Drought effects in Biosphere 2
pp. 1442 & 1514

Grouping cancer-associated T cells pp. 1446 & 1462

All-season thermal regulation pp. 1501 & 1504

Science

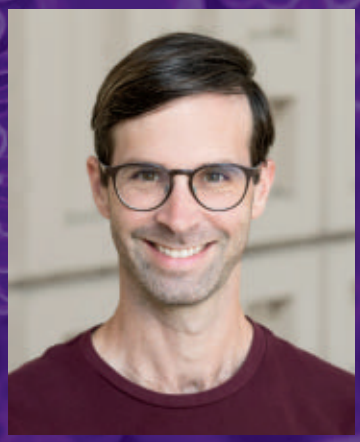
$15
17 DECEMBER 2021
SPECIAL ISSUE
science.org
AAAS

BREAKTHROUGH OF THE YEAR

2021

TRILLIONS OF MICROBES ONE ESSAY

The NOSTER *Science* Microbiome Prize is an international prize that rewards innovative research by investigators, under the age of 35, who are working on the functional attributes of the microbiota. The research can include any organism that has potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent scientist, chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and $25,000 (USD). Submit your research essay today.



Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 1/24/22 at www.science.org/noster

Sponsored by Noster, Inc

EMPOWER WITH EVIDENCE

AAAS | ANNUAL MEETING

Join us for the hybrid 2022 AAAS Annual Meeting, held in Philadelphia, and online, February 17-20.

AAAS requires that all participants in its in-person meetings be fully vaccinated.

Browse the meeting program online and learn about the invited lecturers, scientific sessions, workshops, and other happenings at #AAASmtg that highlight noteworthy advances and new perspectives in STEM.

BE A PART OF THE CONVERSATION

Public Health | Science Communication | Big Data | Artificial Intelligence

Stem Education | Science Policy | Social and Behavorial Sciences | Innovation

Diversity and Equity | Climate and Sustainibilty | Linguistics | Global Perspectives

and more...

Advance registration rates are available now through **January 24, 2022**

aaas.org/meetings | #AAASmtg

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

CONTENTS

17 DECEMBER 2021 • VOLUME 374 • ISSUE 6574

SPECIAL SECTION

Breakthrough of the Year

WINNER

1426 Protein structures for all
AI-powered predictions reveal the shapes of proteins by the thousands
By R. F. Service

RUNNERS-UP

1428 Ancient soil DNA comes of age

1428 Fusion's day in the Sun?

1429 Potent pills boost COVID-19 arsenal

1430 A psychedelic PTSD remedy

1430 Artificial antibodies tame infectious diseases

1431 NASA lander uncovers the Red Planet's core

1432 At last, a crack in particle physics' standard model?

1433 CRISPR fixes genes inside the body

1433 Embryo 'husbandry' opens windows into early development

1434 Missed shots
By doses delivered, the COVID-19 vaccine rollout was a spectacular success. By other measures, it went tragically awry
K. Kupferschmidt

1436 Breakdowns of the year
What went wrong in the world of science

ON THE COVER

Artificial intelligence algorithms can now churn out predictions for the 3D shapes of proteins, key to their function, with a precision matching that of painstaking laboratory techniques. The programs, and the blizzard of protein structures they have revealed, are *Science*'s 2021 Breakthrough of the Year. *Illustration: V. Altounian/*Science*; Data: I. R. Humphreys* et al., Science ***374****, eabm4805 (2021); DOI: 10.1126/science.abm4805*

SEE ALSO EDITORIAL p. 1415 VIDEO PODCAST

NEWS

IN BRIEF

1418 News at a glance

IN DEPTH

1420 Key Antarctic ice shelf is within years of failure
Breakup of shelf holding back Thwaites Glacier will ramp up sea level rise
By P. Voosen

1421 Scientists see a 'really, really tough winter' with Omicron
Another major pandemic wave seems inevitable. The big question is how much severe disease it will bring *By K. Kupferschmidt*

1422 India defuses its population bomb
Sterilization, contraceptives push fertility rate down to two children per woman *By F. Pearce*

1424 To draw down carbon, ocean fertilization gets another look
Panel urges large tests of iron-triggered plankton blooms *By W. Cornwall*

1425 Blood condition may guard against Alzheimer's
Unexpected protection may result from wayward blood cells that enter brain
By M. Leslie

INSIGHTS

POLICY FORUM

1438 COVID-19 impact on infant and adolescent vaccine supplies
Vaccine production is quadrupling rapidly, creating supply chain challenges
By T. Cernuschi et al.

