Z neurons, we also performed the supraspinal labeling and immunostaining with additional group-Z markers, including Zfhx4, Otp, and Foxp2. We found that multiple group-Z markers identify additional long-range projection neurons, providing further evidence that many group-Z cells project long axons (figs. S18 and S19). By contrast, markers for group-N (NeuroD2 and Prox1) and group-E (Ebf1) rarely (0 to 7.9%) labeled long-range neurons (Fig. 5, D to G, and figs. S17 to S20). As a positive control, injection of CTB directly into the cord labeled group-N neurons near the injection site (fig. S20). Taken together, our labeling studies indicate that

group-Z neurons are composed of long-range projecting cells, whereas group-N and group-E neurons appear to be preferentially composed of local interneurons.

Although most of our studies focused on the lumbar spinal cord, we found one counterexample to the group-N–group-Z correlation with axon projection in the thoracic region. Namely, a fraction of spinocerebellar projection Clarke's column neurons was found to express the group-N marker NeuroD2 (fig. S18) (*28*). The reason for this discrepancy is unclear; however, these cells form a distinct medial nucleus with unusual molecular features, such as the expression of Vglut1 rather than Vglut2 (*29*, *30*). Previous gene profiling of Clarke's column neurons also revealed that they cluster separately from other thoracic spinal neurons (*28*). The non–Clarke's column spinocerebellar neurons in the thoracic cord express group-Z genes that are typical for projection neurons (fig. S18) (*28*).

In principle, the correlation between group-Z markers and projection neuron labeling might have arisen as a byproduct of long-range neurons predominantly being located laterally within the spinal cord where group-Z markers are expressed for independent reasons. Therefore, we sought to define the axon projection pattern of group-Z neurons in atypical locations within the spinal cord. A sparse population of Zfhx3$^+$ neurons were situated in the very superficial lamina of the spinal cord (Figs. 2D and 3E). The superficial Zfhx3$^+$ neurons were labeled after injection of CTB into the thalamus (Fig. 5D). These unusual Zfhx3$^+$ projection neurons were further confirmed to represent spinothalamic cells by using Nk1r/Tacr1 labeling and were also born early like other group-Z neurons (Fig. 3E and fig. S17) (*5*, *31*). Near the midline, the boundary between group-N and group-Z is not absolute, and there is sparse intermingling of NeuroD2$^+$ and Zfhx3$^+$ neurons. To further test the relationship between group-N and group-Z markers and their axonal projection patterns, we examined the marker profile of projection neurons in this region of the spinal cord. Retrograde fills of projection neurons preferentially labeled the Zfhx3$^+$ neurons but not the NeuroD2$^+$ neurons (Fig. 5, E to G). Thus, the molecular features that distinguish group-N from group-Z neurons reveal a core phenotypic difference: Group-N (and group-E) neurons are restricted to forming local connections within the spinal cord, whereas group-Z neurons can also form long-range connections with distant targets (Fig. 6).

Discussion

In this study, we sought to identify a unifying logic for spinal neuron diversification. We used a top-down approach to identify the conserved genetic relationships among heterogeneous cell types and used the branch points of this hierarchical system to find marker genes that were linked to neuronal birth date, cytoarchitecture, neurotransmitter, and projection status. The divide between group-E neurons marked by Ebf1 and group-H cells marked by Hoxc10 corresponded to a split between sensory and motor circuits, respectively (Fig. 2H). The H-neuron population was further divided into group-N and -Z types, in which NeuroD2 labeled medial, late born, locally projecting interneurons and Zfhx3 marked lateral, early born, projection neurons (Figs. 2H, 3G, and 6, A and B).

The N-Z division giving rise to local versus projection neurons may have occurred early in evolution because the genetic signatures for these two groups are conserved across broad populations of spinal neurons. Consistent with this, group-N and -Z genes are conserved in many vertebrates, and functional studies of these factors have revealed roles in neural diversification (*32*–*34*). The correlation between birth date and N-Z identity is consistent with neuronal diversity being tied to neurogenesis (*9*, *22*, *35*, *36*).

We anticipated that the most basic characteristic associated with neurons is their neurotransmitter phenotype (*11*, *15*, *19*) but instead found that this was superseded by the distinctions between motor-sensory and local-projection neuron types (Fig. 2H). Motor neurons in different species, although homologous cells, use different neurotransmitters, suggesting that there is flexibility and adaptability in the signals for neurotransmission (*15*). At a molecular level, the same neurotransmitter phenotype across cardinal classes can be specified by different transcriptional pathways (*2*). These findings suggest that spinal neuron neurotransmitter identity can arise through convergent evolutionary processes, causing the genetic signatures for neurotransmitter identity to appear lower in the hierarchical relationships because there are fewer cells with similar genetic characteristics.

Because the lineage-based markers of cardinal interneurons are orthogonal to the conserved markers for the N-Z neuronal sets, we found that the intersection of these two systems systematically labeled neurons with specific dorsoventral and mediolateral positions, birth dates, neurotransmitter phenotypes, ipsilateral-contralateral projections, and local–long-range connectivity (Fig. 6, A and B, and fig. S21). In isolation, neither cardinal class nor N-Z markers precisely define the postsynaptic targets of spinal neurons (*2*, *9*, *10*). However, the spinal neurons innervating the lateral reticular nucleus, thalamus, cerebellum, and long-range spinal targets (propriospinal) can be predicted on the basis of the intersection of their cardinal and N-Z genetic status (Fig. 6B and fig. S21). By identifying markers associated with diversification for the purpose of expanding cellular functions—in this case, evolution and development—it may be possible to create more unified and practical systematics for identifying functionally different cell types in many systems.

REFERENCES AND NOTES

1. D. C. Lu, T. Niu, W. A. Alaynick, *Front. Mol. Neurosci.* **8**, 25 (2015).
2. H. C. Lai, R. P. Seal, J. E. Johnson, *Development* **143**, 3434–3448 (2016).
3. O. Kiehn, *Nat. Rev. Neurosci.* **17**, 224–238 (2016).
4. S. C. Koch, D. Acton, M. Goulding, *Annu. Rev. Physiol.* **80**, 189–217 (2018).
5. A. J. Todd, *Nat. Rev. Neurosci.* **11**, 823–836 (2010).
6. J. Jiang, E. Azim, C. F. Ekerot, B. Alstermark, *Front. Comput. Neurosci.* **9**, 75 (2015).
7. P. J. Osseward 2nd, S. L. Pfaff, *Curr. Opin. Neurobiol.* **56**, 175–184 (2019).
8. J. Ericson *et al*., *Cell* **90**, 169–180 (1997).
9. M. Hayashi *et al*., *Neuron* **97**, 869–884.e5 (2018).
10. J. B. Bikoff *et al*., *Cell* **165**, 207–219 (2016).
11. M. Häring *et al*., *Nat. Neurosci.* **21**, 869–880 (2018).
12. A. Sathyamurthy *et al*., *Cell Rep.* **22**, 2216–2225 (2018).
13. J. Delile *et al*., *Development* **146**, dev173807 (2019).
14. D. Arendt *et al*., *Nat. Rev. Genet.* **17**, 744–757 (2016).
15. O. Hobert, P. Kratsios, *Curr. Opin. Neurobiol.* **56**, 97–105 (2019).
16. D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).
17. H. Zeng, J. R. Sanes, *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
18. D. R. Ladle, E. Pecho-Vrieseling, S. Arber, *Neuron* **56**, 270–283 (2007).
19. B. Tasic *et al*., *Nature* **563**, 72–78 (2018).
20. M. I. Gabitto *et al*., *Cell* **165**, 220–233 (2016).
21. A. E. Stepien, M. Tripodi, S. Arber, *Neuron* **68**, 456–472 (2010).
22. M. Tripodi, A. E. Stepien, S. Arber, *Nature* **479**, 61–66 (2011).
23. A. J. Levine *et al*., *Nat. Neurosci.* **17**, 586–593 (2014).
24. F. J. Alvarez, R. M. Villalba, R. Zerda, S. P. Schneider, *J. Comp. Neurol.* **472**, 257–280 (2004).
25. A. Yuan, M. V. Rao, R. A. Veeranna, R. A. Nixon, *Cold Spring Harb. Perspect. Biol.* **9**, a018309 (2017).
26. C. Pivetta, M. S. Esposito, M. Sigrist, S. Arber, *Cell* **156**, 537–548 (2014).
27. E. Azim, J. Jiang, B. Alstermark, T. M. Jessell, *Nature* **508**, 357–363 (2014).
28. M. Baek, V. Menon, T. M. Jessell, A. W. Hantman, J. S. Dasen, *Cell Rep.* **27**, 2620–2635.e4 (2019).
29. R. Yuengert *et al*., *Cell Rep.* **13**, 1258–1271 (2015).
30. A. W. Hantman, T. M. Jessell, *Nat. Neurosci.* **13**, 1233–1239 (2010).
31. K. Nishida, S. Ito, *Eur. J. Neurosci.* **46**, 2608–2619 (2017).
32. Z. Zhang *et al*., *Exp. Neurol.* **322**, 113055 (2019).
33. M. Tambalo, R. Mitter, D. G. Wilkinson, *Development* **147**, dev184143 (2020).
34. S. A. Wilke *et al*., *Neural Dev.* **7**, 9 (2012).
35. L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, J. D. Macklis, *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
36. C. Q. Doe, *Annu. Rev. Cell Dev. Biol.* **33**, 219–240 (2017).
37. R. B. Roome *et al*., *Cell Rep.* **33**, 108425 (2020).
38. L. Ruder, A. Takeoka, S. Arber, *Neuron* **92**, 1063–1078 (2016).

ACKNOWLEDGMENTS

We thank K. Lettieri, B. O'Leary, H. Forman, A. J. Levine, M. J. Sternfeld, K. L. Hilde, members of the Pfaff laboratory, the FCCF core (C. O'Connor and C. Fitzpatrick; NIH-NCI CCSG: P30 014195), the NGS core (N. Hah; NIH-NCI CCSG: P30 014195, the Chapman Foundation and the Helmsley Charitable Trust), the UCSD IGM Genomic Center (K. Jepsen and other members), the Biophotonics core (NIH-NCI CCSG: P30 014195 and the Waitt Foundation), the GT3 core (J. Naughton, J. Marlett, and R. Armendariz; NIH-NCI CCSG: P30 014195, NINDS R24 Core Grant, funding from NEI), and the Janelia Viral Vector Core for technical support and advice; M. Goulding, G. Gatto, members of the Goulding laboratory, E. J. Kim and E. M. Callaway for providing reagents and advice; and S. Ackerman, K. Asahina, K. Baldwin, T. Komiyama, and Y. Zou for discussions and advice. **Funding:** P.J.O. was supported by the Christopher and Dana Reeve Foundation. N.D.A. is supported by the NIMH T32 Training Grant T32MH019938 J.D.M. is supported by NIH-NICHD K99 HD096512. B.K.B. is supported by the UCSD Cell

and Molecular Genetics Training Program (T32 GM007240) and the Timken-Sturgis Foundation. M.H. was supported by the Japanese Ministry of Education, Culture, Sports, Science, and Technology Long-Term Student Support Program and the Timken-Sturgis Foundation. S.L.P. is a Benjamin H. Lewis chair in neuroscience. This research was supported by funding from the Howard Hughes Medical Institute, the Christopher and Dana Reeve Foundation, the Sol Goldman Charitable Trust, and National Institutes of Health (1 U19 NS112959-01). **Author contributions:** Conceptualization: M.H., P.J.O., and S.L.P. Methodology: M.H., P.J.O., J.D.M., and S.P.D. Software: S.P.D., J.D.M., and P.J.O. Validation: M.H., P.J.O., L.C.B., and R.C.C. Formal analysis: S.P.D., J.D.M., P.J.O., and M.H. Investigation: M.H., P.J.O., N.D.A., B.A.T., B.K.B., L.C.B., F.B., and R.C.C. Resources: M.G., S.P.D., J.D.M., P.J.O., and M.H. Data curation: S.P.D., J.D.M., and P.J.O. Writing, original draft: M.H., P.J.O., and S.L.P. Writing, review and editing: M.H., P.J.O., N.D.A., J.D.M., B.A.T., B.K.B., S.P.D., and S.L.P. Visualization: M.H., P.J.O., and S.P.D. Supervision: M.H., P.J.O., and S.L.P. Project administration: M.H., P.J.O., and S.L.P. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The accession number for the scRNA-seq data reported in this paper is NIH GEO: GSE171607, GSE171605, and GSE171606.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/385/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S5
References (*39*–*52*)
MDAR Reproducibility Checklist

29 July 2020; accepted 15 March 2021
10.1126/science.abe0690

REPORTS

HEAT TRANSPORT

Broadband directional control of thermal emission

Jin Xu, Jyotirmoy Mandal, Aaswath P. Raman*

Controlling the directionality of emitted far-field thermal radiation is a fundamental challenge. Photonic strategies enable angular selectivity of thermal emission over narrow bandwidths, but thermal radiation is a broadband phenomenon. The ability to constrain emitted thermal radiation to fixed narrow angular ranges over broad bandwidths is an important, but lacking, capability. We introduce gradient epsilon-near-zero (ENZ) materials that enable broad-spectrum directional control of thermal emission. We demonstrate two emitters consisting of multiple oxides that exhibit high (>0.7, >0.6) directional emissivity (60° to 75°, 70° to 85°) in the p-polarization for a range of wavelengths (10.0 to 14.3 micrometers, 7.7 to 11.5 micrometers). This broadband directional emission enables meaningful radiative heat transfer primarily in the high emissivity directions. Decoupling the conventional limitations on angular and spectral response improves performance for applications such as thermal camouflaging, solar heating, radiative cooling, and waste heat recovery.

Thermally generated light is a fundamental feature of nature (*1*–*5*). The ubiquity of thermal emission makes its control of central importance to a broad range of technologies, from energy conversion (*6*, *7*) to imaging (*5*, *8*) and sensing (*9*–*11*). Thermal emitters are, in general, incoherent sources that lack directionality. However, high thermal emission at unwanted directions can lead to low efficiency and effectiveness for many thermal devices and applications (*12*). Controlling the angular response of thermal emission has thus been investigated using a range of strategies (*13*), including surface plasmon polariton gratings and metasurfaces (*14*–*17*), phonon polariton–based gratings (*18*), or photonic crystals (*14*). Although these past approaches have demonstrated anomalous angular control of thermal emission, this has been over narrow bandwidths, with the direction of emission varying as a function of wavelength. Thermal radiation, however, is intrinsically a broadband phenomenon. The ability to constrain broad-spectrum thermal emission to a fixed, narrow set of directions would be a fundamentally enabling capability (*19*), as it would facilitate high radiance and heat transfer in terms of bandwidth while restricting this heat transfer to specific angular ranges. This capability is of particular relevance to a range of sensing and energy applications, including thermal imaging (*20*), thermophotovoltaics (*21*, *22*), and radiative cooling (*23*–*25*), where avoiding directional noise and parasitic heat gain or loss can result in substantial gains in device efficiency.

Broadband angular selectivity in the propagation and absorption of electromagnetic waves has proven to be a difficult task (*13*). Advancements to enable angular selectivity in the transmission of light over optical wavelengths have been accomplished by exploiting a Brewster angle–enabled effect (*26*, *27*). However, broadband angular control of thermal emission from a bulk or nanostructured material remains challenging. Theoretical proposals have investigated the use of hyperbolic metamaterials and plasmonic metasurfaces that support a broad-spectrum Brewster angle to enable this capability (*19*, *28*).

An alternate approach for absorption and emission angular control lies in using thin, subwavelength films of materials exhibiting epsilon-near-zero (ENZ) behavior (*29*, *30*). These materials are known to support a leaky p-polarized electromagnetic mode near the material resonance pole, called a Berreman mode (*31*–*35*). This mode can couple to propagating free-space modes at a range of wave vectors and thus certain angles of incidence. Prior work has characterized the angular characteristics of this mode and its coupling to free space, both for absorption over near-infrared (NIR) wavelengths (*32*, *36*–*39*) and for thermal emission and long-wave infrared (LWIR) absorption from a range of oxides (*31*, *33*, *40*). This angular response, while intriguing, is fundamentally narrowband in nature, with no known mechanism for how it can be extended into a broadband response over the same sets of angles and thereby function as a broadband directional emitter or absorber.

To enable broadband directional thermal emission, we introduce the concept of gradient ENZ materials whose permittivity crosses zero at a range of frequencies that vary across a spatial gradient. We show that a gradient ENZ film along the depth dimension supports broad-spectrum leaky modes that couple to free-space propagating modes at nearly the same angles of incidence over that bandwidth, thus functioning as a broadband directional thermal emitter whose angular range of "thermal beaming" can be controlled by total film thickness. We fabricated two deep-subwavelength photonic structures as experimental realizations of the conceptual gradient ENZ film. Each structure consists of three polaritonic oxide materials that collectively support gradient ENZ behavior at different and complementary parts of the LWIR spectral range. We experimentally characterized the spectral and angular responses of both photonic structures; the first structure demonstrated broadband thermal beaming in the p-polarization from 7.7 to 11.5 μm over an angular range of 70° to 85°, and the second, from 10.0 to 14.3 μm over an angular range of 60° to 75°. We performed direct emission measurements at fixed temperatures to verify the directional nature of thermal emission at different angles for each structure.

Consider a thermal emitter with low emissivity over all wavelengths, $\varepsilon(\lambda,\theta) \sim 0$, at all polar angles of incidence θ except between a fixed, narrow set of angles $\theta_{emit} \in [\theta_1, \theta_2]$, where it exhibits higher emissivity, $\varepsilon_{high}(\lambda,\theta_{emit})$. Such a broadband directional thermal emitter or broadband thermal antenna (Fig. 1, A and B)

Department of Materials Science and Engineering, University of California, Los Angeles, CA, USA.
*Corresponding author. Email: aaswath@ucla.edu

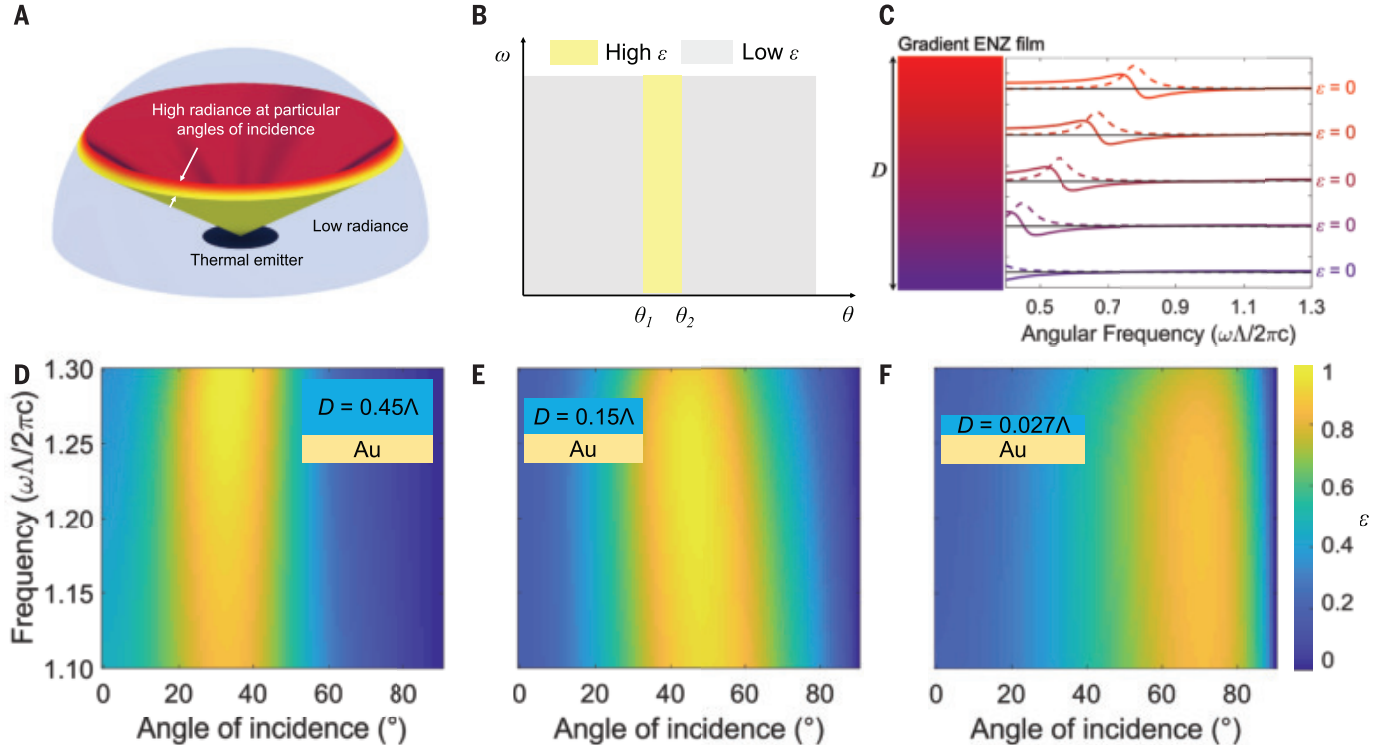


Fig. 1. Broadband directional thermal emitters enabled by gradient ENZ materials. (**A** and **B**) Schematic capability of a broadband directional thermal emitter. It has low emissivity, except at certain angles of incidence and emission, where it shows high emissivity across the entire thermal wavelength range of interest. (**C**) Schematic representation of the spatial gradient in the permittivity of the gradient ENZ film, whose dielectric permittivity is effectively near zero across a wide dimensionless frequency range from 1.1 to 1.3. (**D** to **F**) Emissivity of the gradient ENZ film varying with the total thickness in p-polarization as a function of angle of incidence and angular frequency. As the total thickness of the gradient ENZ materials increases from 0.027Λ to 0.15Λ and 0.45Λ, the center of the high emissivity band moves from 70° to 50° and 30°, respectively (where Λ, a scaling factor, is the mean of the wavelength range).

would then exhibit high radiance over a particular range of angles defined by θ_{emit} but at all other angles would, like a mirror, appear reflective, with low emissivity. This situation is in contrast to conventional bulk materials as well as finite-size nanostructures, where light typically couples to propagating free-space modes at particular wave vectors **k**, and thus angles of incidence θ, that vary as a function of frequency ω (including materials that support the Berreman mode). Although metals have a well-known polarization-dependent reduction in grazing angle reflection, which corresponds to an absorption and emission peak, this effect is only applicable at very large grazing angles (~85°). We instead seek to enable optical structures whose electromagnetic modes couple to free-space modes at angles defined by θ_{emit} over a large bandwidth. Central to this broadband thermal beaming capability is the emitter's ability to present high emissivity over a broad range of wavelengths while maintaining the same angular selectivity and directionality throughout that wavelength range.

Theoretically, a material whose dielectric permittivity is near zero over a broad range of wavelengths can act as a broadband directional thermal emitter in the p-polarization (fig. S1). While this is intriguing, ENZ behavior in real materials is tied to material dispersion and a resonance (through either electron or phonon interactions) and is therefore inherently narrowband, as has been observed with free-space coupling to the Berreman mode. To overcome this limitation, we introduce the concept of gradient ENZ materials whose permittivity is defined by a material resonance pole that changes its frequency across a spatial gradient. More specifically, we consider one-dimensional gradient ENZ films whose permittivity we express by modifying the standard model for the dielectric function of a polaritonic material with loss

$$\varepsilon(\omega, d) = \varepsilon_\infty\left(1+\frac{{\omega_L}^2-{\omega_T}^2}{{\omega_T}^2-\left(\omega-\left(\frac{d}{D}\right)\omega_{range}\right)^2-i\gamma\left(\omega-\left(\frac{d}{D}\right)\omega_{range}\right)}\right), \quad 0 \le d \le D \qquad (1)$$

Here, $\omega_{range} = \omega_L - \omega_T$, where ω_T and ω_L are phonon frequencies that correspond to out-of-phase atomic lattice vibrations with wave vectors aligned parallel (L, longitudinal) and perpendicular (T, transverse) to the incident field; ε_∞ is the permittivity at infinite frequency; γ is the damping rate; and d is the depth in the gradient ENZ film and is bounded by 0 and the film's total thickness D. At each point along the depth dimension, we thus have the permittivity crossing zero at different frequencies, establishing a range of ENZ frequencies that vary spatially (shown schematically in Fig. 1C). As the depth increases, this crossing point shifts to smaller frequencies. The gradient ENZ film shown in Fig. 1C exhibits ENZ values across a wide dimensionless frequency range from 1.1 to 1.3, a range that can be tuned by altering the overall spatial gradient of the material pole response.

To show that gradient ENZ materials can enable broadband directionality in absorption and emission, we characterized the spectral and angular response of three different total thicknesses of gradient ENZ film in the p-polarization using transfer-matrix simulations (Fig. 1, D to F). The thinnest film (Fig. 1F) displays the desired behavior shown schematically (Fig. 1B), as it has high emissivity in the p-polarization over a narrow range of angles, centered at 70°, with low emissivity at other angles, in the dimensionless frequency range from 1.1 to 1.3, where the entire gradient ENZ film supports near-zero permittivity across its depth. As the total thickness of the gradient ENZ materials increases, the angular range of high emissivity moves toward normal incidence

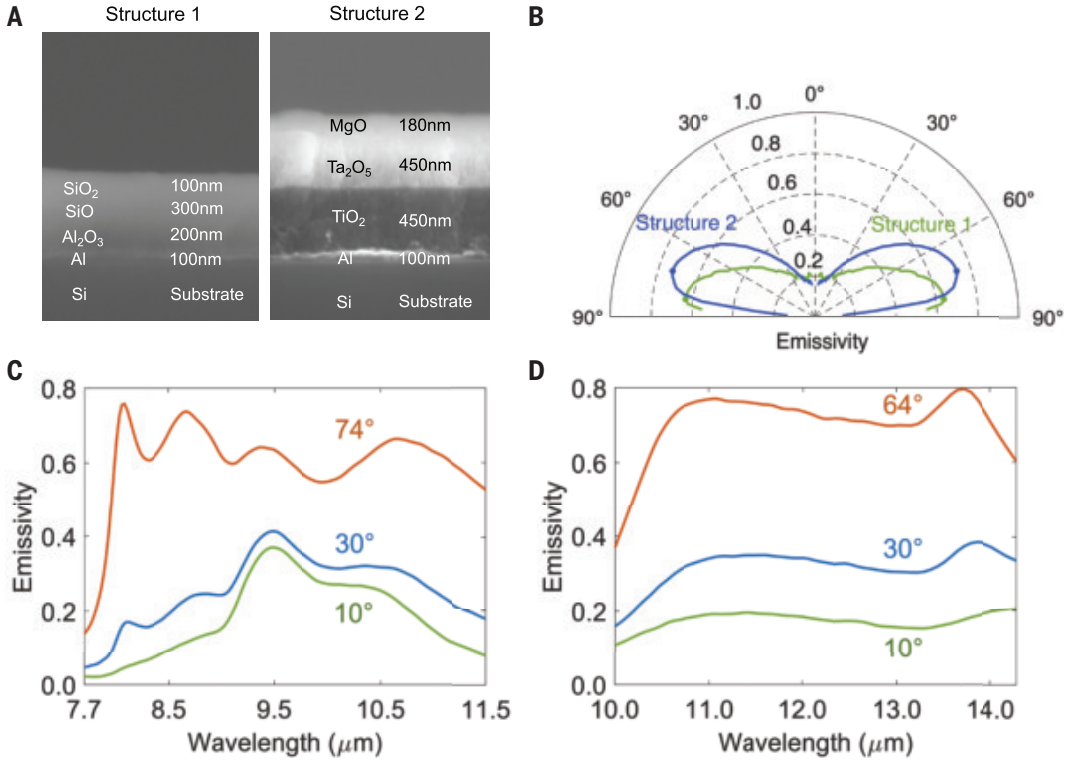


Fig. 2. Experimental characterization of two fabricated broadband directional thermal emitters. (**A**) SEM images of the experimentally fabricated multilayer photonic film structures, with labels identifying the materials used and layer thicknesses. (**B**) Polar plot of the measured average emissivity over broad wavelength ranges of operation (7.7 to 11.5 μm and 10.0 to 14.3 μm, respectively) varying with angle of incidence for p-polarization in the two structures, with high emissivity observed only at angles of incidence of 70° to 85° and 60° to 75° for structures 1 and 2, respectively. (**C** and **D**) Measured emissivity of the two photonic structures varying with wavelength at three different angles in the p-polarization. Both photonic structures exhibit a strong contrast between emissive and reflective behavior as a function of the angle of incidence.

(up to ~25° in our example here) but consistently over the entire bandwidth (Fig. 1, D and E), indicating the ability of gradient ENZ films to "beam" their thermal radiation at a range of angles in the p-polarization. To better understand the source of this behavior, we calculated the dispersion relations of the gradient ENZ films with different thicknesses (fig. S5) in the p-polarization. We observed that gradient ENZ films support distinctive, broadband Berreman modes to the left of the light line, which couple to free-space modes of nearly the same angle of incidence over the entire frequency range (1.1 to 1.3). In particular, as a function of thickness, the gradient ENZ film's supported mode is nearly parallel to the corresponding propagating free-space mode for a fixed angle of incidence over the entire frequency range, which is a different and unexpected behavior in conventional materials as well as photonic structures. Moreover, we again observed that as the film's thickness increases, the mode moves to smaller **k** vectors, nearly parallel to the propagating free-space modes at near-normal angles of incidence. Finally, the ideal ENZ material for broadband directional control should also have a slowly varying permittivity function around the ENZ frequency $\left[\frac{\partial \varepsilon}{\partial \omega}(\omega_{ENZ}) \to 0\right]$ and a small imaginary part of the permittivity at the ENZ frequency $\mathrm{Im}[\varepsilon(\omega_{ENZ})] \to 0$ to enable maximal directional contrast and tunability to near-normal angles of incidence (*41*).

We experimentally realized two gradient ENZ structures targeting the LWIR wavelength range that corresponds to the blackbody spectral radiance of room temperature–range objects (300 K). For both structures, we exploited the phonon polariton response of a range of semiconductor and metal oxides throughout this spectral range that in turn yield ENZ frequencies. The first structure was composed of silicon dioxide (SiO_2), silicon monoxide (SiO), and aluminum oxide (Al_2O_3), which presented ENZ wavelengths of 8.0, 8.7, and 10.7 μm, respectively. We plotted the permittivity of the three oxides on the basis of tabulated values (*42*, *43*) (fig. S3). Layering these oxides results in a gradient ENZ thin film because of their complementary resonances in the LWIR part of the spectrum. For the second structure, we used tantalum pentoxide (Ta_2O_5) (*44*), titanium oxide (TiO_2) (*42*), and magnesium oxide (MgO) (*45*), which presented ENZ wavelengths at 10.7, 12.0, and 13.5 μm, respectively (fig. S3). The thicknesses of both structures were optimized before fabrication through numerical simulations. We collected scanning electron microscope (SEM) images of the fabricated gradient ENZ structures and the layer thicknesses (Fig. 2A), with both structures deposited atop a base layer of aluminum on a 100-mm (4-inch) silicon wafer (*41*).

We measured an average emissivity in the p-polarization for the first structure of >0.6 over its wavelength range of operation (7.7 to 11.5 μm) within the angular range from 70° to 85° (Fig. 2B, green curve). Outside of this angular range, average emissivity is less than 0.3. This high emissivity wavelength range accounts for ~35% of the radiative heat flux of an object at 300 K and also overlaps with the atmospheric window responsible for radiative cooling. We measured average emissivity of the second gradient ENZ structure over its wavelength range of operation (10.0 to 14.3 μm) for p-polarization above 0.7 between 60° and 75° (Fig. 2B, blue line). Outside of this angular range, the emissivity drops drastically to 0.4 at 40°.