PERSPECTIVES

1442 Ecosystem effects of environmental extremes
A large-scale experimental facility reveals tropical rainforest responses to drought
By N. Eisenhauer and A. Weigelt
REPORT p. 1514

1443 Toward single-molecule proteomics
Nanopore rereading of single proteins opens a pathway to next-generation proteomics
By F. Bošković and U. F. Keyser
REPORT p. 1509

1445 Nitride perovskite becomes polar
An oxygen-free polar perovskite offers several advantages over perovskite oxides *By X. Hong*
REPORT p. 1488

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*; DATA: I. R. HUMPHREYS *ET AL.*, *SCIENCE* **374**, EABM4805 (2021); DOI: 10.1126/SCIENCE.ABM4805

1446 An atlas of intratumoral T cells
Intratumoral T cell composition is relevant for disease outcome across tumor types
By A. M. van der Leun and T. N. Schumacher
RESEARCH ARTICLE p. 1462

1447 Radical quantum oscillations
Laser spectroscopy reveals spin quantum beats in electron transfer reactions
By P. J. Hore
RESEARCH ARTICLE p. 1470

1449 Biological basis of cannabinoid medicines
Mechanistic insights into cannabinoid signaling could improve therapeutic applications *By E. Keimpema* et al.

BOOKS *ET AL.*

1451 Climate change in the classroom
Inconsistent and agenda-driven K–12 curricula leave US citizens ill-equipped to confront environmental problems *By A. Huderson*

1452 Searching for solutions to our soil woes
Could a controversial carbon storage plan help restore degraded lands? *By D. D. Richter*

LETTERS

1454 Editorial Expression of Concern
By H. H. Thorp

1454 Track Omicron's spread with molecular data
By L. Scott et al.

1455 News stories must account for gender bias
By J. G. Hering et al.

1455 Past as prologue: A uranium miner's daughter
By T. J. Gallegos

1456 Editor's Note
By T. Appenzeller

RESEARCH

IN BRIEF

1458 From *Science* and other journals

RESEARCH ARTICLES

1461 Genomics
Pangenomics enables genotyping of known structural variants in 5202 diverse genomes
J. Sirén et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG8871

1462 Cancer immunology
Pan-cancer single-cell landscape of tumor-infiltrating T cells *L. Zheng* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABE6474
PERSPECTIVE p. 1446

1463 Coronavirus
Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant
P. Elliott et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABL9551

1464 Carbon capture
A scalable metal-organic framework as a durable physisorbent for carbon dioxide capture *J.-B. Lin* et al.

1470 Spin chemistry
Readout of spin quantum beats in a charge-separated radical pair by pump-push spectroscopy *D. Mims* et al.
PERSPECTIVE p. 1447

REPORTS

1474 Quantum simulation
Many-body–localized discrete time crystal with a programmable spin-based quantum simulator *J. Randall* et al.

1479 Quantum simulation
Information scrambling in quantum circuits *X. Mi* et al.

1484 Magnetism
Atomic-scale visualization of topological spin textures in the chiral magnet MnGe
J. Repicky et al.

1488 Perovskites
Synthesis of $LaWN_3$ nitride perovskite with polar symmetry *K. R. Talley* et al.
PERSPECTIVE p. 1445

1492 Neuroscience
Supramammillary regulation of locomotion and hippocampal activity
J. S. Farrell et al.

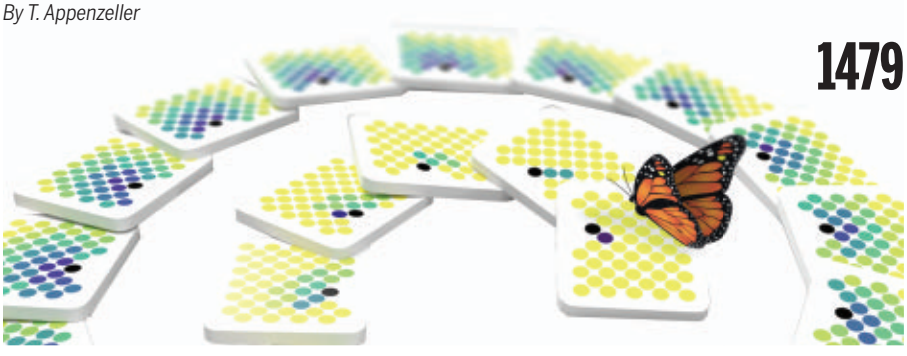
1496 Physiological stress
Physiological costs of undocumented human migration across the southern United States border
S. C. Campbell-Staton et al.

Radiative cooling
1501 Scalable thermochromic smart windows with passive radiative cooling regulation
S. Wang et al.