To better characterize the beaming effect, we plotted the spectral emissivity in the p-polarization for angles of incidence of 10°, 30°, and 74° (Fig. 2C). We observed a clear and consistent emissivity difference between different angles of incidence over the spectral range from 8.0 to 11.5 μm, with as much as a 76.6% contrast obtained near 8 μm. The spectral nature of the angular selectivity of structure 2 is shown in Fig. 2D. We observed consistent, broadband (from 10.5 to 14.3 μm) contrast between the measured emissivity at angles of incidence of 10°, 30°, and 64° for p-polarization (for all angles, p- and s-polarization, see fig. S4). The fabricated structures present very low emissivity at all angles in the s-polarization (fig. S4) owing to their deep subwavelength nature. Thus, while the polarization-averaged emissivity is lower than the measured p-polarization emissivity, the angular contrast and thermal beaming effect are maintained, as it is the photonic structure's p-polarization response that determines its angular response. Furthermore, we observed that, as expected, structures 1 and 2 have differing peak angles of

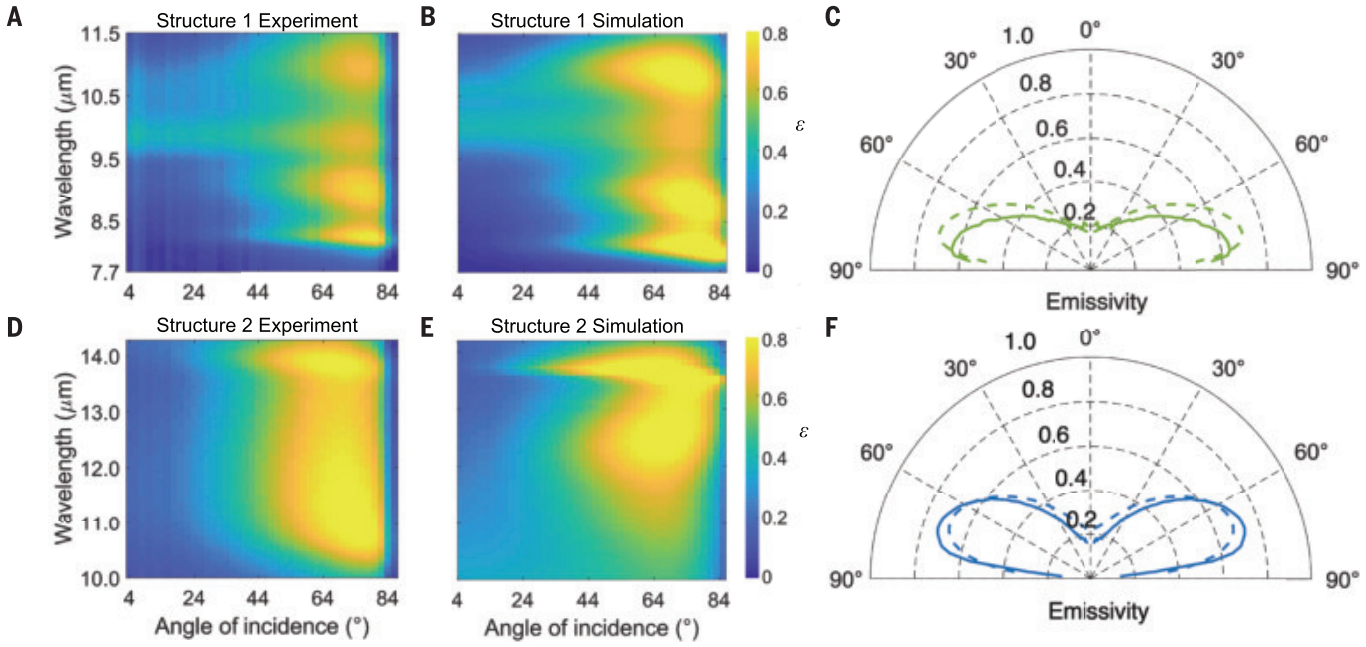


Fig. 3. Comparing experimental measurements with simulated results. (**A** and **D**) Measured emissivity spectra varying with angle and wavelength of the two fabricated photonic structures in p-polarization, and (**B** and **E**) simulation results of the emissivity of the two structures for p-polarization using the transfer matrix method. The simulation results show a broadband angularly selective behavior in emissivity and match well with the experimental measurements. (**C** and **F**) Average emissivity of the two structures from both experiments and simulations varying with the angle of incidence for p-polarization over wavelength ranges of 7.7 to 11.5 μm and 10.0 to 14.3 μm, respectively.

emission, at 82° and 72°, respectively, with the overall angular range of high emissivity also shifting from between 70° and 85° to between 60° and 75°. These results, in addition to demonstrating the core capability of broadband directional thermal emission, also highlight that through deliberate material choice and control of physical dimensions, gradient ENZ films can deliver directional control at a range of angles of incidence with materials easily accessible today.

To validate our experimental results, we show the measured as well as simulated emissivity spectra of both structures for the p-polarization varying with wavelength and angle of incidence (Fig. 3). For structure 1, the spectrum exhibits strong peaks at 8.0, 8.7, and 10.7 μm, respectively, which correspond to the ENZ wavelengths of SiO_2, SiO, and Al_2O_3, respectively (Fig. 3A). The overall p-polarized emissivity throughout the 7.7 to 11.5 μm range, within the angular range from 70° to 85°, is above 0.6 (Fig. 3C). Transfer matrix simulations of the fabricated structure (Fig. 3B) agree well with the experimental results and highlight clear angular regions of high and low emissivity that persist throughout the broad spectral range of operation. We plot the average emissivity of structure 1 over the wavelength range of operation (7.7 to 11.5 μm) from both measured and simulated data in polar coordinates (Fig. 3C). Although the measured emissivity is lower than what was expected from simulated data, the angles of operation and peak angle of incidence agree well. Similarly, we compared the measured and simulated emissivity spectra of structure 2 (Fig. 3, D and E) for the p-polarization varying with wavelength and angle of incidence. Both measured and simulated spectra exhibit continuous, strong directional emission across the wavelength range from 10.5 to 11.3 μm owing to the slowly varying permittivity of Ta_2O_5, TiO_2, and MgO (fig. S3). We again note a 10° shift in peak angle of incidence between structures 1 and 2, which we verified in the numerical modeling. We show good agreement in the average emissivity of structure 2 from both measured and simulated data in polar coordinates over the wavelength range of operation (10.0 to 14.3 μm) (Fig. 3F).

To understand the physical origin of the broadband thermal beaming effect observed in the fabricated gradient ENZ structures, we numerically calculated the dispersion relation for the two photonic structures (*41*). In addition to the two fabricated nanostructures, we also calculated the dispersion relation of alternative structures with differing thicknesses (fig. S7). For both structures, we observed frequency ranges above the light line that correspond to the ENZ wavelength ranges of SiO_2, SiO, and Al_2O_3 in structure 1 and Ta_2O_5, TiO_2, and MgO in structure 2. Other frequency ranges lying to the right of the light line cannot be coupled to from free space. As the total thickness increases, the dispersion curve moves to the left and thus will couple to modes from angles of incidence that are closer to normal incidence, agreeing well with the theoretical analysis of the conceptual gradient ENZ film (fig. S6) and indicating that angular response can be straightforwardly tuned to smaller angles of incidence by increasing the thickness of each layer.

One notable consequence of the broadband nature of the directional thermal emitters we demonstrated is their ability to emit a substantial amount of heat radiatively at tailored angles of incidence. We experimentally probed this behavior through direct radiometric measurements of their directional thermal emission. Both structures, deposited on 100-mm (4-inch) silicon wafers, were heated to 100° and 110°C, for structures 1 and 2, respectively, and emitted thermal radiation detected at fixed angles with a high-resolution microbolometer array (*41*). Thermographs of both structures are shown at a range of angles (Fig. 4, A and B). We observed that at angles of incidence equaling 77.5° and 72.5° for structures 1 and 2, respectively, the wafer appears drastically warmer than at smaller angles of incidence (10° and 45°), even though the actual temperature of the wafer is unchanged.

We plotted the measured directional emissivity for the two structures over the 7.5 to

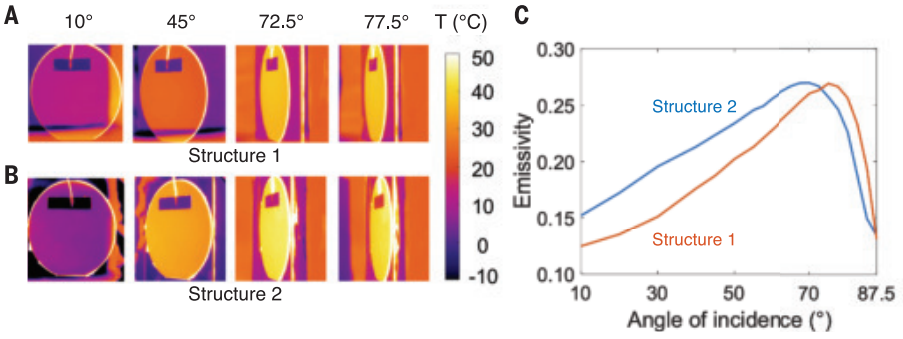


Fig. 4. Radiometric measurements. (**A** and **B**) Thermal images of the fabricated photonic thin film structures at a range of angles of incidence. At angles of 77.5° and 72.5° for the two structures, the wafer appears warmer than when the angle of incidence is at 10° and 45°, even though the actual temperature of the wafer is maintained. (**C**) Inferred average directional emittance by radiometrically derived temperature measurements over the 7.5 to 13.5 μm detector wavelength range for the two thermal emitter structures (*41*), at different angles of incidence.

13.5 μm wavelength range, as obtained through radiometry (*41*) (Fig. 4C). Even though the measured emission by the detector is polarization-averaged, the directionality of the thermal emission is clear. In agreement with the measurements (Figs. 2 and 3), we observed that the emissivity is high between 70° and 82° and between 60° and 75° for structures 1 and 2, respectively, with a peak emissivity value of 0.28 for both structures. The peak directional emissivity is smaller relative to the p-polarization directional emissivity data (Figs. 2 and 3), because the measurement is polarization-averaged and it determines average emissivity over the entire 7.5 to 13.5 μm range of the detector. The detector's range is 1.5 times larger than the designed wavelength range of operation of each structure. Despite this, a nearly 2:1 emissivity contrast was observed between normal and peak angles of incidence for both structures at 100° and 110°C, respectively. Thus, even though the enabled directional control is in the p-polarization, it enables a marked and anomalous contrast in the polarization-averaged thermal emission as a function of angle because the emissivity is minimal at all angles in the s-polarization over the relevant broadband spectrum (fig. S4). This capability highlights immediate applications for near-normal incidence heat signature minimization while maintaining substantial radiative heat transfer at other angles of incidence.

We have proposed and experimentally demonstrated that gradient ENZ materials enable broadband directional control of thermal emission and absorption. Our utilization of conventional oxides in fabricating our gradient ENZ films makes this approach immediately useful for large-area heat transfer applications for near–room temperature applications, because they are low-cost and amenable to well-established manufacturing processes. We emphasize that the temperature and wavelength ranges of operation can be tailored by material choice, for instance, using indium tin oxide (*32*, *36*) at NIR wavelengths, but this tailoring requires careful examination of the temperature dependence of the material's permittivity. While our approach enables angular selectivity in the p-polarization, higher total radiance can be achieved through simultaneous broadband angular selectivity in the s-polarization using effective epsilon- and mu-near-zero (MNZ) materials (*46*, *47*), which might be developed using a metamaterial approach. Alternatively, using gradient ENZ materials alone already delivers broad-spectrum, directional polarized light, a distinct capability that may itself be useful for a range of sensing and device applications. Our results suggest different opportunities to leverage the capabilities of gradient ENZ materials and the optical modes they support for broadband applications. In particular, we believe that these effects may have near-term impacts on heat transfer and energy applications, including directionality control for radiative cooling (*24*), thermophotovoltaics (*48*), and nascent approaches to waste heat recovery from radiative sources (*49*). More broadly, these results highlight that by decoupling angular and spectral response in photonic nanostructures, distinctive light emission and absorption capabilities can emerge.

REFERENCES AND NOTES

1. T. Inoue, M. De Zoysa, T. Asano, S. Noda, *Nat. Mater.* **13**, 928–931 (2014).
2. Z. J. Coppens, J. G. Valentine, *Adv. Mater.* **29**, 1701275 (2017).
3. C. W. J. Beenakker, *Phys. Rev. Lett.* **81**, 1829–1832 (1998).
4. C. Luo, A. Narayanaswamy, G. Chen, J. D. Joannopoulos, *Phys. Rev. Lett.* **93**, 213905 (2004).
5. M. Vollmer, K.-P. Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications* (Wiley, ed. 2, 2017).
6. J. A. Mason, S. Smith, D. Wasserman, *Appl. Phys. Lett.* **98**, 241105 (2011).
7. D. Jung, S. Bank, M. L. Lee, D. Wasserman, *J. Opt.* **19**, 123001 (2017).
8. G. C. Hulley *et al.*, *Atmos. Meas. Tech.* **9**, 2393–2408 (2016).
9. Y. Gong, S. S. Oh, D. L. Huffaker, N. Copner, "Novel mid-infrared metamaterial thermal emitters for optical gas sensing" in *Frontiers in Optics/Laser Science*, OSA Technical Digest (Optical Society of America, 2018), paper JTu3A-89.
10. Z. Pan *et al.*, *Sensors* **17**, 473 (2017).
11. L. Palchetti, G. Di Natale, G. Bianchini, *J. Geophys. Res. Atmos.* **121**, 10,804–10,819 (2016).
12. D. Costantini *et al.*, *Phys. Rev. Appl.* **4**, 014023 (2015).
13. Y. Shen, C. W. Hsu, Y. X. Yeng, J. D. Joannopoulos, M. Soljačić, *Appl. Phys. Rev.* **3**, 011103 (2016).
14. M. Laroche *et al.*, *Opt. Lett.* **30**, 2623–2625 (2005).
15. T. K. Wang, J. N. Zemel, *Infrared Phys.* **32**, 477–488 (1991).
16. P. J. Hesketh, J. N. Zemel, B. Gebhart, *Nature* **324**, 549–551 (1986).
17. P. J. Hesketh, J. N. Zemel, B. Gebhart, *Phys. Rev. B* **37**, 10803–10813 (1988).
18. J.-J. Greffet *et al.*, *Nature* **416**, 61–64 (2002).
19. G. Barbillon, E. Sakat, J.-P. Hugonin, S.-A. Biehs, P. Ben-Abdallah, *Opt. Express* **25**, 23356–23363 (2017).
20. E. Lucchi, *Renew. Sustain. Energy Rev.* **82**, 3077–3090 (2018).
21. E. Sakr, P. Bermel, *Opt. Express* **25**, A880–A895 (2017).
22. R. Sakakibara *et al.*, *J. Photonics Energy* **9**, 32713 (2019).
23. A. P. Raman, W. Li, S. Fan, *Joule* **3**, 2679–2686 (2019).
24. A. P. Raman, M. A. Anoma, L. Zhu, E. Rephaeli, S. Fan, *Nature* **515**, 540–544 (2014).
25. X. Yin, R. Yang, G. Tan, S. Fan, *Science* **370**, 786–791 (2020).
26. Y. Shen *et al.*, *Science* **343**, 1499–1501 (2014).
27. Y. Qu *et al.*, *ACS Photonics* **5**, 4125–4131 (2018).
28. C. Argyropoulos, K. Q. Le, N. Mattiucci, G. D'Aguanno, A. Alu, *Phys. Rev. B* **87**, 205112 (2013).
29. D. I. Vulis, O. Reshef, P. Camayd-Muñoz, E. Mazur, *Rep. Prog. Phys.* **82**, 012001 (2019).
30. I. Liberal, N. Engheta, *Science* **358**, 1540–1541 (2017).
31. J. Rensberg *et al.*, *Phys. Rev. Appl.* **8**, 014009 (2017).
32. J. Yoon *et al.*, *Sci. Rep.* **5**, 12788 (2015).
33. T. Shaykhutdinov, A. Furchner, J. Rappich, K. Hinrichs, *Opt. Mater. Express* **7**, 3706 (2017).
34. W. Streyer *et al.*, *Appl. Phys. Lett.* **104**, 131105 (2014).
35. W. D. Newman *et al.*, *ACS Photonics* **2**, 2–7 (2015).
36. M. A. Badsha, Y. C. Jun, C. K. Hwangbo, *Opt. Commun.* **332**, 206–213 (2014).
37. Z. Wang, P. Zhou, G. Zheng, *Results Phys.* **14**, 102376 (2019).
38. L. J. Krayer, J. Kim, J. L. Garrett, J. N. Munday, *ACS Photonics* **6**, 2238–2244 (2019).
39. Y. Guo *et al.*, *Appl. Phys. Express* **12**, 071005 (2019).
40. J. M. Ane, M. Huetz-Aubert, *Int. J. Thermophys.* **7**, 1191–1208 (1986).
41. Materials and methods are available as supplementary materials.
42. J. Kischkat *et al.*, *Appl. Opt.* **51**, 6789–6798 (2012).
43. G. Hass, C. D. Salzberg, *J. Opt. Soc. Am.* **44**, 181–187 (1954).
44. T. J. Bright *et al.*, *J. Appl. Phys.* **114**, 083515 (2013).
45. A. M. Hofmeister, E. Keppel, A. K. Speck, *Mon. Not. R. Astron. Soc.* **345**, 16–38 (2003).
46. I. Liberal, N. Engheta, *Nat. Photonics* **11**, 149–158 (2017).
47. I. Liberal, A. M. Mahmoud, Y. Li, B. Edwards, N. Engheta, *Science* **355**, 1058–1062 (2017).
48. N. Horiuchi, *Nat. Photonics* **14**, 66 (2020).
49. P. S. Davids *et al.*, *Science* **367**, 1341–1345 (2020).

ACKNOWLEDGMENTS

We thank P. McNeil and B. Dunn for experimental assistance. **Funding:** Supported by the Sloan Research Fellowship in Physics from the Alfred P. Sloan Foundation (A.P.R. and J.X.) and by the Schmidt Science Fellows, in partnership with the Rhodes Trust (J.M.). **Author contributions:** A.P.R. and J.X. conceived of the idea, developed the theoretical model, and designed and analyzed the experiments. J.X. designed the broadband thermal emitters and conducted the experiments. J.M. designed and performed the radiometric measurements with the assistance of J.X. J.X. and A.P.R. wrote the original draft, and all authors reviewed and edited the manuscript. A.P.R. supervised the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All associated data and materials are available in the manuscript and supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/393/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2

30 April 2020; resubmitted 4 February 2021
Accepted 22 March 2021
10.1126/science.abc5381

ORGANIC CHEMISTRY

C(sp^3)–H methylation enabled by peroxide photosensitization and Ni-mediated radical coupling

Aristidis Vasilopoulos[1], Shane W. Krska[2], Shannon S. Stahl[1]*

The "magic methyl" effect describes the change in potency, selectivity, and/or metabolic stability of a drug candidate associated with addition of a single methyl group. We report a synthetic method that enables direct methylation of C(sp^3)–H bonds in diverse drug-like molecules and pharmaceutical building blocks. Visible light–initiated triplet energy transfer promotes homolysis of the O–O bond in di-*tert*-butyl or dicumyl peroxide under mild conditions. The resulting alkoxyl radicals undergo divergent reactivity, either hydrogen-atom transfer from a substrate C–H bond or generation of a methyl radical via β-methyl scission. The relative rates of these steps may be tuned by varying the reaction conditions or peroxide substituents to optimize the yield of methylated product arising from nickel-mediated cross-coupling of substrate and methyl radicals.

The synthesis and structural modification of bioactive organic molecules is a central focus of pharmaceutical research and development (*1*, *2*). Even minor changes to the structure of the molecule can improve the activity or pharmacological properties of a drug candidate. This principle is clearly evident in the "magic methyl" effect (*3*), which describes the numerous cases in which replacing a hydrogen atom with a methyl group leads to higher potency, reduced toxicity, and/or increased stability of a molecule (Fig. 1A) due to changes in lipophilicity, three-dimensional shape, or conformational flexibility (*4*, *5*). The introduction of methyl groups into molecules is often nontrivial, requiring de novo chemical synthesis using a methylated starting material (*6*, *7*), and numerous efforts have been initiated to develop synthetic methods for methylation of bioactive molecules (*8*–*12*). Methods for direct methylation of C–H bonds would be particularly valuable because they would provide efficient access to structures that might not otherwise be available for biological testing (*3*, *8*, *13*, *14*). The present state of the art for C(sp^3)–H methylation features a two-step C–H oxygenation/methyl-nucleophile addition sequence (*15*). This method has some drawbacks [cryogenic temperatures in both steps; pyrophoric trimethylaluminum ($AlMe_3$) as the source of methyl nucleophile], but it provides a unique strategy to incorporate methyl groups adjacent to heteroatoms in pyrrolidines and other saturated heterocycles.

Here, we report a different strategy that leverages modular control over the generation and reactivity of organic radicals to achieve C(sp^3)–H methylation in a single step under mild conditions (Fig. 1B), taking advantage of tertiary alkyl peroxides as an oxidant capable of cleaving C–H bonds and as a source of methyl radicals (Fig. 1C). We postulated that successful reactivity could be achieved by combining the three fundamental reactivity concepts illustrated in Fig. 1D:

(i) Homolytic cleavage of the O–O bond of a peroxide, such as di-*tert*-butyl peroxide (DTBP), generates a pair of alkoxyl radicals, and this step could be promoted under mild conditions by means of visible-light triplet sensitization. This approach would avoid high-temperature thermolysis or direct, high-energy ultraviolet-light photolysis of the O–O bond.

(ii) The alkoxyl radicals generated in (i) could then undergo either hydrogen-atom transfer (HAT) from the substrate C–H bond to generate the substrate radical and tertiary alcohol by-product, or β-methyl scission to generate a methyl radical and ketone by-product.

(iii) Balancing the rates of the radical-forming reactions in (ii) should enable radical coupling to afford the desired methylated product; however, direct coupling of two transient radicals is challenging. We therefore envisioned that a transition metal catalyst could be used to generate comparatively long-lived organometallic intermediates capable of promoting selective carbon-carbon bond formation (*16*).

Peroxide reagents (e.g., DTBP and *tert*-butyl peroxybenzoate) have been widely used in C(sp^3)–H oxidation methods (*17*–*22*); however, precedents typically use a large excess of C–H substrate (e.g., ≥10 equiv). We speculated that this limitation reflects the competition between HAT and β-methyl scission (Fig. 1D, ii) at the elevated temperatures used to promote O–O homolysis (>90°C). Excess substrate is needed to enhance the rate of bimolecular HAT in order for this process to compete with unimolecular β-methyl scission. Several decades ago, fundamental studies of alkoxyl radicals showed that HAT becomes more favorable, relative to β-methyl scission, at lower temperatures (*23*, *24*). This insight suggested that operating the reaction at lower temperatures could support C–H activation without requiring excess C–H substrate. A strategy to generate alkoxyl radicals under mild conditions was inspired by studies in the 1980s, which showed that photosensitizers with suitable triplet energies induce O–O bond homolysis via energy transfer to DTBP (*25*–*27*).

Our initial experiments probed photochemical DTBP activation with a series of photosensitizers at room temperature in the presence of *N*-(3-phenylpropyl)phthalimide as an alkylarene C–H substrate (conditions: 1:2 ratio of alkylarene:DTBP, 1 to 10 mol % photosensitizer, irradiation using a "tuna blue" 400- to 470-nm LED lamp, acetonitrile solvent). The consumption of substrate was monitored by ^{1}H nuclear magnetic resonance (NMR) spectroscopy and showed a strong correlation with the excited-state triplet energy of the photosensitizer (Fig. 2A). Only 12% conversion of the alkylarene was observed in the absence of a photocatalyst, indicating that direct photolysis of the peroxide contributes little to the observed reactivity (*25*), whereas high conversion of alkylarene was observed using Ir photosensitizers with triplet energies of >55 kcal/mol. Thioxanthone, which has a triplet energy of 65 kcal/mol, was not effective with this light source, but it showed good HAT reactivity with a higher-energy violet LED lamp (395 to 400 nm). Ir[dF(CF_3)ppy]$_2$($^{t\text{Bu}}$bpy)PF_6 (Ir-F) readily absorbs blue and violet light and has been shown to be a robust photocatalyst in other synthetic methods (*28*, *29*). Consequently, Ir-F was chosen as photosensitizer for further development of the C(sp^3)–H methylation reaction.

Effective C–H methylation requires that β-methyl scission and HAT operate in parallel, although the optimal relative rates for these processes are not clear a priori. Therefore, we investigated the effect of different reaction parameters on the relative rates of β-methyl scission and HAT (Fig. 2B). For the benchmark reaction conditions, we used a 1:1 ratio of C–H substrate:DTBP and 1 mol % Ir-F in MeCN (0.3 M substrate) at room temperature under "tuna blue" light irradiation (see supplementary materials for experimental details). No reactivity of the peroxide or alkylarene was observed in the absence of light at room temperature. Variation of the reaction temperature revealed trends consistent with the studies noted above (*23*, *24*), with β-methyl scission favored at high temperatures and HAT favored at room temperature. The bimolecular HAT reaction benefited from higher concentrations, with substrate conversion progressing from 37% to 74% when the reaction concentration was increased from 0.15 M to 0.6 M. Solvents have been shown to affect both the absolute and relative rates of β-methyl scission and HAT (*30*, *31*) and were investigated as

[1]Department of Chemistry, University of Wisconsin, Madison, WI 53706, USA. [2]Merck & Co. Inc., Kenilworth, NJ 07033, USA.
*Corresponding author. Email: stahl@chem.wisc.edu

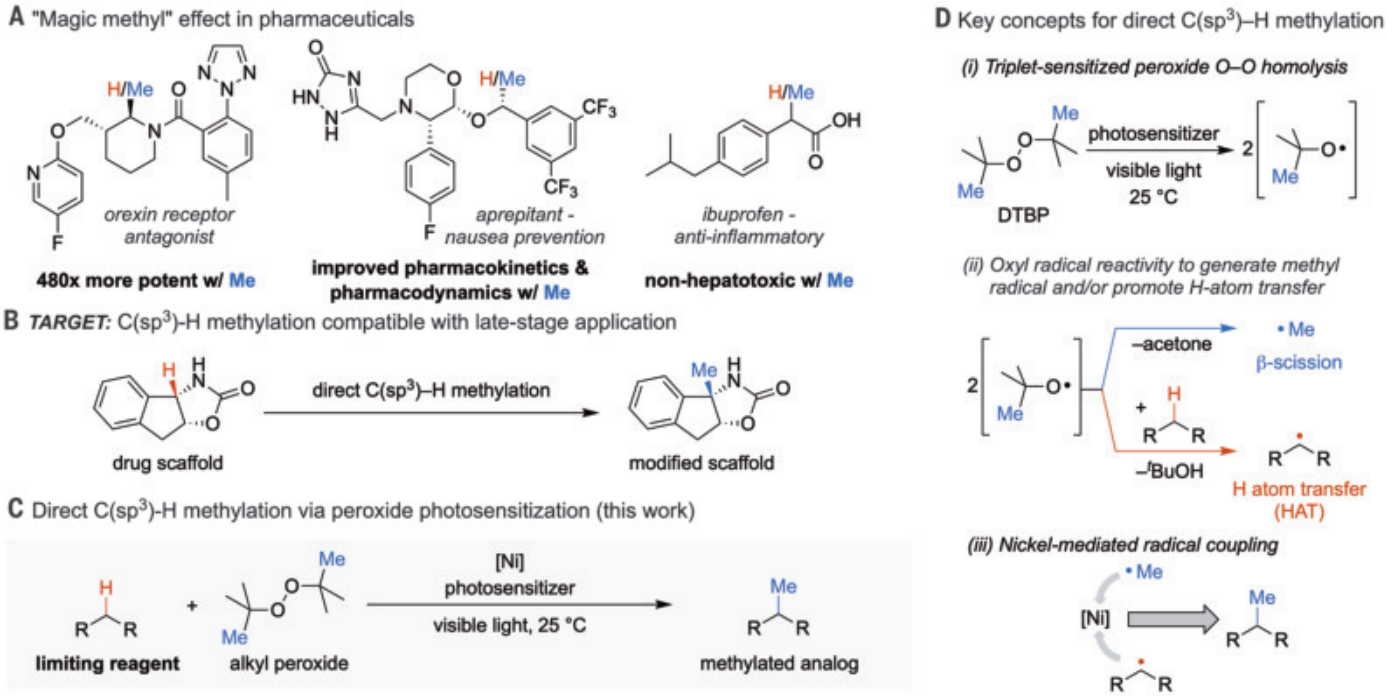


Fig. 1. Direct C(sp^3)–H methylation via photochemical activation of peroxides. (**A** and **B**) Benefits of methyl groups for drug properties (A) motivate the development of a synthetic method for direct C(sp^3)–H methylation (B). (**C**) General representation of our method. (**D**) Concepts that underlie Ni-catalyzed C(sp^3)–H methylation via photosensitization of di-*tert*-butyl peroxide.

another key variable. Dichloromethane attenuated the reactivity of both processes, whereas the polar aprotic solvent dimethyl sulfoxide inhibited β-methyl scission. In contrast, the polar protic solvent 2,2,2-trifluoroethanol (TFE) strongly enhanced β-methyl scission. Use of dicumyl peroxide (DCP) instead of DTBP also strongly favored β-methyl scission (*31*).

The reactions presented in Fig. 2, A and B, generated a broad mixture of mostly unidentified products, confirming that direct radical-radical coupling does not lead to the desired outcome. We postulated that a transition metal catalyst could template radical-radical coupling and lead to productive methylation of the C–H bond (Fig. 2C). Empirical evaluation of various metal salts, ligands, additives, and reaction conditions identified $NiCl_2$•dimethoxyethane ($NiCl_2$•dme) and 4,4′,4″-tri-*tert*-butyl-terpyridine (^{tBu}tpy) as the optimal radical catalyst, when used in combination with six equivalents of DTBP in TFE solvent and 0.5 equiv of trifluoroacetic acid (TFAH). Under these conditions, benzylic C–H methylation proceeded in 60% yield. The balance of C–H substrate primarily consisted of unreacted starting material (20%) and a benzylic dimethylation product (18%).

In addition to forming substrate methylation products, the reaction generated gaseous products, as was evident from a pressure build-up within the reaction vial. Therefore, the reaction was repeated in a reactor equipped with a pressure transducer to quantify the amount of gas generated, and after the reaction the headspace was analyzed by gas chromatography to establish gas identity (Fig. 2D). An initial experiment probed the reaction of DTBP in the absence of $NiCl_2$•dme and the benzylic C–H substrate. This reaction led to substantial formation of methane with a trace amount of ethane (0.36 mmol of total gas; Fig. 2D, orange time course). A second experiment was conducted in the presence of $NiCl_2$•dme but without the benzylic C–H substrate. A similar amount of gas was formed, but it mostly consisted of ethane (C_2H_6:CH_4 = 10.5:1; Fig. 2D, purple time course). Finally, the experiment was conducted in the presence of both $NiCl_2$•dme and the benzylic C–H substrate. This reaction produced less gas, which again consisted mostly of ethane (C_2H_6:CH_4 = 4.7:1; Fig. 2D, green time course) in addition to generating the C–H methylation product. The amount of methylation product (0.11 mmol) directly matched the reduction in gas quantity. Collectively, these experiments show that the Ni catalyst plays a key role in promoting C–C coupling. In the absence of Ni, methyl radicals generate methane, presumably by promoting HAT from the solvent or other species in the reaction mixture. In the presence of Ni, the methyl radicals can undergo homocoupling to form ethane (Fig. 2D, path B) or undergo cross-coupling with a benzyl radical (derived from HAT by the *tert*-butoxyl radical) to form the desired product (Fig. 2C). A structurally characterized [(^{tBu}tpy)Ni^{II}(Me)]$^+$ species has been reported previously and exhibits its reactivity consistent with this proposal (*32*).

The results outlined in Fig. 2 show how the reaction conditions may be tuned to balance the relative rates of HAT and methyl-radical formation, providing a foundation for optimizing C–H methylation of other substrates. For example, substrates that undergo slower HAT reaction with the alkoxyl radical should benefit from conditions that exhibit lower relative rates of β-methyl scission. Guided by these and related considerations, we used high-throughput experimentation methods to survey reaction conditions, catalyst compositions, and additives, using a diverse collection of substrates (Fig. 3). Four general reaction conditions were identified for C–H methylation of these molecules (Fig. 3A, conditions a to d; see supplementary materials for screening data). The methylation of most benzylic and α-amido substrates proceeded effectively with conditions a or b, which use ^{tBu}tpy as the ligand paired with either DTBP/TFE/TFAH or DCP/MeCN/$B(OH)_3$ as the peroxide, solvent, and acid source, respectively. Electron-rich C–H substrates, such as carbamates and ethers, exhibit better outcomes with tripicolylamine (TPA) as the ligand, with either DTBP/TFE or DCP/MeCN as the peroxide and solvent. Reaction concentration represents an additional variable to optimize the reactions (Fig. 2B), with a higher concentration (0.6 M) used for substrates that show

recalcitrance to methylation, and a lower concentration (0.15 M) used for substrates susceptible to dimethylation or overoxidation (e.g., dehydrogenation). The four general sets of reaction conditions, together with concentration modifications, provided the basis for methylation of the substrates in Fig. 3B.

Benzylic and related activated C–H substrates showed good reactivity (Fig. 3B, i). Alkylarenes with protected alcohol and amine functional groups generated the desired methylated products

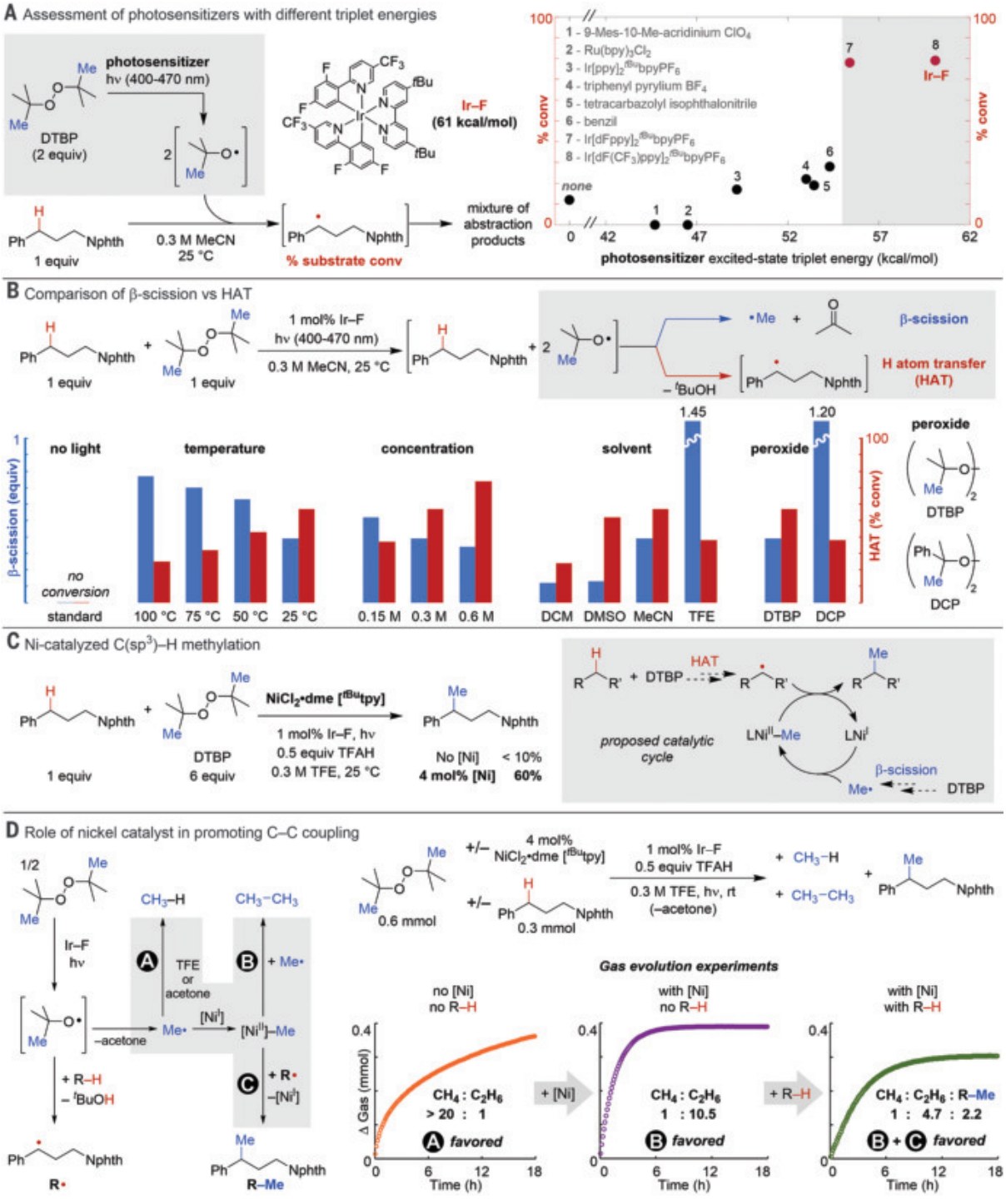


Fig. 2. Mechanistic insights and reaction guide for photochemical Ni-catalyzed C–H methylation. (**A**) Assessment of photosensitizers. (**B**) Survey of different reaction parameters, showing their influence on the relative rates of β-methyl scission (blue) and HAT from the substrate C–H bond (red). (**C**) Optimized conditions for Ni-catalyzed cross-coupling of substrate and methyl radicals. (**D**) Mechanistic insights into the fate of the methyl radical in the absence and presence of the Ni catalyst and the substrate. Reactions in (A) to (C) were conducted on 0.1-mmol scale and yields were determined by ^{1}H NMR (mesitylene standard). Nphth, phthalimide; ppy, 2-phenylpyridine; dFppy, 2-(2,4-difluorophenyl)pyridine; dF(CF$_3$)ppy, 2-(2,4-difluorophenyl)(5-trifluoromethyl) pyridine; ^{tBu}bpy, di-*tert*-butyl-bipyridine; ^{tBu}tpy, tri-*tert*-butyl-terpyridine; dme, dimethoxyethane.

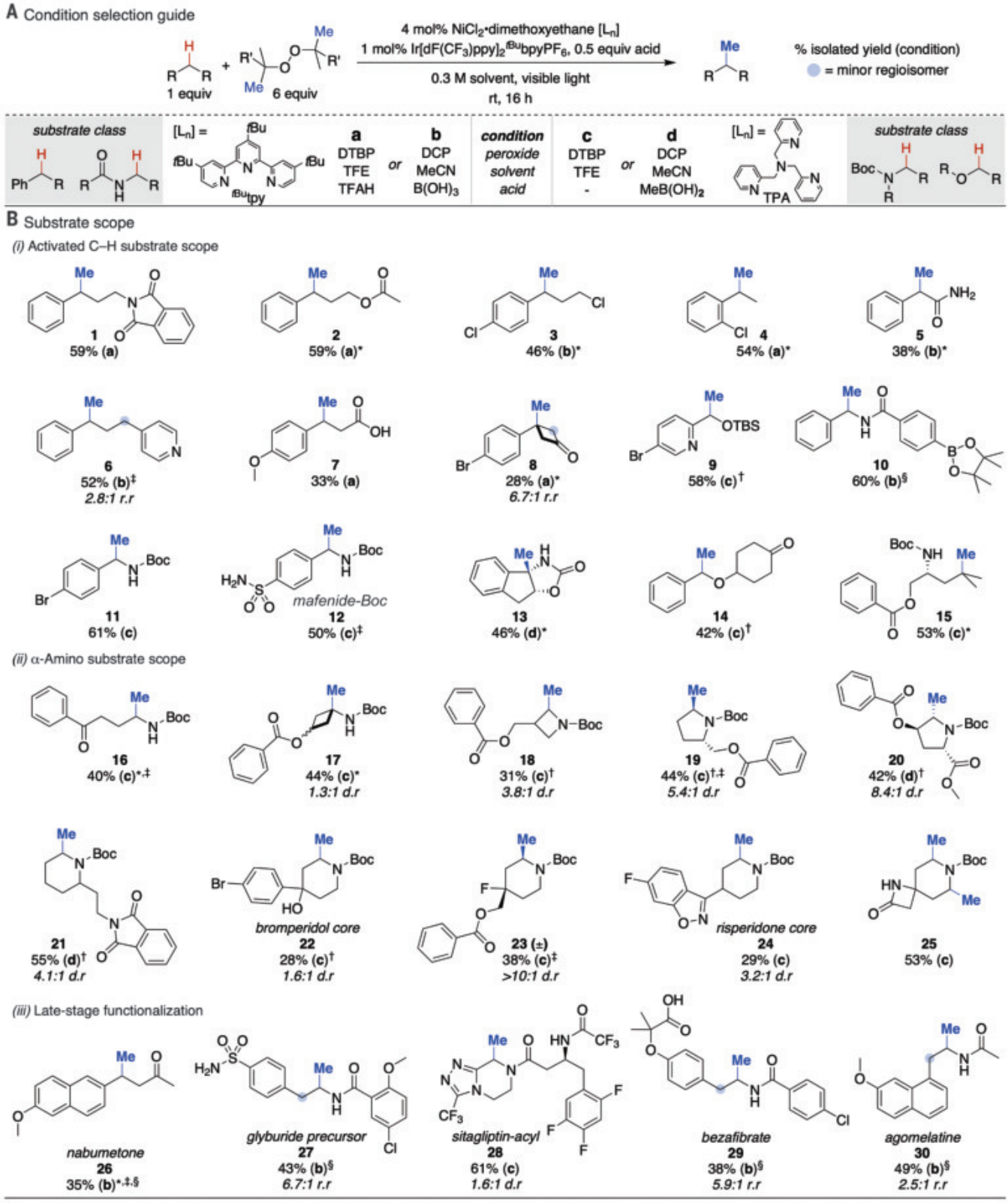


Fig. 3. Methylation substrate scope. (**A**) Different substrate classes and their optimal reaction conditions. (**B**) Reaction outcome with various substrates. Reactions were run on 0.3-mmol scale; isolated yields of combined regio- and stereoisomers are shown. Sites of secondary methylation are designated by blue circles. *0.6 M concentration. †0.15 M concentration. ‡0.5 equiv $MeB(OH)_2$ used as acid. §1:1 MeCN:DMSO used as solvent. r.r, regioisomeric ratio; d.r, diastereomeric ratio.

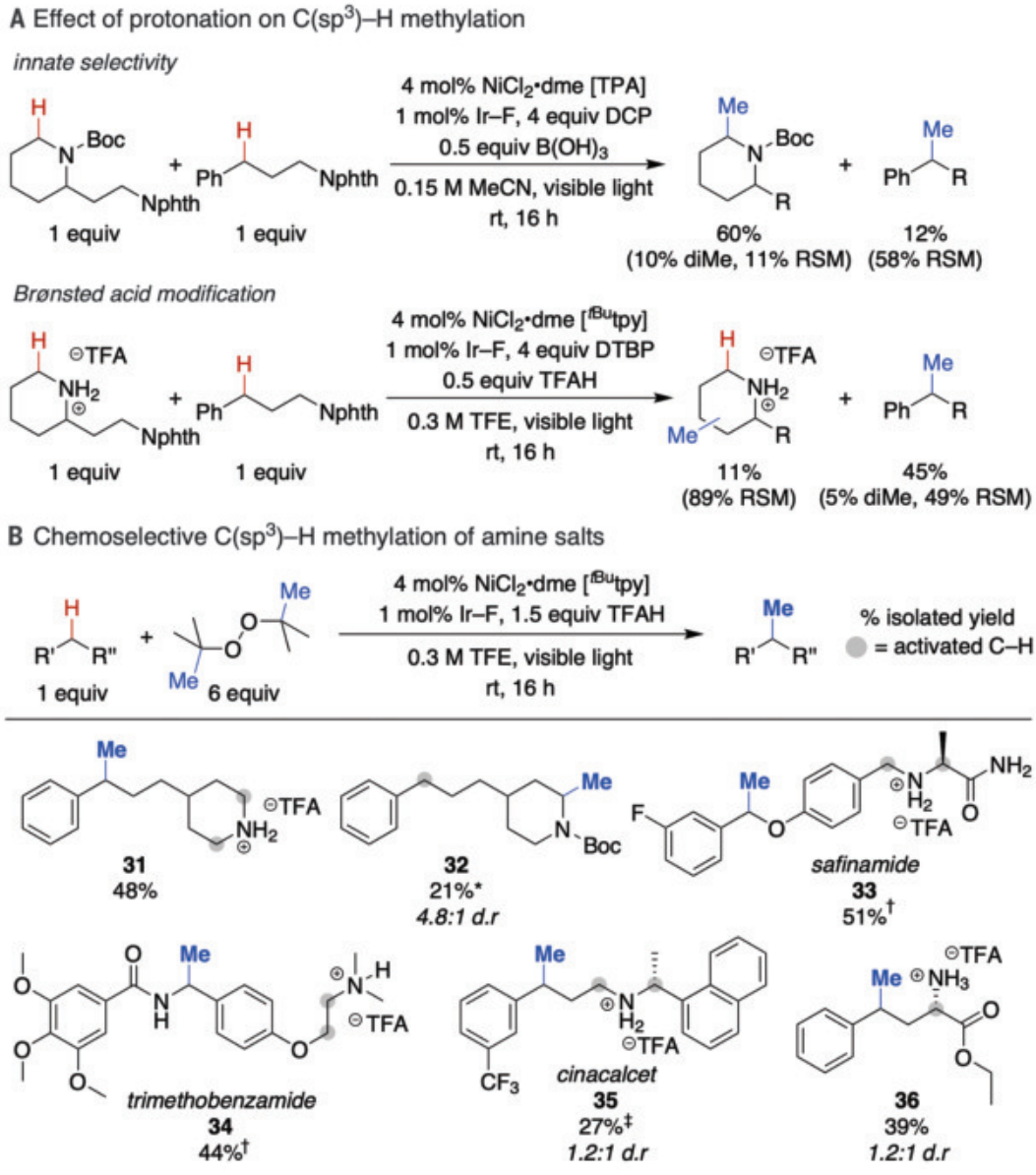


Fig. 4. Control over site selectivity with a Brønsted acid (**A**) α-Amino C–H sites are deactivated by protonation. (**B**) Chemoselective C–H methylation. Reactions were run on 0.3-mmol scale; isolated yields of combined stereoisomers are shown. *No acid added; used 3 equiv of DCP, TPA as the ligand, and 0.15 M MeCN as the solvent. †4 equiv of DCP and MeCN used in place of DTBP and TFE. ‡0.6 M concentration. RSM, recovered starting material.

in good isolated yields (**1** and **2**, 59% each). In these reactions and others, the remaining mass balance primarily consisted of unreacted starting material and dimethylation/multimethylation products. The reaction tolerated both aryl and alkyl chlorides (**3** and **4**). Formation of **5**, bearing a primary amide, benefited from using the DCP/MeCN conditions (b) rather than DTBP/TFE (a), possibly reflecting the influence of the reaction medium on C–H reactivity adjacent to carbonyl groups (*33*). Methylation of benzylic C–H bonds leading to **7** and **12** was favored over reaction with the acidic O–H/N–H bonds, in contrast to the outcome expected from other metal-catalyzed methylation reactions (*8*). The compatibility of aryl boronates and aryl bromides with the reaction conditions (**8**, **9**, **10**, **11**) is especially useful in elaborating lead compounds in medicinal chemistry. A silyl-protected alcohol (**9**) was unaffected by the methylation reaction. Minisci-type addition of the methyl radical to heteroarenes did not compete with C(sp^3)–H methylation of pyridine-containing substrates (**6** and **9**), despite the mildly acidic conditions (*14*, *34*). The methylation reaction was also capable of reacting with tertiary C–H bonds to generate quaternary centers (**8**, **13**, **15**).

C–H sites adjacent to the nitrogen atom of *tert*-butyloxycarbonyl (Boc)–protected amines were highly reactive toward C–H methylation (Fig. 3B, ii). For these substrates, condition c was typically most effective. This preference could reflect partial deactivation of the highly reactive C–H sites from a hydrogen-bonding interaction of the carbamate with TFE (*33*). Use of TPA as a ligand appeared to minimize formation of undesired side products (e.g., from dehydrogenation). Methylation of C–H sites in acyclic substrates and in four-, five-, and six-membered *N*-heterocyclic rings proceeded effectively (**16** to **25**). In these reactions, conversion of the starting material must be balanced against formation of dimethylation or multimethylation by-products. These reactions showed good compatibility with other functional groups. Methylation reactions of protected piperidines included substrates with a quaternary alcohol, an ester, a benzisoxazole, and a lactam (**22** to **25**). In substrate **25**, dimethylation was favored.

This methylation reaction also may be applied to late-stage functionalization of drug molecules (Fig. 3B, iii). High densities of potentially reactive functional groups are compatible, including a ketone, primary sulfonamide, trifluoroacetamide, and carboxylic acid, among others (**26** to **30**). Methylation of the piperazine C–H site in TFA-protected sitagliptin led to a 61% yield of product. Dimethyl sulfoxide was

used as co-solvent in some reactions to improve substrate solubility. In the glyburide precursor, bezafibrate, and agomelatine, the α-amino C–H sites underwent methylation in preference to the benzylic C–H sites (**27**, **29**, **30**).

Changes to the substrate and/or the reaction medium sometimes may be used to alter the site selectivity of reactions initiated by HAT. For example, protonation of basic amines by Brønsted acids deactivates adjacent C–H sites toward HAT with electrophilic radicals, such as *t*BuO• (*33*, *35*, *36*). This concept was investigated for the methylation reaction by conducting a competition experiment between methylation of a piperidine and an alkylarene substrate. In a non–hydrogen-bonding solvent, a Boc-protected piperidine underwent preferential methylation over the benzylic C–H substrate (5:1 selectivity; Fig. 4A). In contrast, the analogous competition with the protonated ammonium salt of piperidine exhibited the opposite selectivity, favoring the benzylic C–H site (1:4 selectivity; Fig. 4A). This reactivity was extended to the methylation of a substrate containing both a benzylic C–H site and a piperidine (**31** and **32**, Fig. 4B). Similar to the intermolecular competition experiment, protonation of the piperidine resulted in preferential reactivity at the benzylic position, whereas the Boc-protected piperidine favored methylation adjacent to nitrogen. Results obtained with a series of other substrates show that protonation of basic amines provides a general strategy to attenuate the reactivity of activated C–H bonds in complex molecules. Examples include methylated products of safinamide (**33**), trimethobenzamide (**34**), cinacalcet (**35**), and homo-phenylalanine (**36**). This Brønsted acid deactivation strategy should find broad utility in the methylation of pharmaceutical building blocks and drug molecules, which commonly feature basic amines.

Overall, this one-step method for C(sp^3)–H methylation of medicinally relevant building blocks and drug molecules greatly expands opportunities to exploit the "magic methyl" effect in drug discovery and development.

REFERENCES AND NOTES

1. P. N. Mortenson, D. A. Erlanson, I. J. P. de Esch, W. Jahnke, C. N. Johnson, *J. Med. Chem.* **62**, 3857–3872 (2019).
2. T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **45**, 546–576 (2016).
3. H. Schönherr, T. Cernak, *Angew. Chem. Int. Ed.* **52**, 12256–12267 (2013).
4. E. J. Barreiro, A. E. Kümmerle, C. A. M. Fraga, *Chem. Rev.* **111**, 5215–5246 (2011).
5. S. Sun, J. Fu, *Bioorg. Med. Chem. Lett.* **28**, 3283–3289 (2018).
6. J. Quancard *et al*., *Chem. Biol.* **19**, 1142–1151 (2012).
7. P. J. Coleman *et al*., *ChemMedChem* **7**, 415–424, 337 (2012).
8. G. Yan, A. J. Borah, L. Wang, M. Yang, *Adv. Synth. Catal.* **357**, 1333–1350 (2015).
9. Y. Chen, *Adv. Synth. Catal.* **362**, 998–1014 (2020).
10. Y. Chen, *Chem. Eur. J.* **25**, 3405–3439 (2019).
11. Z. Liang, W. Xue, K. Lin, H. Gong, *Org. Lett.* **16**, 5620–5623 (2014).
12. R. W. Pipal *et al*., *Nature* **589**, 542–547 (2021).
13. S. D. Friis, M. J. Johansson, L. Ackermann, *Nat. Chem.* **12**, 511–519 (2020).
14. D. Aynetdinova *et al*., *Chem. Soc. Rev.* 10.1039/D0CS00973C (2021).
15. K. Feng *et al*., *Nature* **580**, 621–627 (2020).
16. S. Wang, S. Tang, A. Lei, *Sci. Bull.* **63**, 1006–1009 (2018).
17. M. S. Kharasch, G. Sosnovsky, *J. Am. Chem. Soc.* **80**, 756 (1958).
18. G. Pelletier, D. A. Powell, *Org. Lett.* **8**, 6031–6034 (2006).
19. R. T. Gephart 3rd *et al*., *J. Am. Chem. Soc.* **134**, 17350–17353 (2012).
20. B. L. Tran, B. Li, M. Driess, J. F. Hartwig, *J. Am. Chem. Soc.* **136**, 2555–2563 (2014).
21. A. Vasilopoulos, S. L. Zultanski, S. S. Stahl, *J. Am. Chem. Soc.* **139**, 7705–7708 (2017).
22. H. Yi *et al*., *Chem. Rev.* **117**, 9016–9085 (2017).
23. P. Wagner, C. Walling, *J. Am. Chem. Soc.* **87**, 5179–5185 (1965).
24. R. D. Small, J. C. Scaiano, *J. Am. Chem. Soc.* **100**, 296–298 (1978).
25. J. C. Scaiano, G. G. Wubbels, *J. Am. Chem. Soc.* **103**, 640–645 (1981).
26. M. V. Encinas, E. A. Lissi, *J. Photochem.* **20**, 153–161 (1982).
27. P. S. Engel, T. L. Woods, M. A. Page, *J. Phys. Chem.* **87**, 10–13 (1983).
28. K. L. Skubi, T. R. Blum, T. P. Yoon, *Chem. Rev.* **116**, 10035–10074 (2016).
29. J. Twilton *et al*., *Nat. Rev. Chem.* **1**, 52 (2017).
30. C. Walling, P. J. Wagner, *J. Am. Chem. Soc.* **86**, 3368–3375 (1964).
31. D. V. Avila, C. E. Brown, K. U. Ingold, J. Lusztyk, *J. Am. Chem. Soc.* **115**, 466–470 (1993).
32. G. D. Jones, C. McFarland, T. J. Anderson, D. A. Vicic, *Chem. Commun.* **33**, 4211–4213 (2005).
33. M. Bietti, *Angew. Chem. Int. Ed.* **57**, 16618–16637 (2018).
34. D. A. Dirocco *et al*., *Angew. Chem. Int. Ed.* **53**, 4802–4806 (2014).
35. M. Lee, M. S. Sanford, *J. Am. Chem. Soc.* **137**, 12796–12799 (2015).
36. J. M. Howell, K. Feng, J. R. Clark, L. J. Trzepkowski, M. C. White, *J. Am. Chem. Soc.* **137**, 14590–14593 (2015).

ACKNOWLEDGMENTS

We thank B. Li, K. Dykstra, M. Uehling, S. Borges, and D. Schultz (Merck & Co. Inc., Kenilworth, NJ, USA) for valuable discussions, feedback, and training pertaining to the implementation of experimentation techniques used for development of this chemistry; C. Salazar and W. McDermott for assistance with gas evolution experiments; and I. Hermans (UW-Madison) for use of his gas chromatograph for head-space analysis. **Funding:** Supported by NIH grants R01 GM126832 and R35 GM134929. Spectroscopic instrumentation was partially supported by NIH grant 1S10 OD020022-1 and NSF grant CHE-1048642. **Author contributions:** A.V. conceived the project in collaboration with S.S.S., performed all of the experimental work, and led primary data interpretation and analysis; S.W.K. and S.S.S. consulted with A.V. throughout the project. **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/372/6540/398/suppl/DC1
Materials and Methods
Figs. S1 to S55
Tables S1 to S12
Characterization Data
References (*37*–*55*)

26 February 2021; accepted 25 March 2021
10.1126/science.abh2623

LASERS

Higher-dimensional supersymmetric microlaser arrays

Xingdu Qiao[1]†, Bikashkali Midya[2,3]†, Zihe Gao[2]†, Zhifeng Zhang[1], Haoqi Zhao[1], Tianwei Wu[2], Jieun Yim[2], Ritesh Agarwal[2], Natalia M. Litchinitser[4], Liang Feng[2,1]*

The nonlinear scaling of complexity with the increased number of components in integrated photonics is a major obstacle impeding large-scale, phase-locked laser arrays. Here, we develop a higher-dimensional supersymmetry formalism for precise mode control and nonlinear power scaling. Our supersymmetric microlaser arrays feature phase-locked coherence and synchronization of all of the evanescently coupled microring lasers—collectively oscillating in the fundamental transverse supermode—which enables high-radiance, small-divergence, and single-frequency laser emission with a two-orders-of-magnitude enhancement in energy density. We also demonstrate the feasibility of structuring high-radiance vortex laser beams, which enhance the laser performance by taking full advantage of spatial degrees of freedom of light. Our approach provides a route for designing large-scale integrated photonic systems in both classical and quantum regimes.

The rapid development of integrated photonics, with continuous effort to push the limit of integration density, offers a solution to the future scaling of integrated photonic networks and devices. However, the wave nature of light gives rise to fundamentally inevitable mutual coupling between photonic elements that are closely packed in an array. Control of mutual coupling is thus the key to phase locking all of the elements and further driving them to function collectively. For example, coherence and synchronization are critical for high-radiance optical emitters and lasers, as the radiance scales with the number of elements only if they are mutually coherent (*1*, *2*). Integrated high-radiance sources including synchronized laser arrays (*1*–*9*) and photonic crystal lasers (*10*, *11*) are in high demand and are actively pursued for a wide range of applications, including optical communication, optical sensing, and light detection and ranging (LIDAR). Among various strategies developed to enforce mutual coherence between laser resonators, the

[1]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]Department of Physical Sciences, Indian Institute of Science Education and Research, Berhampur 760010, India. [4]Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA.
†These authors contributed equally to this work.
***Corresponding author. Email: fenglia@seas.upenn.edu**

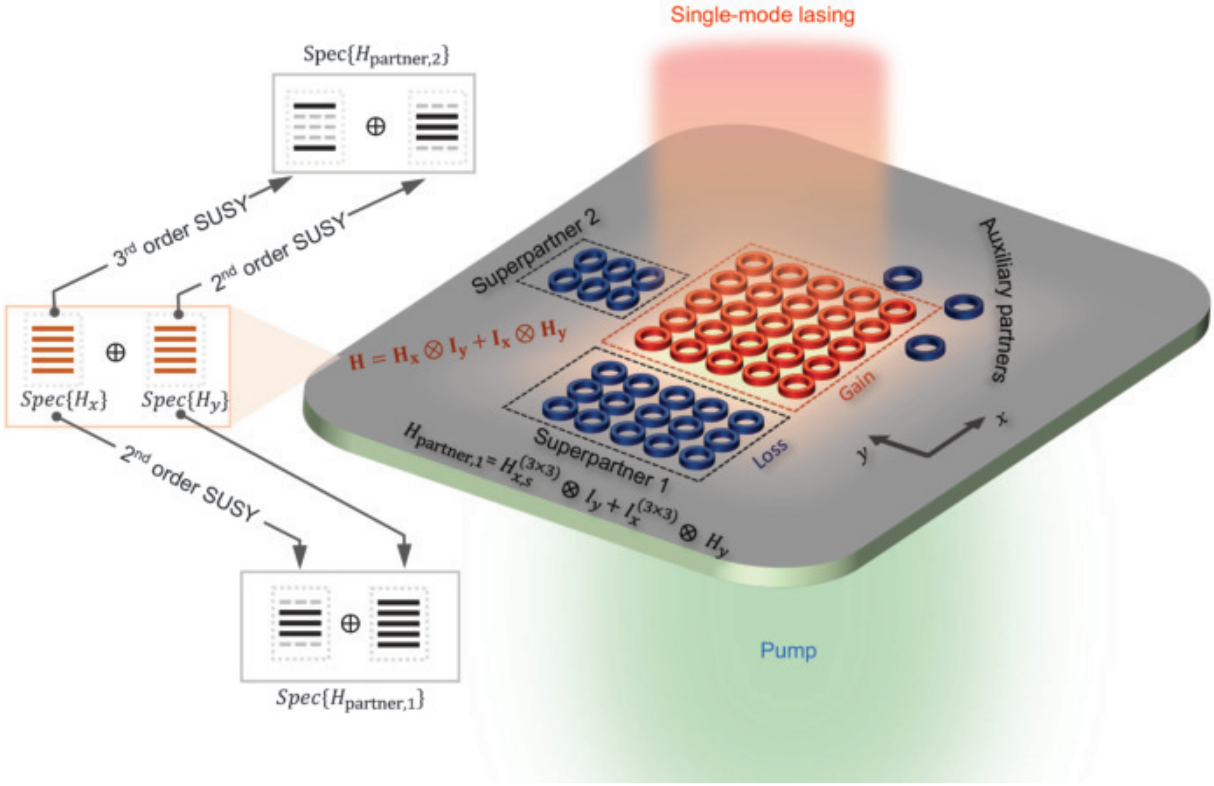


Fig. 1. A higher-dimensional supersymmetric microlaser array. The array consists of a 5 × 5 main array of evanescently coupled microring lasers (red), coupled with its two dissipative superpartner arrays and three auxiliary partner microrings (blue). The second-order SUSY transformation on H_x yields its SUSY partner $H_{x,s}^{(3\times3)}$, a 3 × 3 matrix denoting the coupling strengths in superpartner 1 in the x direction, which is isospectral to H_x except for the highest and lowest energy levels. A Kronecker sum between $H_{x,s}^{(3\times3)}$ and the unvaried H_y generates the Hamiltonian of superpartner 1, corresponding to an array of 3 × 5 coupled microrings. Similarly, a third-order SUSY transformation on H_x and a second-order SUSY transformation on H_y yields the Hamiltonian of superpartner 2, $H_{\text{partner},2} = H_{x,r}^{(2\times2)} \oplus H_{y,s}^{(3\times3)}$, corresponding to a 2 × 3 array. Three auxiliary partner microrings are introduced to couple to three of four remaining energy levels of H (*22*). Altogether, the dissipative supersymmetric and auxiliary partners are isospectral to the main array except for the lowest energy level associated with the fundamental transverse supermode, suppressing all the higher-order transverse supermodes but facilitating a high-radiance single-mode laser action operating exclusively on the fundamental in-phase supermode.

evanescent wave coupling–based strategy—which leverages strong optical confinement (such as micro- and nanoscale resonators and waveguides) and operates in the deep subwavelength regime—is approaching the limit of integration density. A key drawback of evanescent wave coupling is its intrinsically associated energy splitting, which leads to complex mode competition and thus energy inefficiency and irregular, chaotic radiation. The alternative approaches, including antiguided, diffractive, and antenna coupling (*3*–*6*), require delicately designed leakage of optical modes to communicate between elements, which ultimately limits their downsizing and dense packaging as well. Here, we demonstrate a formalism based on higher-dimensional supersymmetry (SUSY) to enforce phase locking and to enable coherent oscillation in a two-dimensional (2D) laser array of evanescently coupled microlasers, opening avenues for the realization of integrated high-radiance sources while maintaining individual controllability of each microlaser.

SUSY was first introduced in string and quantum field theory to unify all physical interactions in nature, including strong, electroweak, and gravitational coupling (*12*). Despite awaiting experimental validation in particle physics, the mathematic framework of SUSY has found its applications in many other branches of modern physics, ranging from nonrelativistic quantum mechanics and condensed-matter physics (*13*, *14*) to optics and photonics (*15*, *16*). SUSY is particularly powerful in tailoring mutual interactions in any arbitrary lattice of evanescently coupled elements, regardless of its complexity and size. With the unbroken SUSY, for example, the lattice Hamiltonian (where couplings are represented by the off-diagonal elements) can be transformed into a new superpartner Hamiltonian with a reduced matrix dimension, which shares almost the same eigenspectrum except for the disappearance of the original fundamental mode. This characteristic has enabled a series of photonic functionalities such as effective mode control, selection, and creation (*15*–*21*), which has facilitated phase-locked 1D laser arrays when strategically performed in a non-Hermitian photonic environment (*7*–*9*). Nevertheless, the factorization technique used in supersymmetric transformation applies only to the Hamiltonians of 1D systems (*13*), which limits the scalability of SUSY photonics and hinders its further promotion into higher dimensions. Here, we have overcome this limitation and demonstrate a generic approach to higher-dimensional supersymmetric microlaser arrays. The dissipative superpartners, coupled to a 2D main laser array, enforce the evanescently coupled microlasers to phase lock and coherently oscillate in the fundamental in-phase supermode, taking the full advantage of dense integration and fulfilling the demand for integrated, high-radiance optical sources with an orders-of-magnitude enhancement in energy density.

We use a homogeneously coupled 2D array of five by five (5 × 5) identical microring lasers made of InGaAsP multiple quantum wells (Fig. 1). Considering the evanescent wave coupling between only the nearest neighbors,

the main microlaser array can be represented by a tight-binding Hamiltonian

$$H = -\sum_{m,n}(\kappa_x a^{\dagger}_{m+1,n} a_{m,n} + \kappa_y a^{\dagger}_{m,n+1} a_{m,n} + h.c.) \quad (1)$$

where κ_x and κ_y denote the nearest-neighbor coupling coefficient between adjacent microrings along the x and y directions, respectively; (m, n) labels the microring sites in the (x, y) plane; $a(a^{\dagger})$ is the photon annihilation (creation) operator of the resonant modes in individual microrings; and *h.c.* denotes the Hermitian conjugate, leading to 25 transverse supermodes with closely spaced eigenfrequencies (fig. S1) (*22*). Despite many transverse supermodes, our objective is to facilitate high-radiance, single-frequency lasing in 2D space with collective in-phase oscillation in only the fundamental transverse mode, while suppressing all other supermodes.

We develop a generic approach to derive the isospectral superpartners of higher-dimensional tight-binding lattices, though the SUSY-based factorization technique is one-dimensionally constrained. The orthogonality inherently associated with the tight-binding Hamiltonian in Eq. 1 allows us to separate it in the form of a Kronecker sum

$$H = H_x \oplus H_y = H_x \otimes I_y + I_x \otimes H_y$$

where $\oplus$ and $\otimes$ denote the Kronecker sum and the tensor (Kronecker) product between matrices, respectively; H_x and H_y represent 1D systems, consisting of five coupled resonators along the x and y directions, respectively, with coupling strengths κ_x and κ_y; and I_y and I_x are 5×5 identity matrices (*22*). Here, the 2D isospectral superpartners can be configured using the tensor product based on two superpartners of H_x and H_y. Specifically, in contrast to nonnegligible onsite frequency detuning across the superpartner array resulting from the first-order SUSY transformation (*8*), we apply the second-order SUSY transformation that can yield a homogeneous superpartner array respecting the particle-hole symmetry and thus consisting of identical elements with the same resonant frequency compared to the main array (*9*), which is experimentally favorable, especially for a large-scale system. To create a symmetric spectrum for the superpartner, in the second-order SUSY transformation we choose to eliminate two levels with the highest and lowest frequencies in the 1D Hamiltonian (i.e., the matrix dimension of the SUSY partner of $H_{x,y}$ reduces from 5×5 to 3×3 in our case). The SUSY transformation thus leads to two superpartner arrays (*22*): a 3×5 array with 15 transverse modes corresponding to $H_{\text{partner},1} = H^{(3\times3)}_{x,s} \otimes I_y + I^{(3\times3)}_x \otimes H_y$, where $H_{x,s}$ is the second-order SUSY transformation of H_x, and a 2×3 array with six transverse modes corresponding to $H_{\text{partner},2} = H^{(2\times2)}_{x,r} \otimes I^{(3\times3)}_y + I^{(2\times2)}_x \otimes H^{(3\times3)}_{y,s}$, where $H_{y,s}$ is the second-order SUSY transformation of H_y, and $H_{x,r}$ (the residual) is an arbitrary 2×2 Hamiltonian that is isospectral to the energy levels originally in H_x but eliminated in achieving $H_{x,s}$. $H_{x,r}$ can also be generated by a third-order SUSY transformation of H_x by deleting three modes apart from the first and last mode (*22*). Superscripts denote the dimensions of the matrices, not to be confused with the size of the resonator array. Together with three individual auxiliary partner microrings (two of three have zero relative frequency

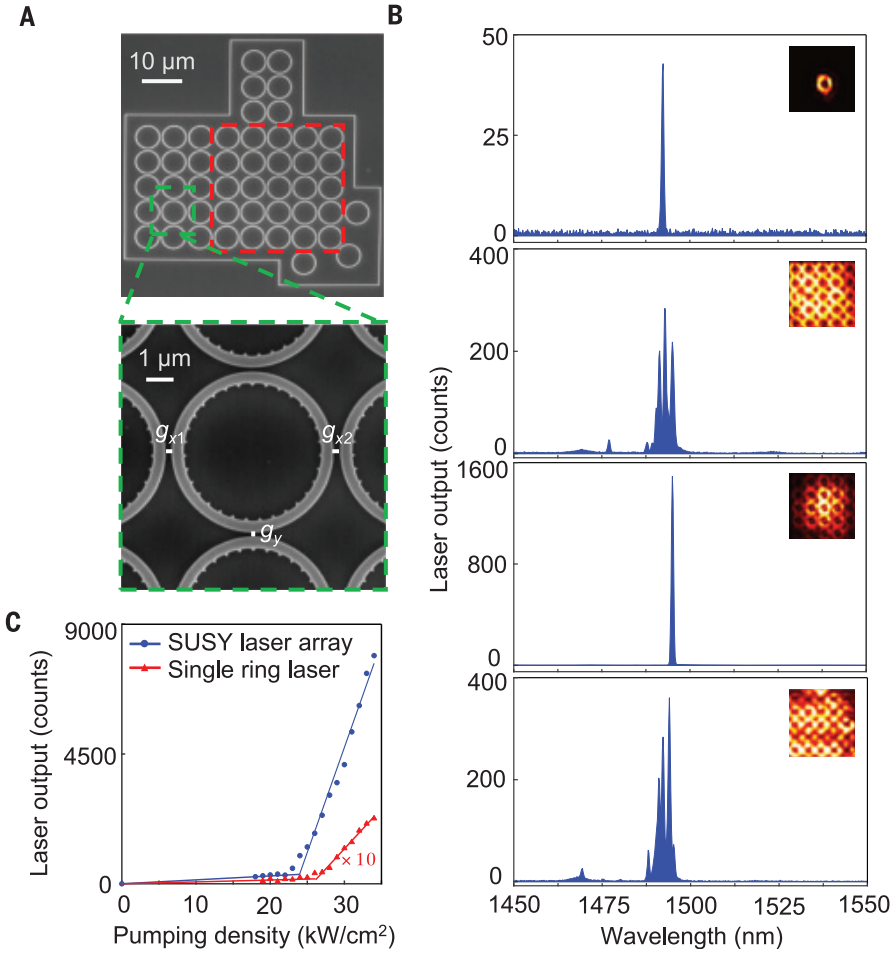


Fig. 2. Experimental characterization of the higher-dimensional supersymmetric microlaser array. (**A**) SEM images of the SUSY microlaser array. The main array is denoted by the red box, where uniform evanescent couplings are introduced with a gap of 200 nm between any pair of adjacent rings. The supersymmetric and auxiliary partners are placed in proximity to the main array, with a gap of 330 nm. The zoom-in image (bottom) shows detailed couplings in superpartner array 1: varying coupling strengths with a gap of g_{x1} = 240 nm and the other g_{x2} = 300 nm in the x direction and uniform coupling equal to that in the main array with a gap of g_y = 200 nm in the y direction. (**B**) Emission spectra, from top to bottom, of a single microring laser (single frequency lasing at 1492 nm), a 5×5 array of microring lasers with identical design parameters but no coupled supersymmetric and auxiliary partners, the supersymmetric microlaser array coupled with dissipative partners by selective pumping (single frequency lasing at 1495 nm), and the supersymmetric microlaser array but with uniform pumping on both the main array and partners, respectively, at the same pumping intensity of 32 kW/cm^2. The SUSY laser array lases in the longest wavelength mode among all the supermodes in the 5×5 array, which confirms that it is the fundamental supermode. In the two measurements for the SUSY array, only the emission from the main array is collected for a fair comparison. The insets show the corresponding emission images at the sample plane, where intensity distribution associated with the selectively pumped SUSY laser array also confirms its operation in the fundamental supermode. (**C**) Light-light curve showing the lowering of the threshold and the enhancement of lasing output (the slope efficiency) in the SUSY microlaser array compared with a single microring laser. The laser output of the single microring laser is magnified by 10× for better visualization.

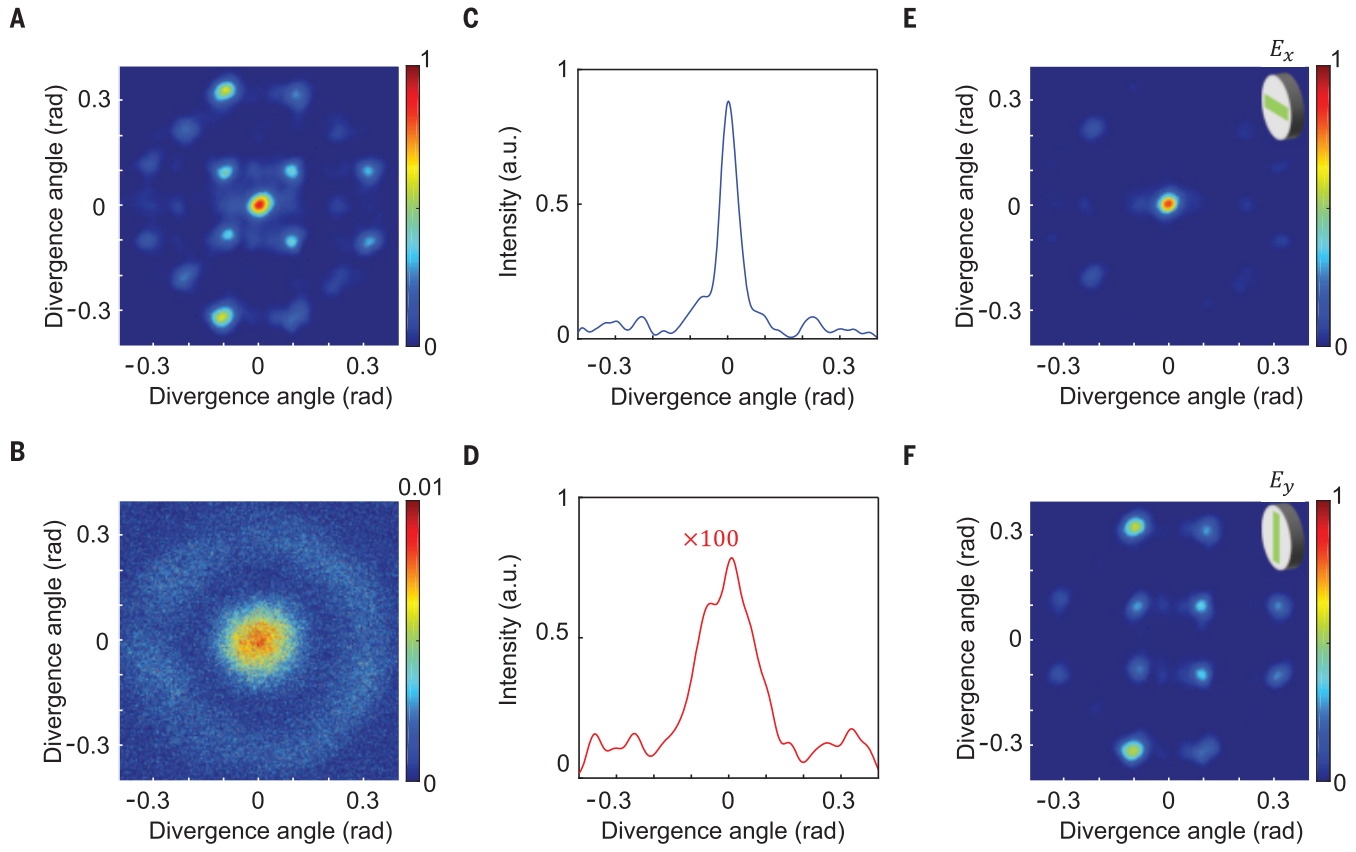


Fig. 3. Far-field characterization of laser emission from the higher-dimensional supersymmetric microlaser array. (**A** and **B**) Far-field diffraction patterns of laser emissions from the SUSY microlaser array (A) and a single microring laser (B). (**C** and **D**) The corresponding intensity distributions of laser emissions from the SUSY microlaser array (C) and single microring laser (D), both along the *x* axis, showing a small divergence of ~2° (versus divergence of ~11° for the single microlaser) and an energy concentration with a two-orders-of-magnitude enhancement in intensity associated with the laser beam from the SUSY microlaser array. Lasing intensity in (D) is magnified by 100× for better visualization. a.u., arbitrary units. (**E** and **F**) *x*- and *y*-polarized diffraction patterns of emission from the SUSY microlaser array, arising from the two degenerate modes in microrings while both oscillating in the fundamental transverse mode.

detuning, and the frequency of the last one matches the out-of-phase supermode with the highest relative frequency among all 25 transverse modes), the spectrum of superpartners is identical to that of the main array, apart from the fundamental in-phase supermode (fig. S3) (*22*). Strategically controlled coupling of the main array with its dissipative superpartners and auxiliary partner microrings, by matching both the eigenfrequencies and mode distributions (fig. S5), therefore ensures the suppression of all but the fundamental transverse supermode, yielding efficient single-supermode lasing.

The scanning electron microscope (SEM) images of the SUSY array sample fabricated on 200-nm-thick InGaAsP multiple quantum wells (*22*) are shown in Fig. 2A. The microring resonators were patterned with an inner radius of 3 μm and a width of the waveguide of 400 nm, operating at a resonance order of N = 32 for the quasi-TE (quasi–transverse-electric) mode. To facilitate surface emission from microring lasers, an angular grating was inscribed on the inner sidewall of each microring. Emission from each scatterer in the angular grating is circularly polarized, resulting from the transverse spin in the evanescent region of the waveguide (i.e., the azimuthal and radial electric field components have a π/2 phase difference) (*23*, *24*). With the transverse spin ($|s| = 1$), we designed the order of the angular grating as $M = N - 1 = 31$, creating phase matching for equiphase emission from all of the scatterers on a microring carrying an orbital angular momentum (OAM) of $l = 0$. The emission spectrum of the SUSY microlaser array was characterized in addition to the emission spectra from three control experiments (Fig. 2B), all pumped optically by a nanosecond pulsed laser at the wavelength of 1064 nm (*22*). Although a single microlaser operates with the designed single-mode laser action, evanescent couplings between microlasers in a 5 × 5 array with the Hamiltonian in Eq. 1 cause energy splitting with a multimode spectrum centered at the resonance frequency of a single microlaser. In the SUSY microlaser array, selective uniform pumping of the main array with gradual intensity decay extending to its surroundings results in high-power, single-frequency lasing in the fundamental transverse supermode. In this manner, the superpartners are pumped below the lasing threshold, so they still remain dissipative, whereas the gain-loss contrast between the main array and superpartners is sufficiently low to maintain efficient couplings between them, which is equivalent to a system operating in the parity-time symmetric phase (*22*, *25*, *26*). The global mode matching with dissipative superpartners ensures the suppression of all but the fundamental supermode, which leads to high-radiance, single-frequency lasing with substantial power enhancement. The importance of the dissipative superpartners was convincingly validated by the control experiment, in which the entire SUSY microlaser array (including both the main array and the superpartners) was uniformly pumped at the same pumping intensity. Emission collected from the main array shows a multimode spectrum similar to that of the 5 × 5 array in terms of the frequencies of lasing peaks with slight power variations. In the in-phase transverse supermode, all 25 individual microlasers in the

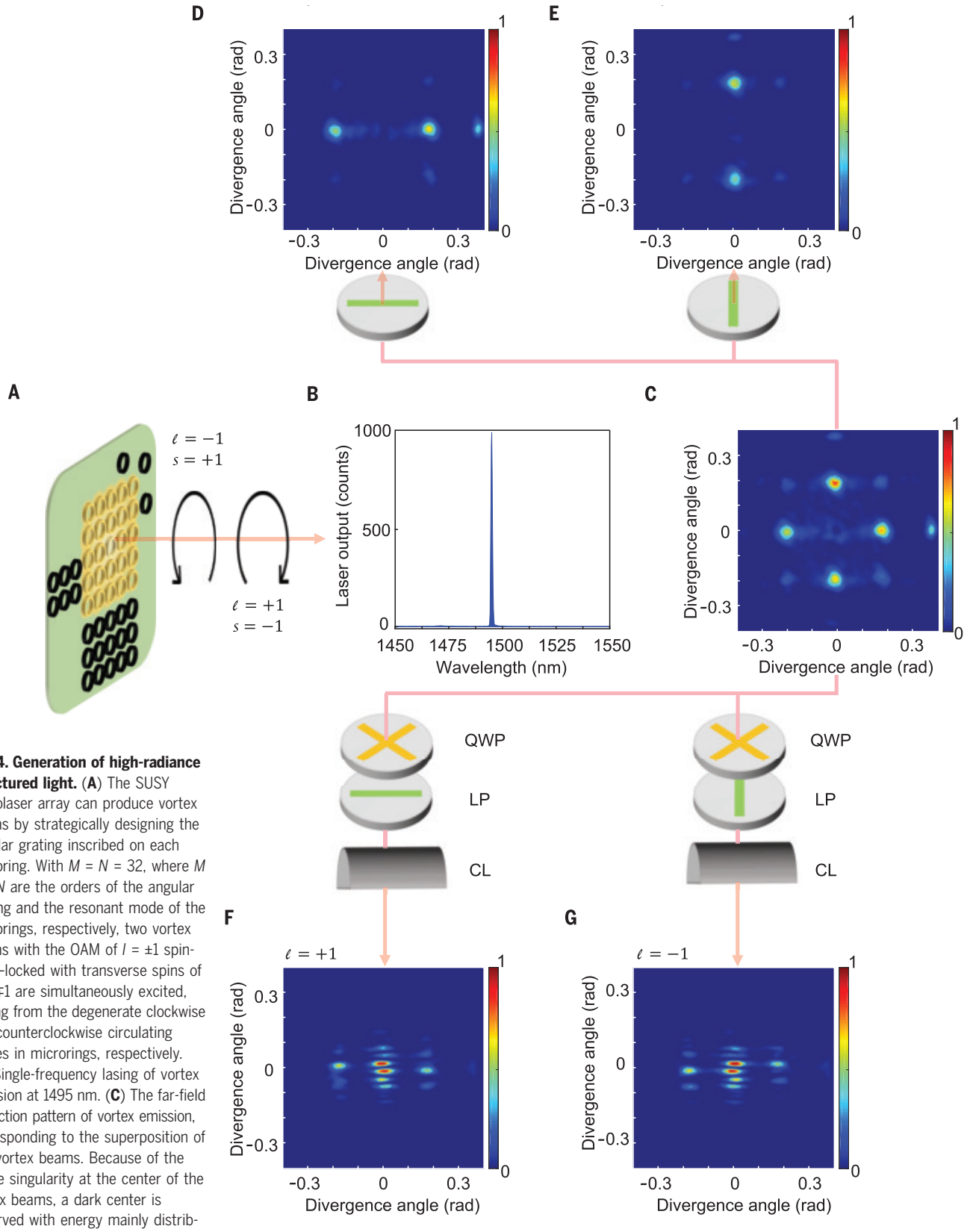


Fig. 4. Generation of high-radiance structured light. (**A**) The SUSY microlaser array can produce vortex beams by strategically designing the angular grating inscribed on each microring. With $M = N = 32$, where M and N are the orders of the angular grating and the resonant mode of the microrings, respectively, two vortex beams with the OAM of $l = \pm 1$ spin-orbit–locked with transverse spins of $s = \mp 1$ are simultaneously excited, arising from the degenerate clockwise and counterclockwise circulating modes in microrings, respectively. (**B**) Single-frequency lasing of vortex emission at 1495 nm. (**C**) The far-field diffraction pattern of vortex emission, corresponding to the superposition of two vortex beams. Because of the phase singularity at the center of the vortex beams, a dark center is observed with energy mainly distributed in the first and second diffraction orders. (**D** and **E**) *x*- and *y*-polarized diffraction patterns of superimposed vortex emissions, showing radially polarized vortex beams. (**F** and **G**) 1D diffraction patterns of the two distinct vortex beams of $l = \pm 1$, filtered by the combined quarter wave plate (QWP) and linear polarizer (LP). The 1D diffraction pattern, equivalent to the 1D Fourier transform of the far-field pattern in (C) along the *y* direction, is captured at the focal plane of a cylindrical lens (CL).

main array oscillate and contribute to power enhancement with a factor of ~25 with respect to emission from a single microlaser, as evidently shown by the light-light curves, where the slope efficiency of the SUSY microlaser array is 26.3 times as high as that of a single microlaser (Fig. 2C). Additionally, the SUSY microlaser array also exhibits a lower lasing threshold because of better optical modal overlap with the gain material.

Beyond power enhancement, the major virtue of the higher-dimensional in-phase supermode lasing is the strong 2D concentration of its emission in the far field, with ultimate energy density quadratically growing with the number of arrayed microlasers. The far-field pattern of the laser beam is a product of the far-field diffraction of the supermode and single microlaser emission (Fig. 3A). As the designed angular grating meets the condition for equiphase emission ($l = 0$), emissions from all scatterers in the single ring add constructively at the center of the far field as they carry the same polarization (Fig. 3B) (*22*). Specifically, the laser radiation from the 2D SUSY microlaser array exhibits a beam divergence of ~2°, compared with that of a single microlaser of ~11°. The combination of power enhancement and narrower divergence results in a two-orders-of-magnitude enhancement in energy density (Fig. 3, C and D). Each microring supports two degenerate modes (clockwise and counterclockwise circulating modes, or their hybridized interfering modes), so each transverse supermode, including the fundamental transverse supermode, carries a twofold degeneracy that contributes to laser actions (*22*). Because of the two degenerate modes being spatially offset by a phase of $\pi/2$, their associated fields are x- and y-polarized as a result from two orthogonal superpositions of the two opposite transverse spins carried by clockwise and counterclockwise modes, respectively. Therefore, the two degenerate modes, leading to different diffraction patterns, can be distinguished by selective measurements of the two polarization states (Fig. 3, E and F). Only the mode with the x polarization, for its in-phase characteristic, contributes to the zero-order diffraction and energy concentration at the center.

It has recently been demonstrated that single microrings offer a convenient strategy for the generation of structured light with spatially inhomogeneous phase variation and polarization distribution, such as optical vortices (*23*, *24*, *27*). Another feature of the higher-dimensional SUSY microlaser array is the ability to maintain the vectorial nature of structured light while improving the total power output. By matching the order of the angular grating with the order of the resonant mode (i.e., $M = N = 32$), the total angular momentum associated with emission becomes zero, which leads to the nonzero OAM of $l = \pm 1$ spin-orbit–locked with transverse spins of $s = \mp 1$ associated with the counterclockwise and clockwise modes in every microring, respectively (Fig. 4A). The desired phase variation and polarization distribution are collectively transferred to the laser beam emitted from the SUSY microlaser array, thereby facilitating single-frequency, high-radiance vortex lasing with a factor of 20.2 in power enhancement (Fig. 4B). Because of the phase singularity at the center of the vortex beams, energy is mainly redistributed to the first and second diffraction orders of the in-phase supermode in the far field (Fig. 4C). The phase fronts of the two vortex beams wind in opposite azimuthal directions, which creates the superposition of left- and right-handed circularly polarized fields with a continuously varying phase delay between them, leading to a vector beam with radial polarizations (*22*); along the horizontal axis, the phase delay is 0, resulting in the x-polarized field, whereas the phase delay becomes $\pm\pi$ in the vertical axis, corresponding to the y-polarized field (Fig. 4, D and E). The two vortex beams with opposite spin-orbit relations can be effectively separated using appropriate combinations of a quarter wave plate and a linear polarizer. Their associated topological charges corresponding to the OAM were characterized by performing 1D Fourier transform of the two vortex beams using a cylindrical lens (Fig. 4, G and F) (*28*). Such astigmatic transformation yields strongly deformed vortices at the focal spot with destructive interference across the 1D diffraction pattern (*22*). Only one dark fringe is clearly observed in the center area, which confirms the topological charge of OAM light of $l = \pm 1$. The orientations of the fringe with respect to the horizontal axis are correlated with the signs of the topological charges.

Our supersymmetric microlaser arrays efficiently produce high-radiance, small-divergence laser beams with an orders-of-magnitude enhancement in energy density. Developing SUSY principles into higher dimensions constitutes a powerful toolbox for effectively tailoring evanescent wave couplings in a large-scale photonic array to synchronize densely packed array elements and prescribe the desired oscillating supermode. Such strategy is generic and applicable to various platforms—for example, coupled vertical-cavity surface-emitting lasers (*2*) or coupled nanolasers (*29*, *30*). Additionally, both phase front and polarization of laser radiation can be spatially structured, taking full advantage of spatial degrees of freedom for deployment in the next generation classical and quantum integrated photonic systems.

REFERENCES AND NOTES

1. D. Botez, D. R. Scifres, *Diode Laser Arrays* (Cambridge Univ. Press, 2005).
2. D. F. Siriani, K. D. Choquette, in *Semiconductors and Semimetals*, vol. 86 of *Advances in Semiconductor Lasers*, J. J. Coleman, A. C. Bryce, C. Jagadish, Eds. (Elsevier, 2012), chap. 6.
3. J. R. Leger, M. L. Scott, W. B. Veldkamp, *Appl. Phys. Lett.* **52**, 1771–1773 (1988).
4. D. Botez, L. Mawst, P. Hayashida, G. Peterson, T. J. Roth, *Appl. Phys. Lett.* **53**, 464–466 (1988).
5. D. F. Siriani, K. D. Choquette, *IEEE J. Quantum Electron.* **47**, 160–164 (2011).
6. T.-Y. Kao, J. L. Reno, Q. Hu, *Nat. Photonics* **10**, 541–546 (2016).
7. R. El-Ganainy, L. Ge, M. Khajavikhan, D. N. Christodoulides, *Phys. Rev. A* **92**, 033818 (2015).
8. M. P. Hokmabadi, N. S. Nye, R. El-Ganainy, D. N. Christodoulides, M. Khajavikhan, *Science* **363**, 623–626 (2019).
9. B. Midya *et al.*, *Photon. Res.* **7**, 363–367 (2019).
10. M. Yoshida *et al.*, *Nat. Mater.* **18**, 121–128 (2019).
11. A. Kodigala *et al.*, *Nature* **541**, 196–199 (2017).
12. M. Dine, *Supersymmetry and String Theory: Beyond the Standard Model* (Cambridge Univ. Press, 2007).
13. F. Cooper, A. Khare, U. Sukhatme, *Phys. Rep.* **251**, 267–385 (1995).
14. N. Sourlas, *Physica D* **15**, 115–122 (1985).
15. M.-A. Miri, M. Heinrich, R. El-Ganainy, D. N. Christodoulides, *Phys. Rev. Lett.* **110**, 233902 (2013).
16. M. Heinrich *et al.*, *Nat. Commun.* **5**, 3698 (2014).
17. S. Longhi, *Phys. Rev. A* **82**, 032111 (2010).
18. M. Heinrich *et al.*, *Opt. Lett.* **39**, 6130–6133 (2014).
19. S. Yu, X. Piao, J. Hong, N. Park, *Nat. Commun.* **6**, 8269 (2015).
20. W. Walasik, B. Midya, L. Feng, N. M. Litchinitser, *Opt. Lett.* **43**, 3758–3761 (2018).
21. S. Barkhofen, L. Lorz, T. Nitsche, C. Silberhorn, H. Schomerus, *Phys. Rev. Lett.* **121**, 260501 (2018).
22. Materials, methods, and additional information are available as supplementary materials.
23. Z. Shao, J. Zhu, Y. Chen, Y. Zhang, S. Yu, *Nat. Commun.* **9**, 926 (2018).
24. Z. Zhang *et al.*, *Science* **368**, 760–763 (2020).
25. L. Feng, R. El-Ganainy, L. Ge, *Nat. Photonics* **11**, 752–762 (2017).
26. B. Peng *et al.*, *Nat. Phys.* **10**, 394–398 (2014).
27. B. Bahari *et al.*, *Nat. Phys.* 10.1038/s41567-021-01165-8 (2021).
28. S. N. Alperin, R. D. Niederriter, J. T. Gopinath, M. E. Siemens, *Opt. Lett.* **41**, 5019–5022 (2016).
29. K. Takeda *et al.*, *Nat. Photonics* **7**, 569–575 (2013).
30. S. S. Deka, S. H. Pan, Q. Gu, Y. Fainman, A. El Amili, *Opt. Lett.* **42**, 4760–4763 (2017).

ACKNOWLEDGMENTS

Funding: We acknowledge support from the US Army Research Office (ARO) (W911NF-19-1-0249 and W911NF-18-1-0348), the National Science Foundation (NSF) (ECCS-1932803, ECCS-1842612, and OMA-1936276), and a Sloan Research Fellowship. This work was partially supported by the NSF through the University of Pennsylvania Materials Research Science and Engineering Center (MRSEC) (DMR-1720530) and was carried out in part at the Singh Center for Nanotechnology, which is supported by the NSF National Nanotechnology Coordinated Infrastructure Program under grant NNCI-1542153. **Author contributions:** B.M., L.F., and N.M.L. conceived the project. B.M. constructed the theoretical model. B.M., X.Q., Z.G., T.W., and L.F. conducted the design and performed numerical simulations. X.Q. fabricated samples. X.Q., Z.G., Z.Z., H.Z., and J.Y. performed measurements. L.F. supervised the study. All authors contributed to data analyses and manuscript preparation. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/403/suppl/DC1
Supplementary Text
Figs. S1 to S20

2 January 2021; accepted 24 March 2021
10.1126/science.abg3904

QUANTUM DEVICES

Josephson junction infrared single-photon detector

Evan D. Walsh[1,2]†, Woochan Jung[3], Gil-Ho Lee[3,4], Dmitri K. Efetov[5], Bae-Ian Wu[6], K.-F. Huang[4], Thomas A. Ohki[7], Takashi Taniguchi[8], Kenji Watanabe[9], Philip Kim[4], Dirk Englund[1,10], Kin Chung Fong[7]*

Josephson junctions are superconducting devices used as high-sensitivity magnetometers and voltage amplifiers as well as the basis of high-performance cryogenic computers and superconducting quantum computers. Although device performance can be degraded by the generation of quasiparticles formed from broken Cooper pairs, this phenomenon also opens opportunities to sensitively detect electromagnetic radiation. We demonstrate single near-infrared photon detection by coupling photons to the localized surface plasmons of a graphene-based Josephson junction. Using the photon-induced switching statistics of the current-biased device, we reveal the critical role of quasiparticles generated by the absorbed photon in the detection mechanism. The photon sensitivity will enable a high-speed, low-power optical interconnect for future superconducting computing architectures.

By exploiting its macroscopic quantum behavior, the Josephson junction (JJ) has become an important superconducting device with a wide array of applications (*1*): high-sensitivity magnetometers, quantum noise limited microwave parametric amplifiers, rapid single-flux quanta in high-performance cryogenic computers, and qubits in superconducting quantum computers. Although JJ performance can be degraded by the quasiparticles (QPs) generated from the breaking of Cooper pairs with radiation, as observed in qubit relaxation (*2*, *3*), the phenomenon can also open opportunities to high-sensitivity photodetection (*4*–*9*). However, despite numerous research efforts to use mechanisms such as nonlinearity in the a.c. Josephson effect, nonequilibrium superconductivity, photo-induced carriers, and bolometric effect (*10*–*19*), single-photon detection with JJs remains challenging. A superconducting tunnel junction (STJ) (*4*) can only be single-photon sensitive by deliberately suppressing the Josephson coupling with an external magnetic field, whereas the quantum capacitance detector requires a large capacitance energy dominating over the Josephson energy for single-photon detection (*9*).

We demonstrate near-infrared (NIR) single-photon detection with current-biased JJs. By coupling photons to a lateral proximity JJ using localized surface plasmons (*20*), we can measure the single-photon–induced JJ switching as a function of current bias, temperature, photon rate, and polarization. The photon-JJ interaction is investigated by illuminating the graphene-based JJ at 27 mK (Fig. 1, A and B) (*17*) [fabrication details are provided in the supplementary materials (*21*)] with a 1550-nm laser. Unlike trilayer JJs, lateral JJs expose directly to the NIR photons. When absorbed into the superconductor, a single NIR photon will break Cooper pairs and generate QPs, which then become a noise source to switch the current-biased JJ. The probability of JJ switching can be described by the resistively and capacitively shunted junction (RCSJ) model (*22*), in which a macroscopic quantum phase particle with a phase difference ϕ between the two superconducting electrodes is subject to a washboard potential (Fig. 1C). When the phase particle is trapped initially in a local minima—$d\phi/dt = 0$—the voltage drop across the JJ is zero (Fig. 1D). The bias current I_b running through the JJ tilts the washboard potential, and the phase particle could escape from the metastable minimum. When it escapes—either through thermal activation (TA) over or macroscopic quantum tunneling (MQT) (*1*) through the barrier, ΔU—the voltage drop across the JJ becomes finite, and the JJ switches to the normal resistive state at a switching current I_s. When the JJ latches into the resistive state after switching, we can register a count, reset the bias current, and over time, measure the switching statistics under different illuminations (Fig. 1E). We studied seven different JJs (*21*) that detect single photons, but we will present results mostly from one device (device A) and compare it with the rest as controls to understand the photon-JJ interaction.

The seemingly random switching with light illumination actually follows the statistics of photon shot noise. After taking 10^4-s time traces at various powers, we produced histograms

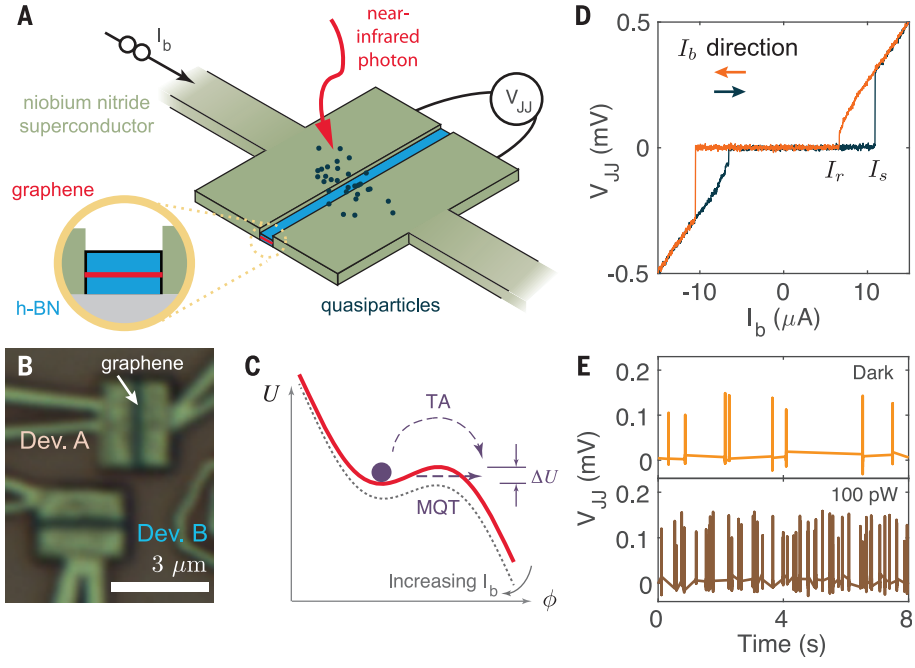


Fig. 1. Current-biased JJ as a SPD. (**A**) QPs are generated as incident photons break Cooper pairs and trigger the JJ to switch. (**B**) Optical image of two graphene-based JJs. (**C**) The behavior of the current-biased JJ can be described by a macroscopic quantum phase particle subjected to a tilted-washboard potential in the RCSJ model. QPs can assist the phase particle to move out of the trapping potential, switching the JJ to a normal resistive state. (**D**) Typical IV curve displaying the switching and retrapping of JJ at I_s and I_r, respectively. (**E**) Switching events in dark and 100 pW of 1550-nm light at I_b = 10.90 μA.

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [3]Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea. [4]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [5]ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain. [6]Air Force Research Laboratory, Wright-Patterson AFB, OH 45433, USA. [7]Raytheon BBN Technologies, Quantum Engineering and Computing Group Cambridge, MA 02138, USA. [8]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, Tsukuba, Japan. [9]Research Center for Functional Materials, National Institute for Materials Science, Tsukuba, Japan. [10]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
†Present address: Systems & Technology Research
*Corresponding author. Email: kc.fong@rtx.com

of switching events using 1-s bins (Fig. 2A). A Poisson distribution (Fig. 2A, solid lines) traces the experimental data closely (Fig. 2A, bars), as expected for uncorrelated switching events. Without any fitting, the data in the Fig. 2A inset also demonstrates the property of Poisson statistics that the variance of the count is equal to the mean. Furthermore, we found the switching probability in a 20-μs-long time window, accounting for an offset due to the false positive (dark) count, depends linearly on the average absorbed photon number, $\langle N_{\text{photon}}\rangle$ (Fig. 2B). $\langle N_{\text{photon}}\rangle$ is calculated by using the rate of absorbed photons by the JJ, $\mathcal{R}_{\text{photon}}$, as calibrated subsequently to be 53 s^{-1} at 100 pW laser power with the polarization measurement. Higher I_b produces higher switching probability because of higher intrinsic quantum efficiency and dark count (*17*). Because the Poisson probability for measuring m photons in a detection time window reduces to $\simeq\langle N_{\text{photon}}\rangle^m/m!$ for $\langle N_{\text{photon}}\rangle \ll 1$, this linearity proves that our JJ detects single photons from a dim coherent source (*5*).

It is imperative to distinguish JJ switching induced by single photons from steady heating from laser power. Therefore, we compared the I_s distributions by ramping up I_b under three conditions 1000 times each: (i) at 27 mK without light, (ii) at 27 mK with light, and (iii) at 325 mK without light (Fig. 3A). The I_s distribution reflects the stochasticity of the phase particle escape in the tilted-washboard potential and measures distinctively the phase particle escape rate without light, $\Gamma_{\text{esc}}^{(\text{dark})}$ (*23*). Although both average values $\langle I_s\rangle_{(\text{ii})}$ and $\langle I_s\rangle_{(\text{iii})}$ are suppressed equally—through photon illumination and operating at a higher temperature, respectively—they do not have the same shape distribution. This result contrasts sharply to the previous experiments in which JJs under photon illumination had I_s distributions matching those of JJs with elevated temperature (*15*, *18*). The difference of I_s distributions in Fig. 3A excludes the steady heating of the JJ from laser power as the cause of increased JJ switching.

To study the switching mechanism, we compared the measured switching rate, Γ_{meas}, versus I_b with and without illumination (Fig. 3B). Without light, $\Gamma_{\text{meas}}^{(\text{dark})}$ directly measures $\Gamma_{\text{esc}}^{(\text{dark})}$ by definition. Up until about 4000 s^{-1} limited by the d.c. bias resetting circuitry, $\Gamma_{\text{meas}}^{(\text{dark})}$ increases with I_b because of a stronger escape tendency of the phase particle as the barrier height $\Delta U \simeq (4\sqrt{2}/3)(\hbar I_c/2e)(1-I_b/I_c)^{3/2}$ decreases, where I_c is the JJ critical current and $\hbar$ and e are Planck's constant h divided by 2π and the electron charge, respectively. The nearly linear dependence of $\Gamma_{\text{meas}}^{(\text{dark})}$ in the semilog plot agrees with an escape rate $\Gamma_{\text{esc}}^{(\text{dark})} \propto e^{-\Delta U/k_B T_{\text{esc}}}$, where k_B and T_{esc} are the Boltzmann constant and escape temperature, respectively. This is a characteristic exponential form in reaction-rate theory with an escape energy of $k_B T_{\text{esc}}$ (*24*). For TA escape, $k_B T_{\text{esc}}$ is given by the thermal energy, whereas MQT escape is given by the energy of the harmonic oscillator at the local minima of the tilted-washboard potential: $k_B T_{\text{esc}} \simeq \hbar\omega_P/7.2(1+0.87/Q)$, where ω_P and Q are the JJ plasma frequency under current bias and the quality factor of the harmonic potential, respectively. $\Gamma_{\text{meas}}^{(\text{dark})}$ is best fit by the MQT theory (Fig. 3B, solid black line) with an I_c of 11.99 μA so that $\omega_P/2\pi$ = 225 GHz at zero I_b and $Q \simeq 1.1$ (*21*).

Whereas $\Gamma_{\text{meas}}^{(\text{dark})} = \Gamma_{\text{esc}}^{(\text{dark})}$, $\Gamma_{\text{meas}}^{(\text{light})}$ in the light does not directly measure $\Gamma_{\text{esc}}^{(\text{light})}$, the single-photon-enhanced escape rate. For a brief moment Δt_1 when $\Gamma_{\text{esc}}^{(\text{light})} \gg \Gamma_{\text{esc}}^{(\text{dark})}$, the phase particle escape probability is $\mathcal{P}_{\text{esc}}^{(\text{light})} = 1 - \exp\left(-\int_0^{\Delta t_1}\Gamma_{\text{esc}}^{(\text{light})}dt\right)$. To simplify our calculation, we considered a constant $\Gamma_{\text{esc}}^{(\text{light})}$ during Δt_1 and focused on the lower I_b regime where $\mathcal{P}_{\text{esc}}^{(\text{light})} \simeq \Gamma_{\text{esc}}^{(\text{light})}\Delta t_1 \ll 1$. With negligible $\Gamma_{\text{meas}}^{(\text{dark})}$, $\Gamma_{\text{meas}}^{(\text{light})} \simeq \mathcal{P}_{\text{esc}}^{(\text{light})}\mathcal{R}_{\text{photon}} \simeq \Gamma_{\text{esc}}^{(\text{light})}\Delta t_1 \mathcal{R}_{\text{photon}}$. Assuming that Δt_1 does not depend on I_b exponentially, the nearly linear $\Gamma_{\text{meas}}^{(\text{light})}$ in the semilog plot suggests that $\Gamma_{\text{esc}}^{(\text{light})}$ follows the reaction-rate theory because the single-photon–induced JJ switching can be described by an enhanced escape energy, $k_B T_{\text{esc}}^*$. More quantitatively, we fit the data using

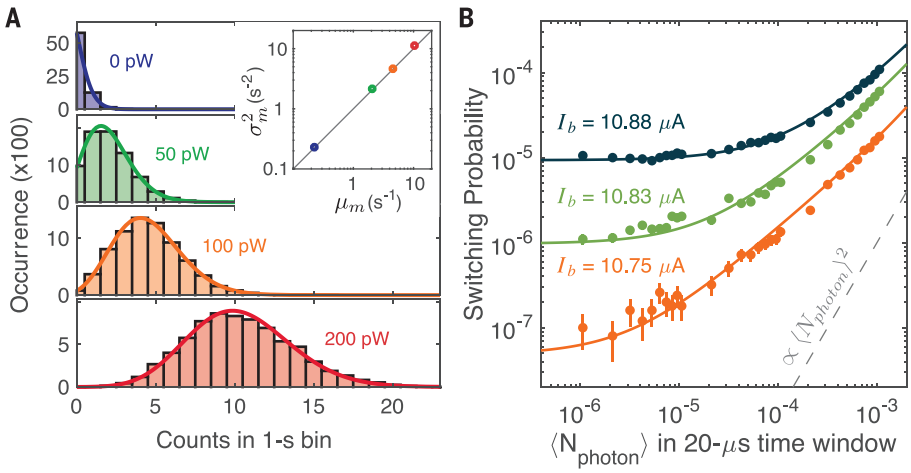


Fig. 2. Switching of JJ by single NIR photon. (**A**) The observed switching events under illumination (0 to 200 pW) follow Poisson statistics. (Inset) The statistical variance of counts σ_m^2, the photon shot noise, equals to its mean μ_m without fitting (solid line). (**B**) The switching probability as a function of average absorbed photon number $\langle N_{\text{photon}}\rangle$ in a 20-μs time window. Switching probability is linearly proportional to $\langle N_{\text{photon}}\rangle$ with an offset owing to dark count. The solid line shows the fitting weighted with the inverse of standard deviation (error bars).

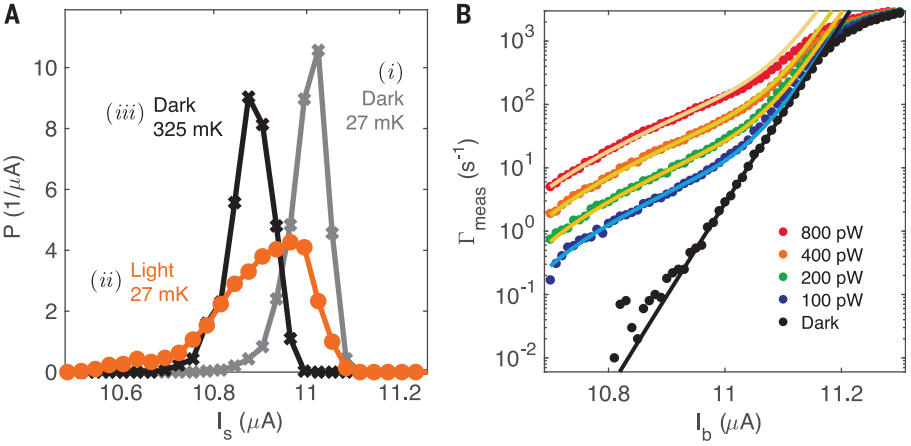


Fig. 3. Switching mechanism. (**A**) The distribution of switching current I_s by ramping up I_b at 1 μA/s under three conditions: (i) at 27 mK in the absence and (ii) the presence (orange) of 100 pW of 1550-nm light and (iii) at 325 mK in the absence of light. (ii) and (iii) have the same averaged I_s but different shapes, thus ruling out steady heating from laser power. (**B**) The switching rate as a function of bias current with various incident laser powers. The points are measured values, the black line is the best fit of the MQT process to Γ_{meas} without light, and the color lines are modeled results from photon-enhanced $\Gamma_{\text{esc}}^{(\text{light})}$. The linearity of the data on a semilog plot indicates a momentary elevation of the escape energy by the photon.

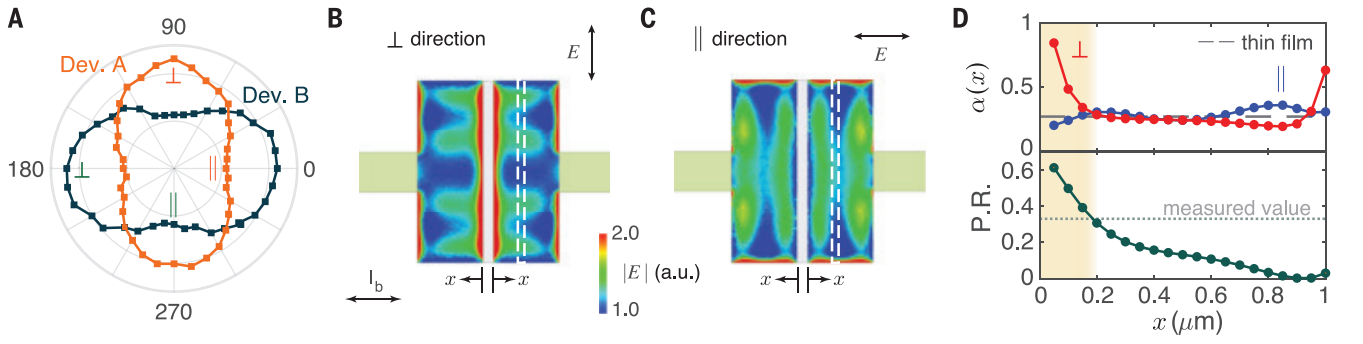


Fig. 4. NIR photon coupling into JJ. (**A**) The polarization dependence of Γ_{meas} for devices A and B (same orientation as in Fig. 1B). (**B** and **C**) The simulated electric field strength for incident light perpendicular and parallel to the supercurrent direction, respectively. The absorption occurs in the NbN contacts, enhanced through coupling to localized surface plasmons (red). (**D**) Averaged volumetric NIR absorption α of the two polarizations versus distance *x* from the graphene (dashed line for α for thin-film NbN) and the polarization ratio versus *x*. The dotted line marks the measured ratio 33% from (A), suggesting an effective absorption length of 190 nm (shaded region) in NbN.

T^*_{esc} and Δt_1 as two free parameters accounting for MQT, TA, and phase diffusion processes (*21*, *24*). Solid color lines in Fig. 3B show the results with $T^*_{esc} \simeq 2.1$ K and $\Delta t_1 \simeq 0.86$ ns. We attribute T^*_{esc} to an increased effective temperature ($>T_{esc}$) caused by QPs generated in the superconducting contacts that will be discussed later. The close fit of the model to the data indicates that a photon induces JJ switching by momentarily increasing the escape energy.

To understand how the JJ absorbs single photons, we studied the dependence of $\Gamma^{(light)}_{meas}$ on the polarization of incident light (*25*). This angular dependence is plotted in Fig. 4A for two equally sized JJs oriented orthogonally to one another on the same substrate (Fig. 1B) with arbitrarily oriented superconducting electrodes connecting the JJs to wire-bonding pads. Both devices exhibited the same photon rate but offset by 90° to each other. This dependence on junction geometry suggests that the detected single photons are absorbed at the JJs, not at the electrodes. By inserting and rotating a polarizer in front of the JJs in a series of experiments (*21*), we found that the maximum $\Gamma^{(light)}_{meas}$ coincides with the polarization perpendicular (⊥) to the JJ supercurrent flow.

Using HFSS (high-frequency structure simulator) (Ansys, Canonsburg, PA) (*21*), we studied the angular dependence by computing the volumetric photo-absorption for the entire junction area with an incident plane wave approximation because the JJ is much smaller than and located at the center of the Gaussian beam waist of 2.41 mm. In Fig. 4, B and C, we plot the field distribution within the niobium nitride (NbN). When the real part of the permittivity of NbN (−6.52 + 60.88*i*) (*26*) and its adjacent medium have opposite signs, they form effectively a capacitor-inductor network that supports a localized surface plasmon along the interface (*20*). With a photon polarization in the ⊥ direction, the plasmon coupling intensifies the field and enhances the photon absorption at the graphene-NbN interface, resulting in the observed polarization-dependent Γ_{meas}.

We can estimate the effective area of the JJ as a single-photon detector (SPD) by matching the measured and simulated polarization ratios. By defining volumetric absorption zones (Fig. 4, B and C, dashed lines) at distance *x* from the graphene-NbN interface with a 50 nm by 2.8 μm area, the spatial dependence of photon absorption coefficient $\alpha_j(x)$ is plotted in Fig. 4D, where $j = \perp, \parallel$ notates the polarization direction. $\alpha_\perp$ and $\alpha_\parallel$ differ considerably near the edge because of the surface plasmon but approach the same value, α = 0.27 (dashed line), further away from the edge. This value agrees with the expected photo-absorption of 50-nm thin-film NbN (*25*) and verifies our calculation. Photons could be absorbed directly into the graphene at rates of 0.07 and 0.27 s^{-1} with 100 pW of laser power in the ⊥ and ∥ directions, respectively. This is much smaller than $\Gamma^{(light)}_{meas}$ and orthogonal to the measured polarization dependence, thus excluding the detected single photons being absorbed directly in graphene. We define the polarization ratio as $[\int_0^x(\alpha_\perp - \alpha_\parallel)dx']/[\int_0^x(\alpha_\perp + \alpha_\parallel)dx']$, which sums up the photon absorption with increasing distance from the graphene-NbN interface (*25*). To match the measured polarization ratio of 33% from Fig. 4A, the simulation suggests that the photons absorbed within 190 nm of the edge are those that trigger the JJ switching events. Therefore, the total effective area of our JJ SPD is about 1.06 μm^2 with an averaged α of 0.58 in the ⊥ polarization direction, resulting in the $\mathcal{R}_{photon}$ value (*21*) used for calculating $\langle N_{photon}\rangle$ in Fig. 2B and fitting data in Fig. 3B.

Analogous to the QP-induced qubit relaxation (*2*, *3*), we argue that QPs mediate the photon-induced JJ switching. When a single NIR photon impinges onto the superconductor, it breaks Cooper pairs and generates ~260 QPs, given by $\eta\hbar\omega_{photon}/\Delta$, where $\omega_{photon}/2\pi$ is the photon frequency, Δ is the superconducting gap energy, and $\eta \simeq 0.57$ is the down-conversion efficiency (*27*). Before recombining through inelastic scattering, these QPs can diffuse across the JJ, resulting in a diffusion current $I_{diff} = (e/4\sqrt{\pi\mathcal{D}t^3})xe^{(-x^2/4\mathcal{D}t)}$ per QP at the JJ, where $\mathcal{D} = 0.55$ cm^2/s is the diffusion constant in NbN (*25*). For x = 100 nm, $I_{diff} \simeq$ 25 nA, which rises and subsides in a characteristic time scale of $x^2/\mathcal{D} \simeq 0.2$ ns. I_{diff} is much smaller than the required $I_c - I_b$ to directly switch the JJ. However, unless scattered inelastically and trapped inside the JJ, these QPs at the gap energy can diffuse across the JJ with a mean-free-path of ~92 nm in graphene through the 160-nm-long channel. The JJ in our experiment is quasi-ballistic, so that the I_{diff} could generate a shot noise $S_I = 2e\mathcal{F}I_{diff}$ with a Fano factor $\mathcal{F}$ on the order of 1. This shot noise can enhance JJ switching by exciting the phase particle to higher-energy states, thus increasing effectively its T_{esc} to ~1 K (*21*, *28*), which is similar to our fitted value of T^*_{esc} in Fig. 3B. I_{diff} might rise above 1 μA for small x—that is, 10 nm. However, the characteristic time scale of I_{diff} is much faster than $1/\omega_P$, invalidating the use of the T^* equation (*28*). Theoretical calculations of $\Gamma^{(light)}_{esc}$ will, because of both adiabatic and nonadiabatic changes of the washboard potential, inform the specific microscopic mechanism of single-photon–induced switching and optimize the SPD efficiency. We could further improve the coupling efficiency by using nanophotonics (*17*). Our result suggests that qubits made of lateral junctions may need additional protection from radiation because of photon coupling enhancement. The critical role of graphene under this hypothetical mechanism is to provide a shunt resistor, quasi-ballistic channel across the JJ and to form a lateral JJ that allows for the efficient coupling of photons by the dissipative surface plasmons. This new mechanism is consistent with the lack of large dependence of gate voltage and graphene thickness in subsequent experiments (*21*). Because we can control the resistance of our graphene-based JJ SPD, this feature will allow

the matching of load impedance with JJ-based computer architectures to enable high-speed, low-power JJ optical interconnects.

REFERENCES AND NOTES

1. T. Francesco Ed., *Fundamentals and Frontiers of the Josephson Effect*, *Springer Series in Materials Science*, vol. 286 (Springer, 2019).
2. D. Ristè *et al*., *Nat. Commun.* **4**, 1913 (2013).
3. A. P. Vepsäläinen *et al*., *Nature* **584**, 551–556 (2020).
4. A. Peacock *et al*., *Nature* **381**, 135–137 (1996).
5. G. N. Gol'tsman *et al*., *Appl. Phys. Lett.* **79**, 705–707 (2001).
6. B. Korzh *et al*., *Nat. Photonics* **14**, 250–255 (2020).
7. J. E. Sadleir, S. J. Smith, S. R. Bandler, J. A. Chervenak, J. R. Clem, *Phys. Rev. Lett.* **104**, 047003 (2010).
8. B. Karasik *et al*., *Appl. Phys. Lett.* **101**, 052601 (2012).
9. P. M. Echternach, B. J. Pepper, T. Reck, C. M. Bradford, *Nat. Astron.* **2**, 90–97 (2018).
10. W. R. McGrath, P. L. Richards, D. W. Face, D. E. Prober, F. L. Lloyd, *J. Appl. Phys.* **63**, 2479–2491 (1998).
11. T. Schäpers, R. P. Müller, A. Kaluza, H. Hardtdegen, H. Lüth, *Appl. Phys. Lett.* **75**, 391–393 (1999).
12. F. Stella *et al*., *Appl. Phys. Lett.* **92**, 202505 (2008).
13. F. Giazotto *et al*., *Appl. Phys. Lett.* **92**, 162507 (2008).
14. G. Oelsner *et al*., *Appl. Phys. Lett.* **103**, 142605 (2013).
15. Y. Wang *et al*., *Physica C* **515**, 49–53 (2015).
16. K. Tsumura *et al*., *Appl. Phys. Lett.* **108**, 033109 (2016).
17. E. D. Walsh *et al*., *Phys. Rev. Appl.* **8**, 024022 (2017).
18. G.-H. Lee *et al*., *Nature* **586**, 42–46 (2020).
19. R. Kokkoniemi *et al*., *Nature* **586**, 47–51 (2020).
20. N. Engheta, *Science* **317**, 1698–1702 (2007).
21. Materials and methods are available as supplementary materials.
22. M. Tinkham, *Introduction to Superconductivity* (Dover, ed. 2, 2004).
23. T. A. Fulton, L. N. Dunkleberger, *Phys. Rev. B* **9**, 4760–4768 (1974).
24. J. Männik *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **71**, 220509 (2005).
25. E. F. C. Driessen *et al*., *Eur. Phys. J. Appl. Phys.* **47**, 10701 (2009).
26. A. Semenov *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **80**, 054510 (2009).
27. T. Guruswamy, D. J. Goldie, S. Withington, *Supercond. Sci. Technol.* **28**, 054002 (2015).
28. J. P. Pekola *et al*., *Phys. Rev. Lett.* **95**, 197004 (2005).
29. B.-I. Wu, K. C. Fong, Simulation file for: Josephson-junction infrared single-photon detector. Zenodo (2021); https://doi.org/10.5281/zenodo.4624530.

ACKNOWLEDGMENTS

We thank L. Levitov, M. Shaw, T. Heikkilä, and L. Govia for valuable discussions. **Funding:** The work of E.D.W. and D.E. was supported in part by the Army Research Laboratory Institute for Soldier Nanotechnologies program W911NF-18-2-0048 and the U.S. Army Research Laboratory (award W911NF-17-1-0435). D.E. was also supported in part by AFOSR program FA2386-20-1-4070 "Developing graphene Josephson microwave single-photon detector for quantum information science." W.J. and G.-H.L. were supported by the National Research Foundation of Korea (NRF) funded by the Korean Government (grants 2016R1A5A1008184, 2020R1C1C1013241, and 2020M3H3A1100839), Samsung Science and Technology Foundation (project SSTF-BA1702-05) and Samsung Electronics Co. (IO201207-07801-01). D.K.E. acknowledges support from the Ministry of Economy and Competitiveness of Spain through the "Severo Ochoa" program for Centres of Excellence in R&D (SE5-0522), Fundació Privada Cellex, Fundació Privada Mir-Puig, the Generalitat de Catalunya through the CERCA program, the H2020 Programme under grant agreement 820378, Project: 2D-SIPC, and the La Caixa Foundation. B.-I.W. was supported by the Air Force Office of Scientific Research under award FA9550-16RYCOR290. T.A.O. and K.C.F. acknowledge support from Raytheon BBN Technologies. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by MEXT, Japan, grant JPMXP0112101001; JSPS KAKENHI grant JP20H00354; and CREST (JPMJCR15F3), JST. P.K. and K.C.F. were supported in part by the Army Research Office under Cooperative Agreement W911NF-17-1-0574. **Author contributions:** G.-H.L., D.E., and K.C.F. conceived the project. W.J., G.-H.L., and K.-F.H. fabricated the samples. T.T. and K.W. provided the hBN crystal. W.J., G.-H.L., E.D.W., and K.C.F. performed the measurements. G.-H.L., E.D.W., B.-I.W., and K.C.F. performed the simulations. E.D.W., G.-H.L., D.K.E., B.-I.W., T.A.O., P.K., D.E., and K.C.F. performed the data analysis and wrote the paper. P.K., D.E., and K.C.F. supervised the project. **Competing interests:** None declared. **Data and materials availability:** All (other) data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. HFSS simulation code is available online (*29*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/409/suppl/DC1
Materials and Methods
Figures S1 to S5
Tables S1 to S4
References (*30*–*44*)

3 November 2020; accepted 22 March 2021
10.1126/science.abf5539

CORONAVIRUS

Timing the SARS-CoV-2 index case in Hubei province

Jonathan Pekar[1,2], Michael Worobey[3]*, Niema Moshiri[4], Konrad Scheffler[5], Joel O. Wertheim[6]*

Understanding when severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged is critical to evaluating our current approach to monitoring novel zoonotic pathogens and understanding the failure of early containment and mitigation efforts for COVID-19. We used a coalescent framework to combine retrospective molecular clock inference with forward epidemiological simulations to determine how long SARS-CoV-2 could have circulated before the time of the most recent common ancestor of all sequenced SARS-CoV-2 genomes. Our results define the period between mid-October and mid-November 2019 as the plausible interval when the first case of SARS-CoV-2 emerged in Hubei province, China. By characterizing the likely dynamics of the virus before it was discovered, we show that more than two-thirds of SARS-CoV-2–like zoonotic events would be self-limited, dying out without igniting a pandemic. Our findings highlight the shortcomings of zoonosis surveillance approaches for detecting highly contagious pathogens with moderate mortality rates.

In late December of 2019, the first cases of COVID-19, the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), were described in the city of Wuhan in Hubei province, China (*1*, *2*). The virus quickly spread within China (*3*). The cordon sanitaire that was put in place in Wuhan on 23 January 2020 and mitigation efforts across China eventually brought about an end to sustained local transmission. In March and April 2020, restrictions across China were relaxed (*4*). By then, however, COVID-19 was a pandemic (*5*).

A concerted effort has been made to retrospectively diagnose the earliest cases of COVID-19 and thus determine when the virus first began transmitting among humans. Both epidemiological and phylogenetic approaches suggest an emergence of the pandemic in Hubei province at some point in late 2019 (*2*, *6*, *7*). The first described cluster of COVID-19 was associated with the Huanan Seafood Wholesale Market in late December 2019, and the earliest sequenced SARS-CoV-2 genomes came from this cluster (*8*, *9*). However, this market cluster is unlikely to have denoted the beginning of the pandemic, as COVID-19 cases from early December lacked connections to the market (*7*). The earliest such case in the scientific literature is from an individual retrospectively diagnosed on 1 December 2019 (*6*). Notably, however, newspaper reports document retrospective COVID-19 diagnoses recorded by the Chinese government going back to 17 November 2019 in Hubei province (*10*). These reports detail daily retrospective COVID-19 diagnoses through the end of November, suggesting that SARS-CoV-2 was actively circulating for at least a month before it was discovered.

Molecular clock phylogenetic analyses have inferred the time of the most recent common ancestor (tMRCA) of all sequenced SARS-CoV-2 genomes to be in late November or early December 2019, with uncertainty estimates typically dating to October 2019 (*7*, *11*, *12*). Crucially, though, this tMRCA is not necessarily equivalent to the date of zoonosis or index case infection (*13*, *14*) because coalescent processes can prune basal viral lineages before they have the opportunity to be sampled, potentially pushing SARS-CoV-2 tMRCA estimates forward in time from the index case by days, weeks, or months. For a point of comparison, consider the zoonotic origins of the HIV-1 pandemic, whose tMRCA in the early 20th century coincided with the urbanization of Kinshasa, in what is now the Democratic Republic of the Congo (*15*, *16*), but whose cross-species transmission from a chimpanzee reservoir occurred in southeast Cameroon, likely predating the tMRCA of sampled HIV-1 genomes by many years (*17*). Despite this important distinction, the tMRCA has been

[1]Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093, USA. [2]Department of Biomedical Informatics, University of California San Diego, La Jolla, CA 92093, USA. [3]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. [4]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA. [5]Illumina, Inc., San Diego, CA 92122, USA. [6]Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA.
***Corresponding author. Email: worobey@arizona.edu (M.W.); jwertheim@health.ucsd.edu (J.O.W.)**

frequently conflated with the date of the index case infection in the SARS-CoV-2 literature (*7*, *18*, *19*).

Here, we combine retrospective molecular clock analysis in a coalescent framework with a forward compartmental epidemiological model to estimate the timing of the SARS-CoV-2 index case in Hubei province. The inferred dynamics during these unobserved early days of SARS-CoV-2 highlight challenges in detecting and preventing nascent pandemics.

We first explored the evolutionary dynamics of the first wave of SARS-CoV-2 infections in China. We used Bayesian phylodynamics (*20*) to reconstruct the underlying coalescent processes using a Bayesian Skyline approach for 583 SARS-CoV-2 complete genomes, sampled in China between when the virus was first discovered at the end of December 2019 and the last of the non-reintroduced circulating virus in April 2020. Applying a strict molecular clock, we inferred an evolutionary rate of 7.90×10^{-4} substitutions per site per year [95% highest posterior density (HPD): 6.64×10^{-4} to 9.27×10^{-4}]. The tMRCA of these circulating strains was inferred to fall within a 34-day window with a mean of 9 December 2019 (95% HPD: 17 November to 20 December) (Fig. 1). This estimate accounts for the many disparate inferred rooting orientations [see supplementary text and (*21*)]. Notably, 78.7% of the posterior density postdates the earliest published case on 1 December, and 95.1% postdates the earliest reported case on 17 November. Relaxing the molecular clock provides a similar tMRCA estimate, as does applying a Skygrid coalescent approach (fig. S1). The recency of this tMRCA estimate in relation to the earliest documented COVID-19 cases obliges us to consider the possibility that this tMRCA does not capture the index case and that SARS-CoV-2 was circulating in Hubei province before the inferred tMRCA.

If the tMRCA postdates the earliest documented cases, then the earliest diverged SARS-CoV-2 lineages must have gone extinct (Fig. 2). As these early basal lineages disappeared, the tMRCA of the remaining lineages would move forward in time (fig. S2). Thus, we interrogated the posterior trees sampled from the phylodynamic analysis to determine whether this time of coalescence had stabilized before the sequencing of the first SARS-CoV-2 genomes on 24 December 2019 or whether this process of basal lineage loss was ongoing in late December and/or early January. Notably, these basal lineages need not be associated with specific mutations, as the phylodynamic inference reconstructs the coalescent history, not the mutational history (*20*).

We find only weak evidence for basal lineage loss between 24 December 2019 and 13 January 2020 (fig. S3A). The root tMRCA is within 1 day of the tMRCA of virus sampled on or after 1 January 2020 in 78.5% of posterior samples (fig. S3B). The tMRCA of genomes sampled on or after 1 January 2020 is 3 days later than the tMRCA of all sampled genomes. By contrast, the mean tMRCA does not change when considering genomes sampled on or after 1 January 2020 versus on or after 13 January 2020. This consistency indicates a stabilization of coalescent processes at the start of 2020, when an estimated total of 1000 people had been infected with SARS-CoV-2 in Wuhan (*22*). Nonetheless, to account for the weak signal of a delay in reaching a stable coalescence (i.e., the point in time at which basal lineages cease to be lost), we identified the tMRCA for all viruses sampled on or after 1 January 2020 (i.e., at the time of stable coalescence) for each tree in the posterior sample.

Phylogenetic analysis alone cannot tell us how long SARS-CoV-2 could have circulated in Hubei province before the tMRCA. To answer this question, we performed forward epidemic simulations (*23*). These simulations were initiated by a single index case using a compartmental epidemiological model across scale-free contact networks (mean number of contacts: 16). This compartmental model was previously developed to describe SARS-CoV-2 transmission dynamics in Wuhan (*22*). This model, termed SAPHIRE, includes compartments for susceptible (S), exposed (E), presymptomatic (P), unascertained (A), ascertained (I), hospitalized (H), and removed (R) individuals. Our simulations used parameters from the time period before COVID-19 mitigation efforts, from 1 January through 22 January 2020 (table S1), based on the work of Hao *et al.* (*22*). We analyzed 1000 epidemic simulations that resulted in ≥1000 total infected people. These simulated epidemics had a median doubling time of 4.1 days (95% range across simulations: 2.7 to 6.7), matching premitigation incidence trends in Wuhan (table S2).

We simulated coalescent processes across the transmission network to determine the tMRCA of the virus at the end of the simulation. This approach allowed us to determine the distribution of the expected number of days between index case infection and the stable coalescence (i.e., tMRCA) (Fig. 2). The median number of days between index case infection and this tMRCA was 8.0 days (95% range: 0.0 to 41.5 days) (Fig. 3A). The median time between index case infection and the first person exiting the presymptomatic phase (i.e., ascertained or unascertained infection) was 5.7 days (95% range: 0.9 to 15.7 days).

As a robustness check, we also simulated epidemics with more densely and more sparsely connected contact networks (mean: 26 and 10 contacts, respectively), rescaling the per-contact transmission rate to maintain empirical epidemic growth dynamics. We also explored the effects of faster (mean: 3.1 days; 95% range: 2.0 to 5.1 days) and slower (mean: 5.3 days; 95% range: 3.6 to 7.5 days) epidemic doubling times (table S2). Slower transmission rates led to more days between the index case and the stable coalescence, but modifying the density of the contact network had minimal effect on this time interval (fig. S4).

To estimate the date of infection for the index case in Hubei province, we combined the retrospective molecular clock analysis with the forward epidemic simulations (fig. S5). We identified the stable tMRCA in the posterior trees as an anchor to the real-world calendar dates and then extended this date back in time according to the number of days between the index case infection and the time of stable coalescence from the compartmental epidemic simulations. However, a random sample of tMRCAs and days from index case infection to coalescence will not produce epidemiologically meaningful results because many of these combinations do not precede the earliest dates of

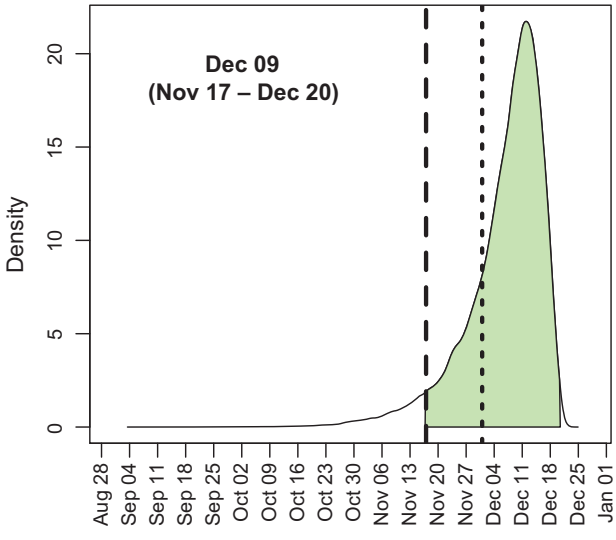


Fig. 1. Posterior distribution for the tMRCA of 583 sampled SARS-CoV-2 genomes circulating in China between December 2019 and April 2020. Inference was performed by using a strict molecular clock and a Bayesian Skyline coalescent prior. The shaded area denotes 95% HPD. The long-dashed line represents 17 November 2019, and the short-dashed line represents 1 December 2019.

reported COVID-19 cases. Therefore, we implemented a rejection sampling–based approach to generate a posterior distribution of dates of infection for the Hubei index case, conditioning on at least one individual who had progressed past the presymptomatic stage in the simulated epidemic before the date of the first reported COVID-19 case (see materials and methods and fig. S6).

In our primary analysis, we assume that 17 November represents the first documented case of COVID-19 (ascertained or unascertained in the SAPHIRE model). Under this assumption, the median number of days between index case infection and stable coalescence after rejection sampling is 37 days (95% HPD: 12 to 55 days) (Fig. 3B). Consequently, the index case in Hubei likely contracted SARS-CoV-2 on or around 4 November 2019 (95% upper HPD: 15 October; 99% upper HPD: 7 October) (Fig. 3C).

This time frame for the Hubei index case is robust (fig. S7). Epidemic simulations with faster or slower transmission rates and more or less densely connected contact networks produce similar date estimates. Furthermore, using the root tMRCA from the sampled posterior trees, rather than adjusting for the shifting coalescence between 24 December 2019 and 1 January 2020, produces a median date of 3 November (95% upper HPD: 14 October; 99% upper HPD: 4 October). Incorporating a relaxed molecular clock or adjusting the coalescent prior assumption (Skyline versus Skygrid) also had minimal effect.

If we enforce that there must be at least one ascertained case in our simulations before 17 November 2019, the median date of the Hubei index case is pushed back about a week to 28 October (95% upper HPD: 11 October; 99% upper HPD: 5 October) (fig. S7). However, the distinction between ascertained and unascertained in the original SAPHIRE model was meant to reflect the probability of missed diagnoses in January 2020 of the Wuhan epidemic and does not account for the investigations that resulted in retrospective diagnoses in November and December 2019.

If we discount the reported evidence of retrospective COVID-19 diagnoses throughout the end of November and instead take 1 December as representing the first confirmed case of COVID-19, then the median time between index case infection and stable coalescence after rejection sampling is 23 days (95% range: 1 to 47 days) (Fig. 3D). Under this scenario, the index case in Hubei would have contracted SARS-CoV-2 on or around 17 November 2019 (95% upper HPD: 24 October; 99% upper HPD: 13 October) (Fig. 3E). Similar dates are inferred with a relaxed molecular clock and conditioning on an ascertained infection by 1 December (fig. S7).

It is reasonable to postulate that the variant of SARS-CoV-2 that first emerged was less fit than the variant that spread through China and that evolutionary adaptation was critical to its establishment in humans (*12*). Therefore, we simulated two-phase epidemics in which the index case was infected with a less-fit variant (i.e., half as transmissible) that went extinct, but not before giving rise to a mutant strain matching the transmission dynamics estimated in Wuhan (figs. S8 and S9). If we condition on an ascertained or unascertained COVID-19 case (due to either the original or adapted variant) by 17 November, the original variant transmits for a median of 5 days (95% HPD: 0 to 36 days) before the adapted strain emerges (Fig. 3F); this adapted strain then circulates for a median of 28 days (95% HPD: 0 to 52 days) before reaching a stable coalescence (Fig. 3G). In this scenario, the index case in Hubei would have likely contracted an unobserved variant of SARS-CoV-2 on 5 November 2019 (95% upper HPD: 11 October; 99% upper HPD: 25 September) (Fig. 3H). If we again discount the reported evidence of COVID-19 in November and take 1 December as the first confirmed case, then the virus would spend less time in humans across both phases of the early epidemic (Fig. 3, I and J), and the index case in Hubei would have acquired SARS-CoV-2 on 18 November 2019 (95% upper HPD: 20 October; 99% upper HPD: 5 October) (Fig. 3K). As in the primary analysis, the inferred date of the index case in the two-phase epidemic was robust to varying model assumptions (fig. S10), including the amount by which viral fitness differed between the two phases (supplementary text and fig. S18).

These one-phase and two-phase epidemic simulations both suggest that the tMRCA is not representative of the emergence of SARS-CoV-2. In the primary analysis, the index case remains infected at the tMRCA (as in Fig. 2, upper left panel) in <1.1% of simulated epidemics (table S3). In the two-phase epidemics, the index case is still infected at the tMRCA in 2.2% of simulations, likely because of an increased variance in the time between index case and tMRCA. The initial, less-fit variant persisted until the tMRCA in 17% of simulated epidemics and until 1 January 2020 in 3.7% of simulated epidemics. However, when this less-fit variant persisted, it was represented by only a single infected individual at the tMRCA; this low frequency suggests that even if this less-fit variant did exist, it could have easily been missed in early genome sequencing efforts. In the two-phase epidemic,

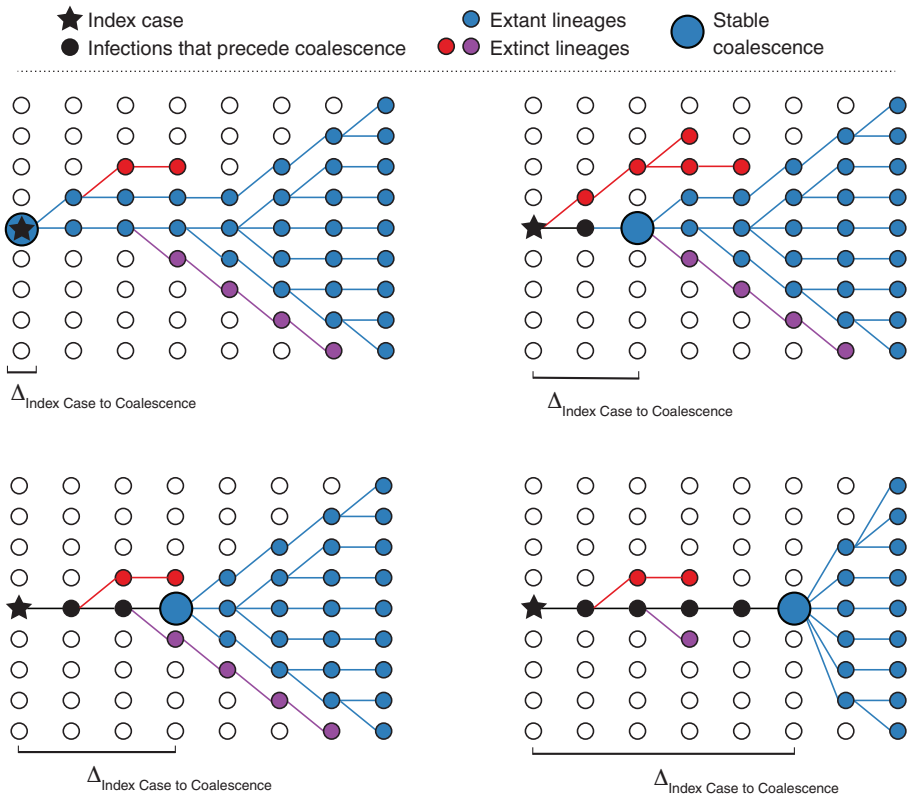


Fig. 2. Hypothetical coalescent scenarios depicting how the time between index case infection and time of stable coalescence can vary on the basis of stochastic extinction events of basal viral lineages. Coalescence can occur within or contemporaneously with the index case (upper left) or, in cases infected later in the course of the epidemic, with one (upper right) or more (lower left) basal lineages going extinct. In extreme cases, the epidemic can persist at low levels for a long time before stable coalescence (lower right).

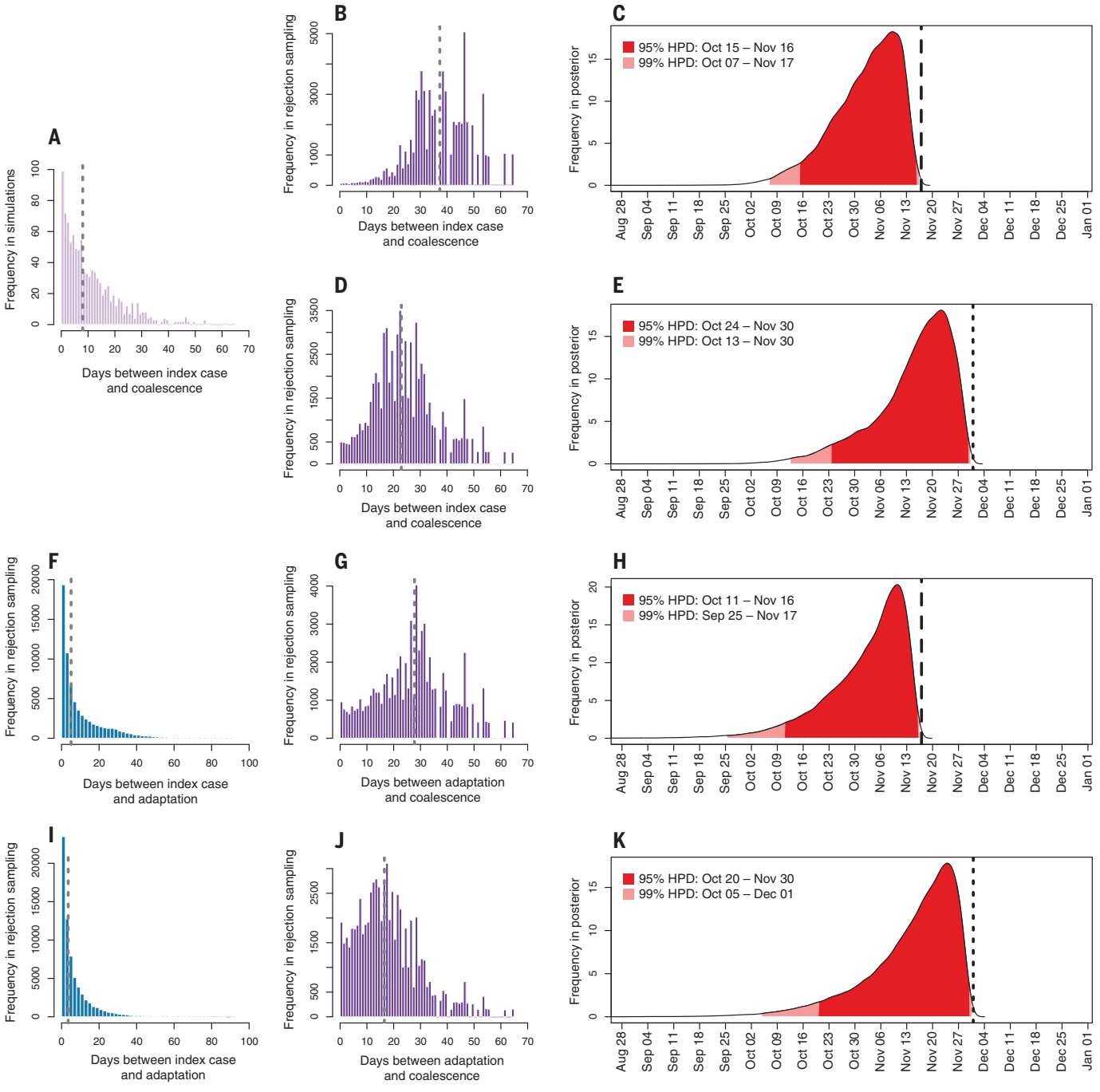


Fig. 3. Forward simulations estimating the timing of the index case in Hubei province. (**A**) Days between index case infection and stable coalescence in forward compartmental epidemic simulations (n = 1000). (**B** and **C**) Days between index case infection and stable coalescence after rejection sampling (B) and posterior distribution for date of index case infection (C), conditioned on an ascertained case by 17 November 2019, which is denoted by a long-dashed line. (**D** and **E**) Epidemic simulation, conditioned on an ascertained case by 1 December 2019, which is denoted by a short-dashed line. (**F** to **H**) Two-phase epidemic (F) days between index case carrying less-fit variant and adaptation (n = 2000), (G) days between adaptation and stable coalescence (n = 1000), and (H) posterior distribution for date of index case infection, conditioned on an ascertained case by 17 November 2019. (**I** to **K**) Two-phase epidemic conditioned on an ascertained case by 1 December 2019. Gray dashed lines indicate median estimates.

the first ascertained (or unascertained) case was due to the less-fit variant in around two-thirds of simulated epidemics.

By anchoring our epidemic simulations to specific tMRCA estimates, we can reconstruct a plausible range for the number of SARS-CoV-2 infections before the discovery of the virus (Fig. 4A). The median number of individuals infected with SARS-CoV-2 in our primary analysis is less than one until 4 November. The median number of infected individuals is four (95% HPD: 1 to 13) on 17 November and reaches nine (95% HPD: 2 to 26) on 1 December. These values are generally robust to model specifications, molecular clock method, and

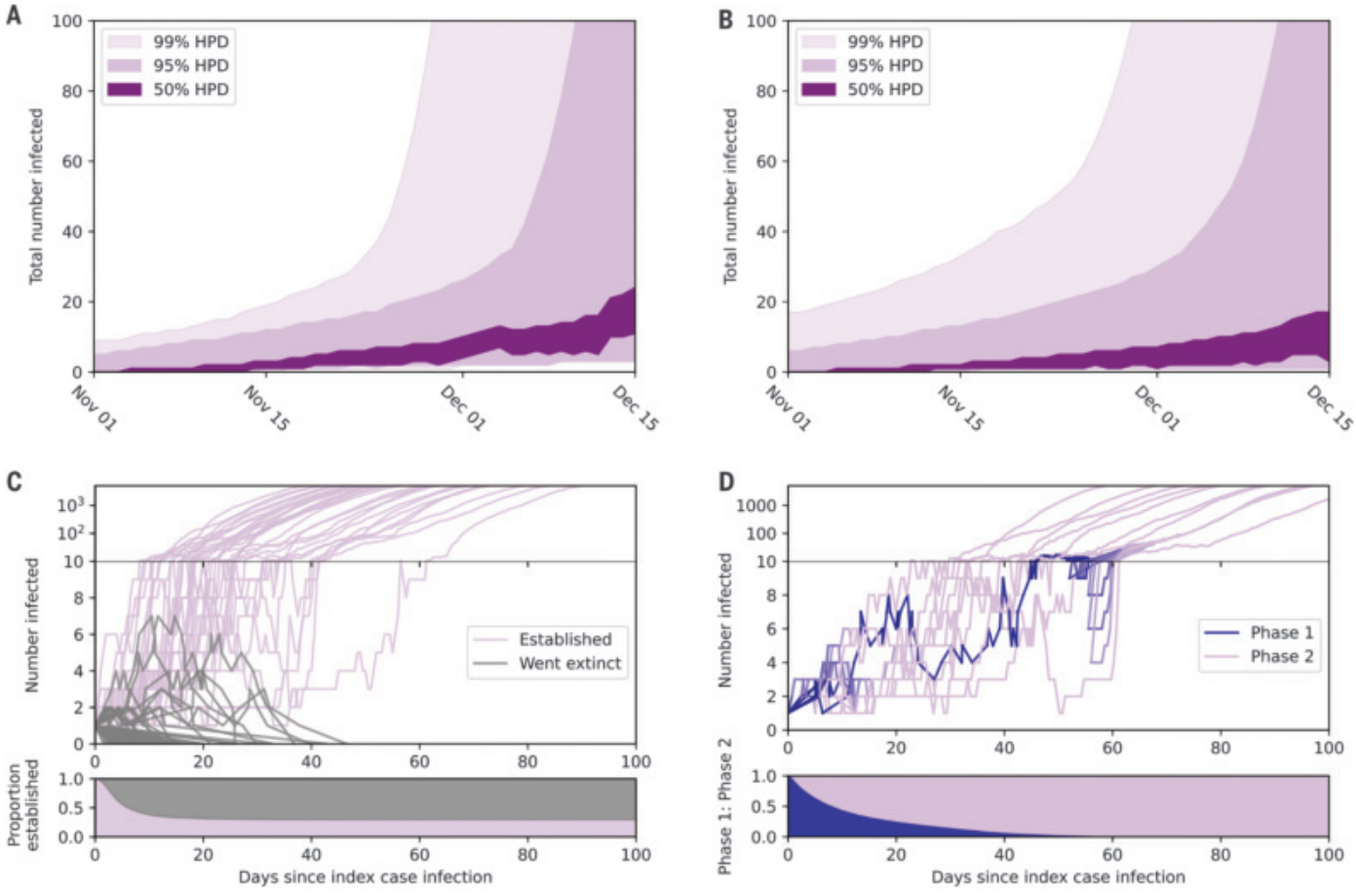


Fig. 4. Epidemic growth in compartmental simulations. (**A**) Estimated total number of people infected in late 2019. Dark purple shading represents central 50% HPD, intermediate purple shading represents central 95% HPD, and light purple represents central 99% HPD. (**B**) Estimated total number of people infected in late 2019 for a two-phase epidemic. (**C**) Number of people infected over time in a sample of epidemic simulations that established (purple; n = 30) and went extinct (gray; n = 70). The y axis transitions to log scale once ≥10 people are infected at any given time. The lower panel shows the proportion of simulations that still have at least 1 infected individual over time (persisting epidemics in purple; extinct epidemics in gray). (**D**) Sample (n = 10) of two-phase epidemic simulations transitioning from less-fit phase 1 (blue) to more-fit phase 2 (purple). Each line represents a single simulation and its transition over time. The lower panel shows the average proportion of phase 1– to phase 2–infected individuals over time.

date of first COVID-19 case (table S4 and fig. S11). The two-phase epidemics tend to exhibit similar growth patterns (Fig. 4B, fig. S12, and table S5). Notably, we do not see any evidence for an increase in hospitalizations until mid- to late December, even when we increase the virulence of the less-fit variant in the two-phase epidemics or increase the probability of hospitalization before the stable coalescence (supplementary text and figs. S13 and S14).

Empirical observation throughout the SARS-CoV-2 pandemic has shown the outsized role of superspreading events in the propagation of SARS-CoV-2 (*24*–*27*), wherein the average infected person does not transmit the virus. Our results suggest that the same dynamics likely influenced the initial establishment of SARS-CoV-2 in humans, as only 29.7% of simulated epidemics from the primary analysis went on to establish self-sustaining epidemics. The remaining 70.3% of epidemics went extinct (Fig. 4C). Simulated epidemics that went extinct typically produced only 1 infection (95% range: 1 to 9) and never more than 44 infections total or 14 infections at any given time (table S2). The median failed epidemic went extinct by day 8. As the contact network became more or less densely connected, the number of epidemics that went extinct was similar: 68.3 and 69.4%, respectively (table S2). However, the percentage of extinct epidemics increased as the transmission rate decreased (80.5%) and decreased as the transmission rate increased (53.6%). In the two-phase epidemic, this original less-fit variant went extinct by day 9 (95% HPD: 2 to 52) and produced a median of one infection (95% HPD: 1 to 13) (Fig. 4D and table S6).

The overdispersed nature of SARS-CoV-2 transmission patterns favors its persistence, as epidemics simulated over random contact networks (with the same mean number of contacts) that are not characterized by superspreading events (*27*) tended to go extinct more frequently, 83.7% of the time. Furthermore, the large and highly connected contact networks characterizing urban areas seem critical to the establishment of SARS-CoV-2. When we simulated epidemics in which the number of connections was reduced by 50 or 75% (without rescaling per-contact transmissibility) to reflect emergence in a rural community, the epidemics went extinct 94.5 or 99.6% of the time, respectively.

Our results highlight the unpredictable dynamics that characterized the earliest days of the COVID-19 pandemic. The successful establishment of SARS-CoV-2 postzoonosis was far from certain, as more than two-thirds of simulated epidemics quickly went extinct. It is highly probable that SARS-CoV-2 was circulating in Hubei province at low levels in November 2019 and possibly as early as October 2019, but not earlier. Nonetheless, the inferred prevalence of this virus was too low to permit its discovery and characterization for weeks or

months. By the time that COVID-19 was first identified, the virus had firmly established itself in Wuhan. This delay highlights the difficulty in surveillance for novel zoonotic pathogens with high transmissibility and moderate mortality rates.

The high extinction rates we inferred suggest that spillover of SARS-CoV-2–like viruses may be frequent, even if pandemics are rare (*28*). Furthermore, the same dynamics that characterized the establishment of SARS-CoV-2 in Hubei province may have played out all over the world, as the virus was repeatedly introduced but only occasionally took hold (*29*, *30*). The reports of cases in December 2019 and January 2020 in France and California that did not establish sustained transmission fit this pattern (*31*–*33*). However, our results suggest that polymerase chain reaction evidence of SARS-CoV-2 in wastewater outside of China before November 2019 is unlikely to be valid (*34*), and the suggestion of international spread in mid-November or early December 2019 should be viewed with skepticism (*35*–*37*), given that our results suggest that fewer than 20 people were infected with SARS-CoV-2 at this time (table S4 and fig. S11). Our results also refute claims (*38*) of large numbers of patients requiring hospitalization because of COVID-19 in Hubei province before December 2019 (figs. S13 and S14). Nevertheless, SARS-CoV-2 may be detectable in archived wastewater samples or other biomaterials from Hubei province from early to mid-November 2019, should they exist, and incorporating these types of data in our model could further refine our timing estimates. Moreover, wastewater detection may present the best chance of early detection of future pandemics during the early phase of spread for which we estimate very low numbers of infections (*39*).

Even though all of the earliest documented cases of COVID-19 were found in Hubei province, we cannot discount the possibility that the index case initially acquired the virus elsewhere. Nonetheless, our dating inference is insensitive to geography. Furthermore, our results suggest that if the virus first emerged in a rural community, it would have needed to migrate to an urban setting to avoid extinction. The lack of reports of COVID-19 elsewhere in China in November and early December suggests that Hubei province is the location where human-to-human transmission chains were first established.

The circumstances surrounding the emergence of SARS-CoV-2 in Hubei province remain shrouded. Although SARS-CoV-2 is repeatedly adapting to spread among humans (*40*, *41*), our findings do not reveal whether the virus that first emerged was less fit than the virus that spread throughout China. Nevertheless, the inferred timing of the index case is generally similar in both of these scenarios because less-fit viruses in our simulations that went extinct tended to do so very quickly. It is yet unknown whether the virus emerged directly from its animal reservoir [presumably horseshoe bats (*42*, *43*)] or first circulated in and possibly adapted to an intermediate host. Our estimates for the timing of the Hubei index case further distance this individual from the outbreak at the Huanan Seafood Wholesale Market. Finding the animal reservoir or hypothetical intermediate host will help to further narrow down the date, location, and circumstances of the original SARS-CoV-2 infection in humans. However, even in the absence of that information, coalescent-based approaches permit us to look back beyond the tMRCA and toward the earliest days of the COVID-19 pandemic. Although there was a pre-tMRCA fuse to the COVID-19 pandemic, it was almost certainly very short. This brief period of time suggests that future pandemics with similar characteristics to those of the COVID-19 pandemic permit only a narrow window for preemptive intervention.

REFERENCES AND NOTES

1. N. Zhu *et al*., *N. Engl. J. Med.* **382**, 727–733 (2020).
2. Q. Li *et al*., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
3. World Health Organization, "Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)" (World Health Organization, 2020); www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19).
4. H. Fei *et al*., *Lancet Reg. Health West. Pac.* **3**, 100032 (2020).
5. World Health Organization, "WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020" (World Health Organization, 2020); www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020.
6. C. Huang *et al*., *Lancet* **395**, 497–506 (2020).
7. X. Zhang *et al*., *Nature* **583**, 437–440 (2020).
8. F. Wu *et al*., *Nature* **579**, 265–269 (2020).
9. R. Lu *et al*., *Lancet* **395**, 565–574 (2020).
10. J. Ma, "Coronavirus: China's first confirmed Covid-19 case traced back to November 17," *South China Morning Post*, 13 March 2020; www.scmp.com/news/china/society/article/3074991/coronavirus-chinas-first-confirmed-covid-19-case-traced-back.
11. A. Rambaut, "Phylodynamic Analysis | 176 genomes | 6 Mar 2020," Virological (2020); https://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356.
12. S. Duchene *et al*., *Virus Evol.* **6**, veaa061 (2020).
13. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, *Nat. Med.* **26**, 450–452 (2020).
14. L. du Plessis, O. Pybus, "Further musings on the tMRCA," Virological (2020); https://virological.org/t/further-musings-on-the-tmrca/340.
15. N. R. Faria *et al*., *Science* **346**, 56–61 (2014).
16. M. Worobey *et al*., *Nature* **455**, 661–664 (2008).
17. B. F. Keele *et al*., *Science* **313**, 523–526 (2006).
18. T. Bedford *et al*., *Science* **370**, 571–575 (2020).
19. Q. Bi *et al*., *Lancet Infect. Dis.* **20**, 911–919 (2020).
20. M. A. Suchard *et al*., *Virus Evol.* **4**, vey016 (2018).
21. L. Pipes, H. Wang, J. P. Huelsenbeck, R. Nielsen, *Mol. Biol. Evol.* 10.1093/molbev/msaa316 (2020).
22. X. Hao *et al*., *Nature* **584**, 420–424 (2020).
23. N. Moshiri, M. Ragonnet-Cronin, J. O. Wertheim, S. Mirarab, *Bioinformatics* **35**, 1852–1861 (2019).
24. R. Laxminarayan *et al*., *Science* **370**, 691–697 (2020).
25. A. Endo, S. Abbott, A. J. Kucharski, S. Funk; Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, *Wellcome Open Res.* **5**, 67 (2020).
26. X.-K. Xu *et al*., *Clin. Infect. Dis.* **71**, 3163–3167 (2020).
27. D. C. Adam *et al*., *Nat. Med.* **26**, 1714–1719 (2020).
28. H. Li *et al*., *Biosaf. Health* **1**, 84–90 (2019).
29. M. Worobey *et al*., *Science* **370**, 564–570 (2020).
30. L. du Plessis *et al*., *Science* **371**, 708–712 (2021).
31. A. Deslandes *et al*., *Int. J. Antimicrob. Agents* **55**, 106006 (2020).
32. X. Deng *et al*., *Science* **369**, 582–587 (2020).
33. M. A. Jorden *et al*., *Morb. Mortal. Wkly. Rep.* **69**, 680–684 (2020).
34. G. Chavarria-Miró *et al*., medRxiv 2020.06.13.20129627 [Preprint]. 13 June 2020. https://doi.org/10.1101/2020.06.13.20129627.
35. G. Fongaro *et al*., *Sci. Total Environ.* **778**, 146198 (2021).
36. G. La Rosa *et al*., *Sci. Total Environ.* **750**, 141711 (2021).
37. A. Amendola *et al*., *Emerg. Infect. Dis.* **27**, 648–650 (2021).
38. E. O. Nsoesie, B. Rader, Y. L. Barnoon, L. Goodwin, J. Brownstein, "Analysis of hospital traffic and search engine data in Wuhan China indicates early disease activity in the Fall of 2019" (Harvard Library Office for Scholarly Communication, 2020); https://dash.harvard.edu/handle/1/42669767.
39. J. Peccia *et al*., *Nat. Biotechnol.* **38**, 1164–1167 (2020).
40. E. Volz *et al*., *Nature* 10.1038/s41586-021-03470-x (2020).
41. B. Korber *et al*., *Cell* **182**, 812–827.e19 (2020).
42. S. Lytras, J. Hughes, W. Xia, X. Jiang, D. L. Robertson, bioRxiv 2021.01.22.427830 [Preprint]. 30 January 2021. https://doi.org/10.1101/2021.01.22.427830
43. M. F. Boni *et al*., *Nat. Microbiol.* **5**, 1408–1417 (2020).
44. J. Pekar, J. Wertheim, Data for "Timing the SARS-CoV-2 Index Case in Hubei Province," Dryad (2021); https://doi.org/10.5061/dryad.4f4qrfjbm.

ACKNOWLEDGMENTS

We gratefully acknowledge the authors from the originating laboratories and the submitting laboratories, who generated and shared via GISAID the viral genomic sequence data on which this research is based. A complete list acknowledging the authors who submitted the data analyzed in this study can be found in data S1. We thank J. Havens, M. Sanderson, and B. Wertheim for fruitful conversations and insight. **Funding:** J.O.W. acknowledges funding from the National Institutes of Health (AI135992 and AI136056). J.P. acknowledges funding from the National Institutes of Health (T15LM011271) and the Google Cloud COVID-19 Research Credits Program. M.W. was supported by the David and Lucile Packard Foundation as well as the University of Arizona College of Science. N.M. acknowledges funding from the National Science Foundation (2028040) and the Google Cloud COVID-19 Research Credits Program. **Author contributions:** Conceptualization: J.O.W.; Methodology: J.P., N.M., M.W., K.S., and J.O.W.; Software: J.P., N.M., and J.O.W.; Validation: J.P. and J.O.W; Formal analysis: J.P. and J.O.W.; Investigation: J.P. and J.O.W.; Resources: J.O.W.; Data curation: J.O.W.; Writing – original draft: J.P. and J.O.W.; Writing – review & editing: J.P., M.W., K.S., J.O.W., and N.M.; Visualization: J.P. and J.O.W.; Supervision: J.O.W.; Project administration: J.O.W.; and Funding acquisition: J.O.W. and M.W. **Competing interests:** J.O.W. has received funding from Gilead Sciences, LLC (completed), and the CDC (ongoing) via grants and contracts to his institution unrelated to this research. **Data and materials availability:** All data used in this analysis are free to access via GISAID. BEAST XML input, FAVITES JSON, and results files are available in Dryad (*44*). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/412/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S7
References (*45*–*54*)
MDAR Reproducibility Checklist
Data S1

20 November 2020; accepted 15 March 2021
Published online 18 March 2021
10.1126/science.abf8003

GROUNDWATER

Global groundwater wells at risk of running dry

Scott Jasechko[1]* and Debra Perrone[2]

Groundwater wells supply water to billions of people, but they can run dry when water tables decline. Here, we analyzed construction records for ~39 million globally distributed wells. We show that 6 to 20% of wells are no more than 5 meters deeper than the water table, implying that millions of wells are at risk of running dry if groundwater levels decline by only a few meters. Further, newer wells are not being constructed deeper than older wells in some of the places experiencing significant groundwater level declines, suggesting that newer wells are at least as likely to run dry as older wells if groundwater levels continue to decline. Poor water quality in deep aquifers and the high costs of well construction limit the effectiveness of tapping deep groundwater to stave off the loss of access to water as wells run dry.

Groundwater provides nearly half of the water used for irrigated agriculture and supplies drinking water for billions of people (*1*, *2*). Its perennial and distributed nature makes groundwater suitable for providing reliable access to potable water and supporting efforts to adapt to climate change (*3*). Nevertheless, access to groundwater to meet water demands may be hindered when groundwater levels decline. Groundwater levels are declining in many aquifers around the globe [e.g., (*4*–*6*)], causing groundwater wells (*7*) to run dry (*8*–*10*) and leaving farmers unable to irrigate and households without access to reliable and convenient drinking water.

When groundwater wells run dry, one common adaptation strategy is to construct deeper wells (*11*). Deeper wells tend to be less vulnerable than shallower wells to climate variability (*12*) and groundwater level declines (*13*), but even relatively deep wells are not immune to long-term reductions in groundwater storage. Despite the role of wells as a basic infrastructure used to access groundwater, information about the locations and depths of wells has never before been compiled and analyzed at the global scale. Analyzing groundwater resources at the global scale is becoming increasingly important because of groundwater's role in virtual water trade, international policy, and sustainable development (*14*, *15*). Nevertheless, we emphasize that groundwater is a local resource influenced by hydrogeologic conditions, water policies, sociocultural preferences, and economic drivers.

Here, we compiled ~39 million records of groundwater well locations, depths, purposes, and construction dates (supplementary materials section S1), which provide local information at the global scale. These observations promote a better understanding of spatiotemporal patterns of well locations and depths (*16*, *17*). Groundwater data are notoriously difficult to collect and collate (*18*, *19*). The ~39 million wells are situated in 40 countries or territories that represent ~40% of global ice-free lands (average data density of 0.7 wells/km^2; Fig. 1). Half of all global groundwater pumping takes place within our study countries or territories [fig. S2, reference (*2*), and

[1]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106, USA. [2]Environmental Studies Program, University of California, Santa Barbara, CA 93106, USA.
*Corresponding author. Email: jasechko@ucsb.edu (S.J.)

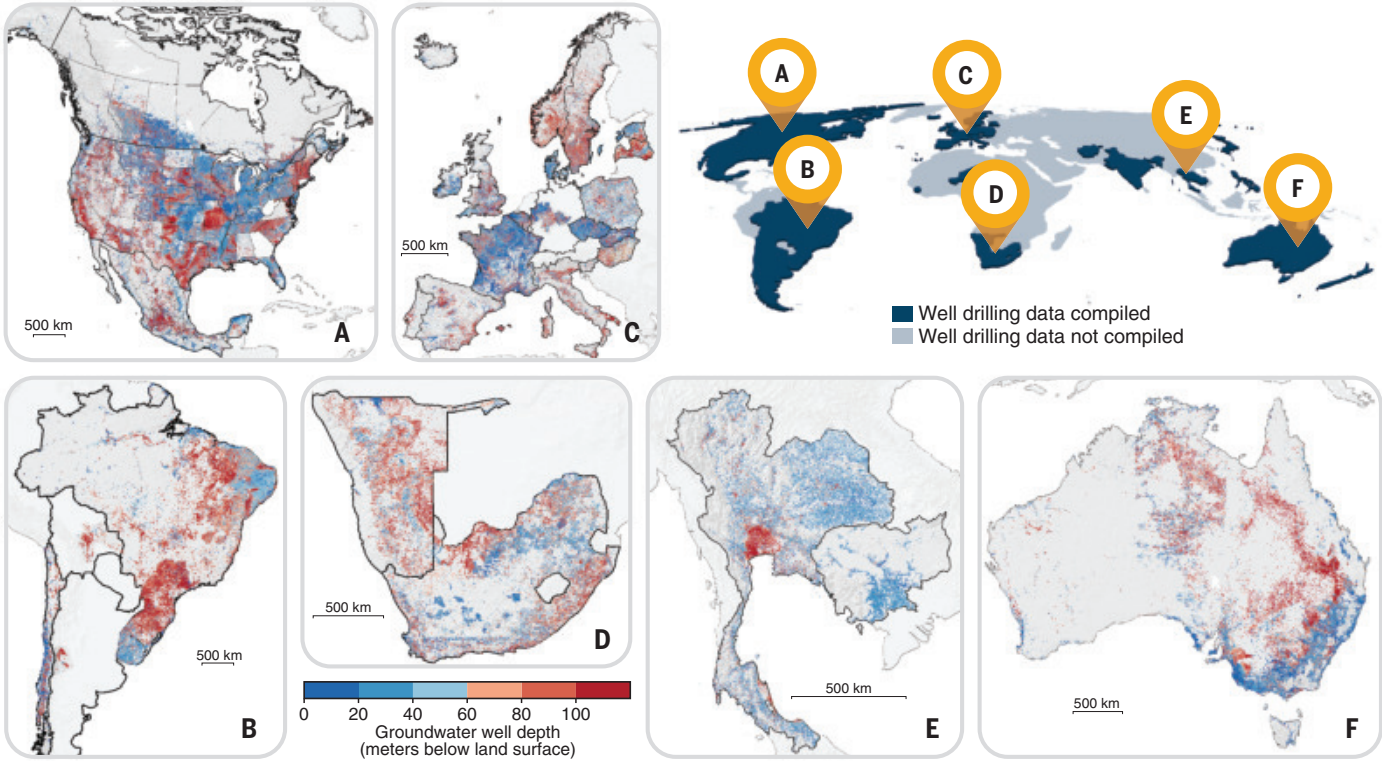


Fig. 1. Groundwater well depths in countries around the globe. The dataset contains ~39 million wells in 40 countries or territories (see inset map). Blue points mark shallower wells and red points mark deeper wells in Canada, the United States (white areas mark states where data are unavailable), and Mexico (**A**); Argentina (partial coverage), Bolivia (partial coverage), Brazil, Chile, and Uruguay (**B**); Iceland, Portugal, Spain, France, Germany (partial coverage), Denmark, Sweden, Norway, Italy, Slovakia, Slovenia, Belgium, Poland, Latvia, Estonia, Czech Republic, Hungary, United Kingdom, and Ireland (**C**); Namibia and South Africa (**D**); Thailand and Cambodia (**E**); and Australia (**F**). National-level analyses are available in figs. S37 to S435.

supplementary materials section S5.1], which are home to >3 billion people (table S1).

The compiled and quality-controlled well-depth data average 46 m and have a median depth of 31 m, a lower-upper quartile range of 16 to 62 m, and a 5th- to 95th-percentile range of 3.9 to 130 m (Fig. 1 and fig. S22); these global summary statistics obscure local variability. To better understand local context, we compared simulated water tables (*20*, *21*) with the depth of each well in our dataset. We show that between one-in-17 and one-in-five wells are no more than 5 m deeper than the water table. This comparison therefore suggests that ~6 to ~20% of wells would be shallower than the water table if it should decline by only a few meters (supplementary materials section S6 and fig. S36).

Wells with depths similar to that of the water table are particularly vulnerable to running dry because of (i) localized declines in the water table induced during pumping (i.e., drawdown), which can cause wells to run dry even if static water tables overlie the bottom of the well (*22*), and (ii) seasonal and interannual water table fluctuations, which can cause wells to temporarily run dry, rendering them unable to provide year-round access to water (*23*, *24*). By compiling and analyzing >100 million water level measurements made in >1 million individual monitoring wells (supplementary materials sections S2 and S6), we identified typical seasonal groundwater level fluctuations of at least ~1 m in half of these monitoring wells (figs. S33 and S34). Our findings suggest that some shallow wells may already run dry for parts of the year because of (i) a combination of drawdown and (ii) seasonal and interannual water table fluctuations.

Next, we analyzed how well construction depths vary over time. We identified locations where newer wells have been constructed deeper than older wells by grouping nearby wells and completing nonparametric regression analyses of well construction date versus well depth (supplementary materials section S7). We explored up to five different time intervals for each country, creating n = 117 individual country–timespan pairings (see points in Fig. 2). Nearly 75% of these pairings (n = 87) show that newer wells were being constructed deeper than older wells (Spearman rank coefficient $\rho > 0$) in the majority (>50%) of 100 km by 100 km areas within the country where sufficient data exist. We obtained similar results if we used two alternative spatial groupings: 10 km by 10 km study areas chosen to capture local contexts of groundwater access and irregular polygons designed to maintain a consistent sample size of ~100 wells (supplementary materials section S3). The similarity of these results demonstrates the robustness of our finding: Newer wells tend to be deeper than older wells in most, but not all, areas.

Although constructing a new, deeper well may provide short-term resilience to declining water levels, it can also increase the operational and capital costs of using groundwater (*25*, *26*), raising equity and adaptation concerns over the long term (*27*). Deeper wells often have deeper water levels and therefore require more energy to lift water from the well to the surface (*17*). High costs associated with well construction already limit some users from accessing groundwater in areas where groundwater is being depleted; that is, groundwater depletion can exclude neighboring groundwater pumpers, allowing well owners with capital to construct deeper wells and outcompete others in the area (*10*, *25*, *28*, *29*). In Punjab, India, well depths correlate with property size and crop production (*28*), highlighting that owners of large and lucrative farms are more likely than owners of smaller farms to be able to afford deep wells. Further, constructing deep wells is only feasible where fresh groundwater is stored in deep, permeable aquifers. In places where deep aquifers contain saline water or yield little water when pumped (*30*), constructing deep wells cannot

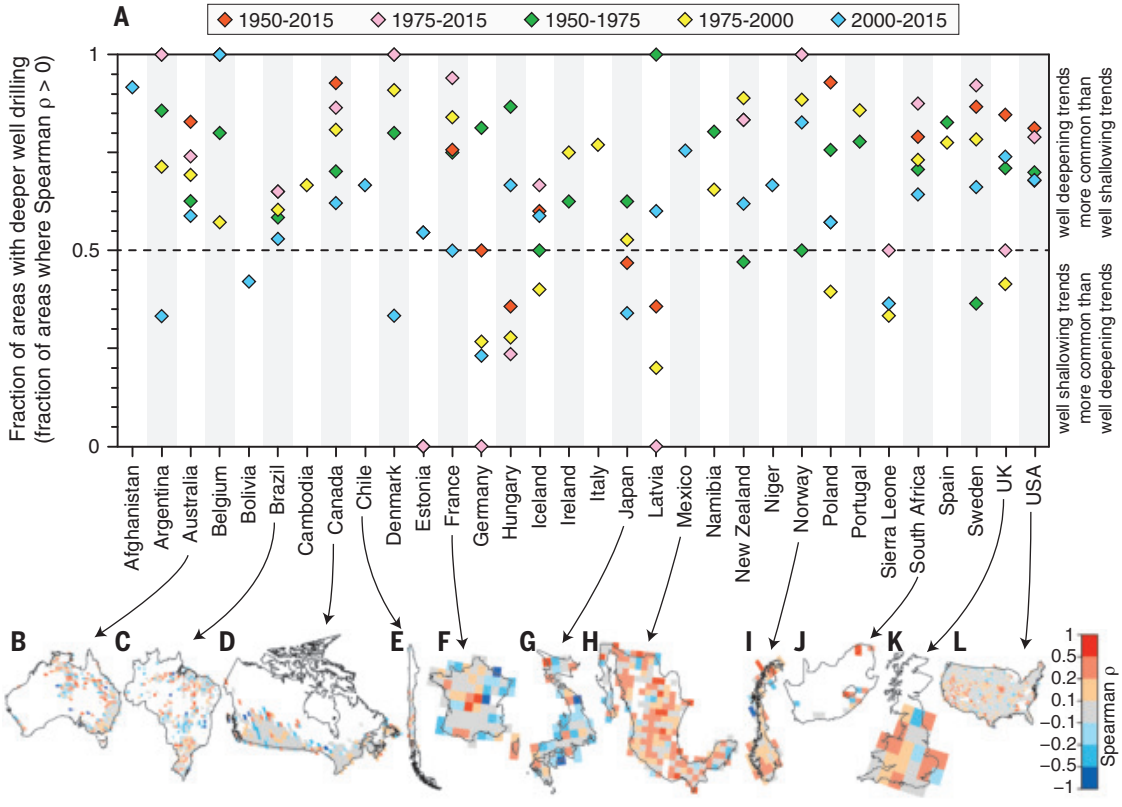


Fig. 2. Fraction of areas with deeper well drilling versus shallower well drilling for five time periods among study countries. Well-deepening trends are more common than well-shallowing trends for the majority of time periods for the majority of countries (national-level analyses in figs. S37 to S435). (**A**) The fraction of 100 km by 100 km study areas where wells are being drilled deeper over time (Spearman $\rho > 0$; regression of well completion depth versus well construction date). Each diamond represents well construction depth variations over time for a given country (e.g., blue diamonds are regressions for years 2000–2015; see legend above figure); some points overlap. (**B** to **L**) Spearman ρ determined by regression of well depth versus construction date for the time interval 2000–2015 in 11 countries [i.e., the fraction of all 100 km by 100 km areas in these maps with Spearman $\rho > 0$ correspond to the blue diamonds in (A)]. Blue shades mark 100 km by 100 km areas where wells are being constructed shallower over time (i.e., Spearman $\rho < -0.1$); orange and red shades mark 100 km by 100 km areas where wells are being constructed deeper over time (i.e., Spearman $\rho > 0.1$).

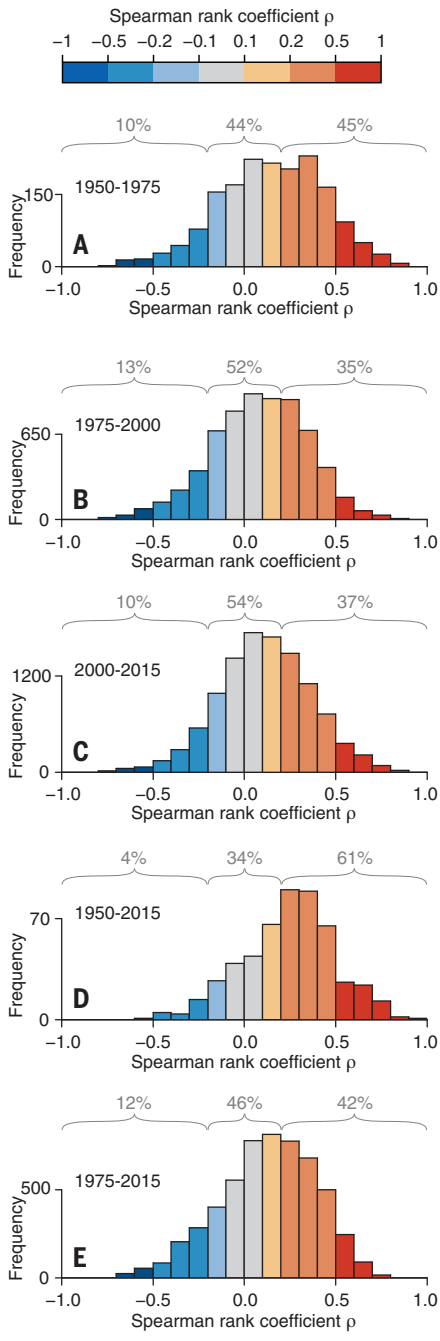


Fig. 3. Well construction depth trends at locations close to (<5 km) a monitoring well where groundwater levels have declined (Spearman ρ > 0) significantly (Spearman *P* < 0.05) over time. Results derive disproportionately from Australia, Canada, France, Mexico, New Zealand, Sweden, and the United States (tables S64 to S69). The histograms are distributions of Spearman rank coefficients as determined through regression analyses of well completion depths versus well completion dates at sites close to a monitoring well where water levels have declined significantly over time (Spearman ρ > 0 and Spearman *P* < 0.05 determined through regression analyses of measured depth to water in monitoring well versus water level measurement date). Each plot in (**A**) to (**E**) represents one study time interval [e.g., (A) presents results for 1950–1975]. Although all distributions skew right (i.e., well deepening where groundwater levels are declining), a substantial number of sites indicate that newer wells are not being constructed substantially deeper than older wells where groundwater levels are declining. Curly brackets above each plot in (A) to (E) provide the percentage of well construction depth trends falling within the following Spearman ρ ranges: −1.0 to −0.2, −0.2 to 0.2, and 0.2 to 1.0.

alleviate water scarcity; for example, some deep aquifers contain brackish or saline groundwater (*31*) (fig. S30).

We found that newer wells were being constructed deeper than older wells in many, but not all, areas (Fig. 2), so it is unclear whether deeper well construction was widespread where groundwater levels were declining. By analyzing the aforementioned compiled monitoring well data (supplementary materials section S2), we identified monitoring wells that recorded a significant (Spearman *P* value < 0.05) groundwater level decline. Next, we analyzed temporal variations in the depths of new wells constructed near (<5 km) each monitoring well where a significant groundwater level decline had been observed and sufficient well construction data were available (see the materials and methods) in the same five time periods noted in Fig. 2. We show that newer wells were being constructed deeper than older wells (Spearman ρ > 0) in 64 to 83% of the areas surrounding a monitoring well where a significant groundwater level decline had been observed (table S69). Using a more stringent threshold (Spearman ρ > 0.2), we found that newer wells were being constructed deeper than older wells in 35 to 61% of the areas surrounding a monitoring well where a significant groundwater level decline had been observed (Fig. 3). We conclude that well construction was targeting similar depths or shallower depths over time in many of the places experiencing long-term groundwater level declines (Fig. 3; supplementary materials section S4), implying that newer wells were at least as vulnerable as older wells to declining groundwater levels in these areas (Fig. 3).

Many countries lack long-term groundwater monitoring programs or do not publish such data online; therefore, our comparison of monitoring well water level time series and well construction activities derives mostly from Australia, Canada, France, Mexico, New Zealand, Sweden, and the United States (tables S64 to S69). To go beyond these countries, we compared well depth data with a global map of 2002 to 2016 mean rates of change ("trends") in terrestrial water storage (*5*) based on gravity field observations from the Gravity Recovery and Climate Experiment (GRACE) satellites. During its 15-year mission, GRACE observed monthly changes in total terrestrial water storage (including groundwater) at a coarse spatial resolution [~200,000 km^2 (*32*)] with global coverage. Our analysis does not reveal a close spatial relationship between declining terrestrial water storage and deeper well construction during 2002 to 2016 (figs. S12 to S16), further supporting our finding that wells are not always being constructed deeper where terrestrial water stores are declining.

Where water levels are declining and wells are not being constructed deeper, newer wells are at least as likely to run dry as older wells, should groundwater levels continue to decline. Declining groundwater levels can result from overpumping (*4*, *6*), reductions in recharge driven by drought or longer-term climate change (*3*, *33*), or a combination of overpumping and reductions in recharge. Regardless of the drivers behind declining groundwater levels, the ramification of these declines is clear: Groundwater levels that decline below well bottoms render these wells unable to access groundwater (Fig. 3).

From India to the United States, wells are already running dry because of groundwater level declines (*22*, *26*, *34*, *35*). When a well runs dry, there are a number of ways to adapt: (i) construct a new and deeper well or deepen the existing well that ran dry (*10*, *13*); (ii) access alternative water sources (*10*, *29*); (iii) conserve water through behavioral changes or technological adoptions [e.g., switching to crops that are less water intensive; adopting irrigation approaches that reduce water losses (*25*)]; (iv) sell the property (*10*); or (v) abandon activities that require high water inputs, such as irrigation (*35*). In Karnataka, India, some households that have had their well run dry

have offset lost agricultural profits by seeking off-farm income (*35*). The transition away from agriculture can have longer-term impacts on the household, such as debt accrual while seeking alternative employment and lost educational opportunities for children who join the workforce early to supplement household income (*35*). In California's Central Valley (United States) and several other agricultural hubs around the globe, typical agricultural wells are deeper than domestic wells (supplementary materials sections S5.5 to S5.6); as a result, rural domestic wells are running dry disproportionately relative to agricultural wells (*9*), even though domestic wells tend to withdraw less water. Unlike urban areas, many rural communities have limited economic, technical, and political capacity to diversify water sources or construct deeper wells, even if fresh groundwater is available at deeper depths.

Groundwater wells supply water to billions of people around the globe (*2*, *13*, *14*, *36*). Groundwater depletion is projected to continue in some areas where it is already occurring, and even expand to new areas not yet experiencing depletion (*37*). Despite this, not all of our study countries have long-term groundwater monitoring networks. There is an opportunity to establish monitoring networks and make the collected data publicly available and shareable to evaluate relationships between groundwater well construction and depletion in the future. Furthermore, identifying areas where constructing deeper wells may remain feasible in the short term will require expanding existing global databases of subsurface permeability and groundwater quality [e.g., salinity (*31*)].

Our work highlights the vulnerability of existing wells to groundwater depletion because (i) many wells are not much deeper than the local water table, making them likely to run dry with even modest groundwater level declines (supplementary materials section S6), and (ii) deeper well construction is common but not ubiquitous where groundwater levels are declining (Fig. 3 and supplementary materials section S4). Incorporating well depth data into management and sustainable development plans will be crucial for projecting limitations to groundwater use imposed by socioeconomics and heterogeneities in aquifer systems.

REFERENCES AND NOTES

1. S. Siebert *et al*., *Hydrol. Earth Syst. Sci.* **14**, 1863–1880 (2010).
2. J. Margat, J. van der Gun, *Groundwater Around the World: A Geographic Synopsis* (CRC, 2013).
3. R. G. Taylor *et al*., *Nat. Clim. Chang.* **3**, 322–329 (2013).
4. L. F. Konikow, E. Kendy, *Hydrogeol. J.* **13**, 317–320 (2005).
5. M. Rodell *et al*., *Nature* **557**, 651–659 (2018).
6. T. Gleeson, Y. Wada, M. F. P. Bierkens, L. P. H. van Beek, *Nature* **488**, 197–200 (2012).
7. For simplicity, we define a "well" as a vertical hole constructed to access groundwater, including holes that are hand dug or drilled with machinery, where water is accessed with or without a pump (terminology varies globally).
8. D. Perrone, S. Jasechko, *Environ. Res. Lett.* **12**, 104002 (2017).
9. S. Jasechko, D. Perrone, *Earths Futur.* **8**, e2019EF001339 (2020).
10. M. Murti *et al*., *BMC Public Health* **16**, 430 (2016).
11. F. Molle, E. Lopez-Gunn, F. van Steenbergen, *Water Altern.* **11**, 445–457 (2018).
12. D. J. MacAllister, A. M. MacDonald, S. Kebede, S. Godfrey, R. Calow, *Nat. Commun.* **11**, 1099 (2020).
13. J. S. Famiglietti, *Nat. Clim. Chang.* **4**, 945–948 (2014).
14. S. Foster, J. Chilton, G. J. Nijsten, A. Richts, *Curr. Opin. Environ. Sustain.* **5**, 685–695 (2013).
15. T. Gleeson, M. Cuthbert, G. Ferguson, D. Perrone, *Annu. Rev. Earth Planet. Sci.* **48**, 431–463 (2020).
16. The depth below the land surface to a pump used to access water in a well is not documented widely, so well depths are used as a proxy for the depth that groundwater can be drawn from a well. The depth to a well bottom is deeper than the depth to a pump, providing a maximum depth to which groundwater can be accessed.
17. D. Perrone, S. Jasechko, *Nat. Sustain.* **2**, 773–782 (2019).
18. W. M. Alley, L. Beutler, M. E. Campana, S. B. Megdal, J. C. Tracy, *Ground Water* **54**, 758–761 (2016).
19. M. Moench, "When the well runs dry but livelihood continues: Adaptive responses to groundwater depletion and strategies for mitigating the associated impacts," in *The Agricultural Groundwater Revolution: Opportunities and Threats to Development* (CABI, Comprehensive Assessment of Water Management in Agriculture Series, 2007).
20. Y. Fan, H. Li, G. Miguez-Macho, *Science* **339**, 940–943 (2013).
21. Y. Zeng *et al*., *J. Adv. Model. Earth Syst.* **10**, 1882–1900 (2018).
22. R. A. Pauloo *et al*., *Environ. Res. Lett.* **15**, 044010 (2020).
23. S. D. Limaye, *Hydrogeol. J.* **18**, 543–558 (2010).
24. R. Johnston, M. Roberts, T. Try, S. J. de Silva, "Groundwater for irrigation in Cambodia," in *IWMI - ACIAR Investing in Water Management to Improve Productivity of Rice-Based Farming Systems in Cambodia Project* (International Water Management Institute, Issue Brief #3, 2013), p. 12.
25. S. S. Sayre, V. Taraz, *J. Environ. Econ. Manage.* **93**, 85–100 (2019).
26. C. A. Scott, T. Shah, *Int. J. Water Resour. Dev.* **20**, 149–164 (2004).
27. K. E. Kemper, "Instruments and institutions for groundwater management" in *The Agricultural Groundwater Revolution: Opportunities and Threats to Development* (CABI, Comprehensive Assessment of Water Management in Agriculture Series, 2007).
28. A. Sarkar, *Econ. Polit. Wkly.* **46**, 59–66 (2011).
29. T. Shah, *Groundwater Markets and Irrigation Development: Political Economy and Practical Policy* (Oxford Univ. Press, 1993).
30. T. Foster, N. Brozović, C. Speir, *J. Hydrol. (Amst.)* **547**, 638–649 (2017).
31. J. Thorslund, M. T. H. van Vliet, *Sci. Data* **7**, 231 (2020).
32. J. S. Famiglietti, M. Rodell, *Science* **340**, 1300–1301 (2013).
33. B. F. Thomas, J. S. Famiglietti, *Sci. Rep.* **9**, 4124 (2019).
34. T. Hora, V. Srinivasan, N. B. Basu, *Geophys. Res. Lett.* **46**, 9602–9611 (2019).
35. D. Blakeslee, R. Fishman, V. Srinivasan, *Am. Econ. Rev.* **110**, 200–224 (2020).
36. A. Mukherji, T. Shah, *Hydrogeol. J.* **13**, 328–345 (2005).
37. S. W. D. Turner, M. Hejazi, C. Yonkofski, S. H. Kim, P. Kyle, *Earths Future* **7**, 123–135 (2019).

ACKNOWLEDGMENTS

We thank agencies for providing well construction data for analyses and for their assistance controlling the datasets for quality; C. Madden and A. Leiby for assistance compiling monitoring well data; and J. W. Kirchner for comments that improved our manuscript. **Funding:** This research was completed without external funding. **Author contributions:** D.P. and S.J. conceived and designed the work. S.J. oversaw, compiled, and quality controlled the well completion and monitoring well data. S.J. completed the geospatial analyses. S.J. and D.P. wrote the manuscript. S.J. wrote the supplementary materials. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Groundwater well data that support our analyses are available through web pages and personal communication with national and regional-scale agencies; some agencies require consent to share well data, some prefer that requests go through their agency, and other agencies require records requests. We direct other researchers interested in accessing well data to the supplementary materials, where we document websites for directly downloading databases and provide contact information for agencies to request access to data.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6540/418/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S435
Tables S1 to S107
References

15 April 2020; accepted 26 February 2021
10.1126/science.abc2755

Microscope 2.0: Artificial intelligence (AI)-powered microscopes are coming to a lab near you.

Small images, big picture:
Artificial intelligence to revolutionize microscopy

Researchers hope that bringing deep learning techniques to cell imaging and analysis could turn messy biological problems into solvable computations.

By Sarah O'Meara

It was 20 years ago– while completing her Ph.D.–that computational biologist Anne Carpenter first realized she needed to learn computer programming.

Carpenter, who runs a laboratory at the **Broad Institute of MIT** and Harvard in Cambridge, Massachusetts, says she remembers facing either three months of manual image analysis, or the choice to enable her microscope to run by itself. She chose the latter. This automated approach has since revealed its potential to solve–or at least begin to tackle–some of the problems that limit scientists who use microscopic techniques to manually observe the workings of cells. For example, automation can reduce the time-consuming, meticulous work of identifying changes in cells, known as cellular morphology.

Carpenter's lab focuses on accelerating drug discovery using software to analyze the cellular morphology data contained in millions of images. "There are so many bottlenecks in the drug discovery pipeline, and the data from these images is proving useful for each of them: from building better disease-relevant assays and better screening libraries, to predicting assay outcomes and toxicity," she says.

Managing the limitations

Rebecca Richards-Kortum, a professor of bioengineering at **Rice University** in Houston, Texas, is currently collaborating with the MD Anderson Cancer Center on addressing some of the fundamental limitations of traditional microscopy. When using a conventional microscope, there is a fixed trade-off between depth-of-field (DOF) and spatial resolution: the higher the desired spatial resolution, the narrower the DOF. Working with Ashok Veeraraghavan at Rice and Ann Gillenwater at MD Anderson, the team has developed a computational microscope called a DeepDOF, which can achieve DOF greater than five times that of conventional microscopes while retaining its resolution, significantly reducing the time needed for image processing.

DeepDOF uses an optimized phase mask placed at the microscope aperture and a deep-learning-based algorithm that turns sensor data into high-resolution, large-DOF images, explains Richards-Kortum.

"Because of its low cost, high speed, and automated analysis capabilities, we hope the DeepDOF scope can greatly expand the number of surgical centers that have the capability to accurately assess oral cancer tumor margins at the time of surgery," she says. "The ability to accurately assess diseased tissue could help optimize the outcome of surgical removal, especially in resource-constrained settings, such as rural areas."

cont.>

IMAGE: © LEONID STUDIO\SHUTTERSTOCK.COM

Upcoming features

- Genetics—September 24

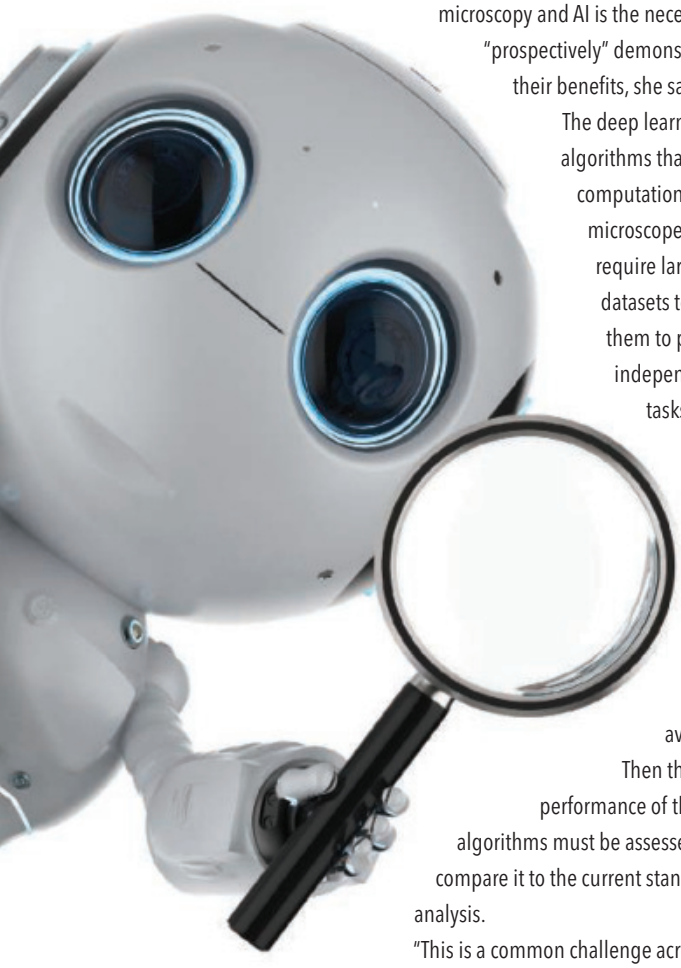


IMAGE: © PHONLAMAI PHOTO\SHUTTERSTOCK.COM

One of the biggest challenges to developing her innovative health care techniques using microscopy and AI is the necessity of "prospectively" demonstrating their benefits, she says. The deep learning algorithms that power computational microscopes require large datasets to train them to perform independent tasks, but such datasets are not always readily available. Then the performance of these algorithms must be assessed to compare it to the current standard of analysis.

"This is a common challenge across the health care technology community," she says.

Working through the challenges

Ricardo Henriques runs the **Optical Cell Biology Laboratory** at the **Instituto Gulbenkian de Ciência**, Portugal. His cross-disciplinary team of optical physicists, computer scientists, and biomedical researchers works on improving the limits of current imaging technology. The team is focused on two key challenges: how to analyze the real-time behavior of viruses infecting living cells and how to establish intelligent microscopy technology that reduces the damage caused by light to biological systems during their observation, known as phototoxicity.

To imagine these ideas working together, he recommends likening cells to football players.

"So you want to film a football match, but there's something about the camera that is toxic to the players," says Henriques. "To reduce the risk to them, you must minimize the amount of time you're filming, but you also need to make good decisions about what key moments to capture to truly understand the game."

Henriques' team is working on developing machine learning algorithms that can make better predictions of when key events will happen in cells as a viral infection progresses, and capture those moments. Simultaneously, these algorithms will attempt to reduce the amount of time spent capturing nonrelevant changes and alleviate the exposure that the cells have in these toxic environments.

For Henriques, it's been important to build a cross-disciplinary team to tackle these ideas, as the work involves multiple scientific skill sets.

"There needs to be a mentality shift to fully bring AI into scientific research across the board," he says.

Many of the disciplines that are inherent to microscopy, such as physics and biology, have a natural tendency to work separately due to linguistic barriers between the fields and how funding is regularly organized to target discrete areas, rather than collaborative projects.

"I think that organizations are slowly investing in building these bridges, but more needs to be done to encourage this," says Henriques.

Building bridges

Geoscientist Matt Andrew works at optics technology company **ZEISS** in Dublin, California, and his research focuses on the flow and transport processes in porous and sedimentary rocks. Andrew's work has increasingly centered on developing technologies to make better use of the data produced by microscopes, he says. He now works with a variety of teams across the company to help colleagues to bring AI into their research practice.

He says that the key to introducing AI into the daily practice of microscopy, whether you're looking at cells or rocks, is to ensure that the technology can be used by any scientist, regardless of their knowledge of deep learning techniques.

"Building workflows that unlock the potential and power of deep learning, and that work quickly and are easy to use, has been critical to their adoption," Andrew says.

For example, Andrew and his team use a process called the Solutions Lab to build workflows that use AI to automatically detect sample regions scientists may wish to investigate. "You can use AI to identify regions that correspond to individual features you then want to image at a much higher resolution," he says, adding, "AI Technologies commonly use open source libraries and shared components; our technologies are so successful because we ensure they are a lot more straightforward to use and comes in a package that's easier to digest."

Andrew believes that we're at the beginning of a revolution in the use and implementation of microscopy data.

"If I think back to 5 years ago, we didn't have a clue that we could use these kinds of techniques for microscopy," he says. "Now, we're moving to a point where we are going to have these algorithms sitting at the heart of every single portion of all of our workplace."

Luciano Guerreiro Lucas, a director at **Leica Microsystems**, headquartered in Wetzlar, Germany, is also focused on creating intelligent software solutions that can solve some of the biggest problems the life sciences and biopharma communities face when it comes to image data. Over the past 4½ years, his team has been building a library of pre-trained deep learning models as well as Aivia software that allow anyone to leverage some key AI microscopy technology.

cont.>

Featured participants

Carpenter Lab, Broad Institute
carpenterlab.broadinstitute.org

Gladstone Institutes
gladstone.org

Leica Microsystems
www.leica-microsystems.com

Nikon Instruments
www.microscope.healthcare.nikon.com

Optical Cell Biology Laboratory, Instituto Gulbenkian de Ciência
gulbenkian.pt/ciencia/research-groups/rhenriques

Rice University
www.rice.edu

ZEISS
www.zeiss.com

"Present day tools ignore the fact that researchers may be experts in biology, or a similar discipline, but have very limited expertise in microscopy or image analysis," says Lucas. "We are creating tools that leverage the biologist's expertise and learn from them. Such tools should gradually learn what a cell is and what it can look like in multiple scenarios and ultimately do the imaging and image analysis autonomously, allowing the researcher to focus on the creative and critical thinking portion of the scientific discovery process."

Lucas says that the current challenge of fulfilling this idea is a limited supply of high-quality, structured data, and a lack of standard image formats.

"These problems make it hard to progress faster in our field. Moreover, the data that does exist tends to be kept in isolated silos. It is hard to coordinate broad agreement around file and data standards in the research space. Researchers all like to do things their own way."

The commercial and academic sectors need to invest more time in educating the community about the benefits of solving these issues, he says.

Data in action

Steven Finkbeiner, a director and senior investigator at the **Gladstone Institutes** in San Francisco, California, has been at the forefront of AI and microscopy research for the past decade. Since inventing a fully automated robotic microscope that can track cells for months at a time, he and his team have generated extraordinarily large amounts of data. This information has given his team the ability to truly explore the potential of AI.

"We have shamelessly been using the petabytes of data we generate," he says.

For example, his team is using facial recognition AI technology–treating a cell's morphology as a face–to identify and track individual cells in complex systems, such as tissues, over time.

"We expect these approaches will open up new possibilities to study processes that involve complex cell–cell interactions such as neuroinflammation," he says.

Finkbeiner is also teaching learning networks to diagnose neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and Parkinson's disease, by showing the neural networks examples of cell images from patients.

"We ask the network in a relatively unbiased way to see if it can find anything in the image that would enable it to make an accurate diagnosis, and we have very encouraging results," he says. "We are hopeful that this could lead to new ways to stratify patients, discover biomarkers, and develop effective individualized therapies. It may even make it possible to diagnose a risk for a disease before symptoms start, which would be transformative."

His team is also using AI to predict the future fate of cells. "To do this, we are leveraging our longitudinal single-cell data and the use of deep learning networks to look for features of a cell at an early time point that would predict its fate. We are using this now with a cancer project that will help us understand why some cells develop drug resistance and some do not," says Finkbeiner.

Bringing AI into labs

Rich Gruskin is senior general manager for software systems at **Nikon Instruments**, headquartered in Melville, New York. He works closely with customers to ensure that researchers are easily able to adopt AI technology.

In a recent case, the customer was looking to identify multiple cell types in their label-free (brightfield) image data. As they were low-contrast images that differed only by sometimes subtle morphological features, several AI networks were trained to work together in one analysis assay to distinguish the different cell types.

"We trained a neural network using the customer's datasets. We ran it and it works very well," he says. "Sometimes if there's resistance to trying something new, we jump in and help clients process the information, build new routines, and show them how it works. Ensuring that the application is pain-free and the results are quickly obtainable is key to building confidence in using new technology."

The future of AI

Across the board, researchers working in academic and commercial settings see the greatest barrier to adopting AI into scientific life as fear of the unknown. Yet its increasing influence is undeniable.

"Change is happening in months, not years," says Henriques. "Look at self-driving cars. They are doing exactly what we want to do with microscopes, observing their environment in real time and making decisions on how to interact with it and about how to keep an organism alive–the person who is in the car."

However, there's also a sense that while change is inevitable, society needs to make a stronger commitment to ensuring all scientists can benefit from these new technologies. Finkbeiner thinks it would be hugely helpful to the field to create some public image repositories that computer scientists could use to develop new algorithms and approaches.

"Kids in college and even in high school could use data like this for education and training as well," says Finkbeiner. "The potential of this field is huge, so it would be great to invest now to train the generation we need to really take us forward."

He also would like to see academic institutions place greater emphasis on fostering collaborations between biologists and computer scientists.

"Universities that have computer science departments and who recognize the value of multidisciplinary science have an opportunity to lead," says Finkbeiner. "The gap between computer science and biology is large, and we need a sustained effort and sustained support to get the spark to ignite."

Sarah O'Meara is a freelance writer who specializes in life sciences.

Compact Microscope

ZEISS has introduced a compact microscope for digital teaching and routine lab work. ZEISS Primostar 3 is a robust upright light microscope developed for daily work in a classroom or in a lab for tissue and sample examination in histology, cell biology, food, or microbiology, etc. It is designed for long-term use and extreme durability. The unit is simple to operate, so that students and laboratory staff can spend more of their time exploring rather than fiddling with knobs. It is easy to learn, and the setup is quick and convenient with a plug-and-play installation. Users can choose either a 30-W halogen bulb or an energy-saving LED bulb with a stable color temperature and illumination intensity. Or, they can add on the fluorescence tube and turn ZEISS Primostar 3 into an LED fluorescence microscope. ZEISS Primostar 3 offers the advantages of an integrated 8-megapixel microscope camera. With the ZEISS Labscope imaging app, it is easy to connect microscopes in classrooms to each other.

ZEISS

For info: +49-(0)-7364-20-3800
www.zeiss.com

3D Cell Analysis Software

Used with Olympus confocal laser scanning microscopes, such as the FLUOVIEW FV3000 system, NoviSight 3D cell analysis software provides images of cell clusters down to the nuclei. The software's True 3D technology uses multiple microplate images to provide accurate morphology data and the ability to quantitatively analyze the effect of medications, including growth suppression and cell-survival rates. A range of parameters can be easily and precisely measured, enabling researchers to count the number of cells that have suppressed growth, proliferated, or died. NoviSight 3D also makes it easy to compare the effects of different medications at various concentrations. Results—including recognition, analysis, and statistics—are displayed on a single screen. Users can view the quantitative data as a scatterplot, heat map, or graph, and click on a point in the graphical display to automatically open the corresponding image. It is easy to switch between 2D and 3D views of the sample, and data are easily exported as CSV or FCS files for further analysis.

Olympus

For info: 301-325-3713
www.olympus-lifescience.com/en

Fixable Mitochondrial Stain

MitoView Fix 640 is a new far-red fluorescent live-cell mitochondrial stain that retains crisp and specific staining after fixation for downstream immunofluorescence protocols. The nontoxic, cell membrane–permeant dye rapidly accumulates in the mitochondria of live cells and has a reactive group that covalently attaches to proteins inside the mitochondria. This results in stable mitochondria staining that lasts for days in live cultures, is well retained after fixation with formaldehyde or methanol, and withstands permeabilization. Staining with MitoView Fix 640 shows better specificity after fixation as compared to MitoTracker Deep Red and other mitochondrial stains currently available. MitoView Fix 640 is ideal for the Cy5 channel and is offered in 1 or 10 units of 50 µg.

Biotium

For info: 800-304-5357
biotium.com

8-Well Chamber Slide

ibidi's top-selling, all-in-one 8-well chamber slide, the µ-Slide 8 Well, now has extra-high individual walls to keep cross-contamination between wells as low as possible when performing high-resolution, cost-effective cell-culture experiments. With its thin ibidi Polymer Coverslip bottom, the µ-Slide 8 Well high allows for excellent cell adhesion onto the tissue culture–treated surface. In addition, it has the highest optical quality and is ideally suited for a variety of microscopic techniques, such as widefield fluorescence, confocal microscopy, and differential interference contrast (DIC). For scientists who perform special microscopic applications, such as total internal reflection fluorescence (TIRF) and superresolution microscopy, ibidi has created the µ-Slide 8 Well high Glass Bottom with a #1.5H D 263 M Schott glass. It is also available as a sticky-Slide 8 Well high without any bottom, enabling the researcher to mount any chosen substrate. The entire µ-Slide 8 Well high family is ideal for experiments with small cell numbers and low reagent volumes. Scientists who would like to test one of the µ-Slide 8 Well high variations with their own experiments can request free samples.

ibidi

For info: 844-276-6363
ibidi.com

SLM-Based All-Optical Solution

The HoloStim-3D is an all-optical solution that facilitates simultaneous imaging and photostimulation of hundreds of individual cells in 3D, enabling novel insights into the roles of neural subtypes and individual neurons in brain circuits and behavior. A compact module, the HoloStim-3D seamlessly integrates with Scientifica's HyperScope, a multiphoton imaging system, creating a fully integrated spatial light modulator (SLM)-based photostimulation solution. With its industry-leading resolution, cell bodies and subcellular structures can be specifically targeted with less off-target activation than other systems on the market, generating cleaner, more accurate data. HyperScope and HoloStim-3D allow different imaging and stimulation wavelengths to freely combine for excitation and imaging of target cells expressing different opsins. Full integration with ScanImage software means that the HoloStim-3D is easier to adopt than ever before, making it an excellent choice for researchers who want to start experimenting right away.

Scientifica

For info: 609-454-6069
www.scientifica.uk.com/products/scientifica-holostim-3d

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

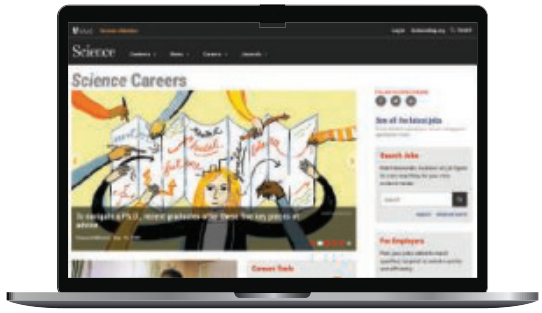
Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Science Careers online @sciencecareers.org

Science Careers helps you advance your career. Learn how!

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP."
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



SCIENCECAREERS.ORG

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Zhengzhou University is one of the key universities of the national "211 Project", and of "World 1st University Project", as well as of universities co-sponsored by the Henan Provincial Government and the Ministry of Education. It was born in 10th, July 2000 when the former ZZU, Zhengzhou University of Technology (ZUT) and Henan Medical University (HMU) merged into one, and now developed into a comprehensive university with 12 major disciplines, namely liberal arts, science, engineering and medicine, etc. It has over 50,000 full-time undergraduates, more than 23,000 full-time postgraduates and nearly 2,500 international students. Meanwhile, there are 30 first-level disciplines authorized to confer doctorate degrees, 3 professional doctorate programs, 28 postdoctoral research stations; 3 "first-class" disciplines and 6 national key (being cultivated) disciplines. ZZU is home to 12 national scientific and research platforms including state key laboratories, engineering research centers, engineering laboratories, etc. Among a total of 5800 faculty members, there are 16 academicians of Chinese Academy of Science and Chinese Academy of Engineering, 2 members of Chinese Academy of Social Sciences and 4 overseas academicians as well as78 elected into national talent programs (projects). They have formed a strong team with academicians and academic masters as the head, distinguished scholars as academic leaders, and outstanding PhDs as core talents.

Rooted in the broad and profound culture of the Central Plains of China and nurtured by blended diversified cultures and disciplines, ZZU people have developed the virtues of tolerance, lenience and can-do spirit and pursued ZZU's mission and spirit of seeking truth and being responsible, thus forming typical school ethos-- Perseverance, Trust-worthiness, Benevolence, Tolerance, Modesty, Prudence, and Critical Thinking and Diligence. Standing at this new starting line, ZZU decided where to go and laid out a three-step strategy to make it to the list of world-class comprehensive & research-oriented universities by the middle of this century. To achieve high-quality development, the university has made efforts to strengthen itself though talents. With the idea of "top-notch talents breakthrough, team integration, talent cultivation and structural optimization", ZZU implemented Distinguished Professor Program to attract more academic masters, and Innovation Talent Cultivation Program to create an ideal environment for the growth of innovative talents.

ZZU, your stage to make a difference!

RECRUITING POSITION

Distinguished Professor, Top Talent

A competitive remuneration package and scientific research services referring to the standards of world-class universities will be provided for the right candidates.

HOW TO APPLY

Please send full curriculum vitae (please note the position and discipline you are applying for) to

rcb@zzu.edu.

THE RESOURCES & MATERIALS DISCIPLINE

The Resources & Materials Discipline covers material science and engineering, metallurgy, and materials physics and chemistry, etc., aiming to address scientific fundamentals essential to technical applications of advanced materials and resources. The discipline is committed to achieve academic excellency through establishing a knowledge center and innovation base towards sustainable exploitation of resources, their green and low-carbon processing and commercialization of advanced materials technologies to impact upon the wider community. The discipline consists of four research divisions: mineral processing and materialization, material design and green metallurgy, advanced material forming and molding, refractory and tribological materials. Currently, the Resource & Materials Discipline hosts eight national platforms of excellency, such as the National Engineering Research Center for Rubber and Plastic Molds, the National and Regional Joint Engineering Research Center for Green Mining, Smelting and Processing, and the Innovation and Talents Recruitment Base for Light and Functional Polymer Molding. In the recent five years, we have engaged in innovative research and development of high-performance materials with lowered consumption of resources and/or energy, leading to technological breakthroughs for green extraction of rare metals of great strategic importance. This has been proactively coupled with advancements in theoretical modeling and process simulation towards cost-effective delivery of technological solutions. Among various achievements are advanced technologies for the fabrication of large-scale targets for functional coatings, and molding technologies for stress-free forming of plastic parts under severe service conditions. Also, in 2018, the discipline launched an international journal, Energy & Environmental Materials, which has been indexed in major databases such as the SCIE and Scopus less than 3 years (July and September, 2020), owing to publication of high quality papers with projected IF over 15 in 2021. The Discipline's world ranking has also been significantly improved, in terms of the global ESI ranking, World Ranking of Universities. We would appreciate your kind support and fruitful collaboration and we sincerely invite talented scholars to join us for even greater success, together!

Contact: Mr. Li Email: clzp@zzu.edu.cn Tel: +86-371-67781590

SOMETIMES THE GRASS REALLY IS GREENER SOMEPLACE NEW.

Find your next job at ScienceCareers.org

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place: sciencecareers.org

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

Who's the top employer for 2020?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers

Science 2020 TOP EMPLOYER

CBC CHICAGO BIOMEDICAL CONSORTIUM

CBC Entrepreneurial Fellowships

Applications are being accepted for the **CBC Entrepreneurial Fellows Award** program. Fellows are exposed to a breadth of real-world experiences across the CBC institutions. Fellows will be recent PhD or MD/PhD recipients or will have recently completed an initial post doc in the biomedical sciences.

- Supports the professional development of academic researchers who are keen to develop the skills and experiences needed to move translational projects from a university lab toward commercialization.
- Exposure to all aspects of drug development, supporting projects at NU, UIC and UChicago having therapeutic and commercial potential.
- Fellows assist in development of faculty proposals, collection and assessment of data supporting the unmet medical need, develop a work plan to advance the science toward pre-clinical proof-of-concept while assessing drug discovery/development risks.

Description of appointment:

- Highly competitive salary with fringe benefits
- Appointments are for 12 months; may be renewed for an additional 12 months by mutual agreement

Learn more about the position, qualifications and application process at:
www.chicagobiomedicalconsortium.org

The mission of the CBC is to stimulate collaboration among scientists at NU, The University of Chicago, UIC and others to accelerate discovery that will transform biomedical research and improve the health of humankind.

THE UNIVERSITY OF CHICAGO
UNIVERSITY OF ILLINOIS AT CHICAGO
NORTHWESTERN UNIVERSITY

For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers | AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

By **Nancy Saana Banono**

When do I tell my supervisor?

There seemed to be no point in mentioning the possibility yet. I was a few years into my Ph.D., and I had just applied for a 3-month industry internship program. Starting even before I joined my supervisor's lab, I had been open with her about my interest in an industry career. She had always been supportive, including sharing her experience transitioning from biotech to academia. But I wasn't entirely sure she would be happy with the internship timing, which would mean taking a break from my Ph.D. during my final year. Anyway, I doubted that I would even get in. It didn't seem worth opening a potential can of worms for something that felt unlikely to happen.

Then, in January 2020, I was shocked and thrilled when I received word that my application was successful! But soon, things took a turn. I could have used my supervisor's help. Yet I still hesitated, feeling I needed to present her with a clean, confident proposal of my plans rather than lean on her for support.

The internship coordinators matched me to a company based on my experience and interests, and it seemed like a great fit at first. But after meeting with the CEO, I began to have second thoughts. The company wanted me to start in the spring. However, I was in the middle of critical experiments for two manuscripts and hoped to start a little later in the year to minimize the disruption to my Ph.D. Also, the company did not have a clear project for me, and I worried that I could end up doing busywork rather than getting an authentic industry research scientist experience.

I spent the next few weeks mulling over my options. Should I accept the offer, knowing I might not find any job satisfaction and that my Ph.D. could be derailed? Or should I gamble on finding an opportunity that was a better fit? The funds had to be used within 2020, so I would be racing against time to find another host company and get everything set up. Moreover, most of the companies from the selection pool were outside Oslo, where I live with my husband, and we did not want to live apart, having already been through that once before.

Ultimately, I rejected the original company's offer, cautiously optimistic that I would be able to find a new host in time that would be a better match. It was a big gamble, and it finally spurred me to face my fears and involve my supervisor. I worried that I would lose this opportunity, and my supervisor seemed to be the best person to help save it. Although I did not know how she would take the news, I mustered the courage to send her an email explaining my predicament.

"Choosing your own path takes determination and resilience, but you don't have to do it alone."

I was surprised and relieved when I received her positive response. She didn't even question me about the late disclosure. She was happy for me to take a 3-month break. We had honest conversations and decided on a reasonable timetable that would make the transition back into my Ph.D. project as smooth as possible and wouldn't significantly delay my degree. Through her contacts, she also found three companies in Oslo that matched my interests, and ultimately I found a suitable host. In hindsight, I wish I had talked with her sooner.

In September 2020, I said goodbye for now to my academic colleagues and headed to industry. The internship gave me time away from the repetitive tasks of my Ph.D. and offered new challenges. It helped me appreciate that much about being a scientist is the same everywhere, and that the skills I had acquired during my Ph.D. were useful in industry. And I built valuable connections that could be helpful in the future. I'm now back in my Ph.D. lab, finishing up my thesis and scheduled to graduate just a few months later than if I had not done the internship.

The experience taught me that you have to push for what you need. Choosing your own path takes determination and resilience, but you don't have to do it alone. Be open with your mentors, and reach out for help. ■

Nancy Saana Banono is a doctoral research fellow at the University of Oslo in Norway. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Science Webinars

AAAS

Science Webinars help you keep pace with emerging scientific fields!

Stay informed about scientific breakthroughs and discoveries.

Gain insights into current research from top scientists.

Take the opportunity to ask questions during live broadcasts.

Get alerts about upcoming free webinars.

Sign up at: webinar.sciencemag.org/stayinformed

CALL FOR NOMINATIONS

2022 AAAI Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity

$1,000,000 Prize

The AAAI Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity recognizes positive impacts of artificial intelligence to protect, enhance, and improve human life in meaningful ways with long-lived effects. The award was given for the first time in 2021.

The award is given annually at the conference for the Association for the Advancement of Artificial Intelligence (AAAI), and is accompanied by a prize of $1,000,000 plus travel expenses to the conference.

Candidates may be individuals, groups, or organizations that are directly connected with the main contribution stated in the nomination. Qualifications and technical knowledge in artificial intelligence are not requirements for nominations. The emphasis is on the significance and impact of the work.

The award is administered by AAAI, with support from the European Artificial Intelligence Association (EurAI). Financial support for the award is provided by Squirrel AI.

DEADLINE for 2022 Award Nominations: **June 15, 2021**

https://www.aaai.org/Awards/squirrel-ai-award.php