1504 Temperature-adaptive radiative coating for all-season household thermal regulation
K. Tang et al.

1509 Biotechnology
Multiple rereads of single proteins at single–amino acid resolution using nanopores
H. Brinkerhoff et al.
PERSPECTIVE p. 1443

1514 Forest ecology
Ecosystem fluxes during drought and recovery in an experimental forest
C. Werner et al.
PERSPECTIVE p. 1442



Investigation of quantum scrambling on a superconducting chip yields observations of the "butterfly velocity."

CREDIT: GOOGLE QUANTUM AI

DEPARTMENTS

1415 Editorial
Proteins, proteins everywhere
By H. Holden Thorp
BREAKTHROUGH OF THE YEAR SECTION p. 1426

1526 Working Life
A path to independence *By Angela Hessler*

Science Staff 1416
Science Careers 1519

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

Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS.

Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

eppendorf

New epT.I.P.S.®
Box 2.0

Clever and Smart

Time for a fresh reboot

Often copied but unrivaled, the epT.I.P.S.® Box has been in laboratories worldwide since 2002. The fresh new epT.I.P.S Box 2.0 design features a bright white tone, clean lines and a redesigned closure button in typical Eppendorf blue, as well as the bluish-tinted transparent lid with indentations on the sides.

> No slits on the back of boxes to reduce risk of contamination
> Autoclavable up to 100 x for a long lifetime
> Lid with indentations on sides to improve stackability
> Small silicone feet in the bottom of the box reduce risk of slipping

www.eppendorf.com/epTIPS-News

Eppendorf®, the Eppendorf Brand Design, epT.I.P.S.® and PhysioCare Concept® are registered trademarks of Eppendorf AG, Germany.
All rights reserved, including graphics and images. Copyright © 2021 by Eppendorf AG.

EDITORIAL

Proteins, proteins everywhere

The first protein structures were determined by x-ray crystallography in 1957 by John C. Kendrew and Max F. Perutz. As a bioinorganic chemist, I was delighted that the structures were myoglobin and hemoglobin, both heme proteins with big, beautiful iron atoms. It must have been an extraordinary experience to stare at a physical model of the structures and see something that had previously only been imagined. Not long afterward, Christian B. Anfinsen Jr. proposed that the structure of a protein was thermodynamically stable. It seemed possible that the three-dimensional structure of a protein could be predicted based on the sequence of its amino acids. This "protein-folding problem," as it came to be known, baffled scientists until this year, when the papers we've deemed the 2021 Breakthrough of the Year were published.

I came of age as a biochemist in an era when scientists could describe but not solve protein folding. Chemical force field and molecular dynamics modeling just couldn't quite get there. The early days of de novo protein synthesis were heady because complicated structures could be designed and synthesized, but elaborate structures and multiprotein assemblies remained out of reach. When I left the lab in 2006 to go fight administrative windmills, I thought the protein-folding problem would never be solved.

A lot has happened in the 15 years since. Structures started coming much faster after the development of cryo–electron microscopy (cryo-EM), a method requiring very expensive instrumentation that can determine protein structures without crystallized samples. Then the Critical Assessment of Protein Structure Prediction (CASP) competition, in which scientists were challenged to match algorithms to known protein structures, really took off. As these annual competitions proceeded, DeepMind—a UK subsidiary of Alphabet, Google's parent company—began showing results from an artificial intelligence algorithm that mined known protein structures to predict unknown structures. Their program, AlphaFold, made a big impression at the fall 2020 CASP conference on Minkyung Baek, a researcher at the University of Washington. Baek devised RoseTTAFold, an algorithm that uses less computing power and can predict structures of protein complexes. The methods of DeepMind and Baek *et al.* were published simultaneously in *Nature* and *Science*, respectively, this year. Both AlphaFold and RoseTTAFold are available free for researchers (as is a database of protein structure predictions), and scientists immediately began obtaining protein structures from these algorithms without having to crystallize their proteins or access cryo-EM tools. Now structures can be obtained for samples that defy experimental methods, and moreover, in labs that can't afford the experimental approaches. It is truly protein structure for all.

This is a breakthrough on two fronts. It solves a scientific problem that has been on the to-do list for 50 years. And just like Fermat's Last Theorem or gravitational waves, scientists kept at it until it was done. Also, it's a game-changing technique that, like CRISPR or cryo-EM, will greatly accelerate scientific discovery. It would have been the Breakthrough of the Year for either reason, but it's a twofer for sure.

> **"The breakthrough in protein folding is one of the greatest ever..."**

The breakdowns of the year are all topics that were covered heavily in *Science*. The disastrous US Food and Drug Administration (FDA) approval for Alzheimer's disease of aducanumab—a costly treatment with little if any clinical efficacy—was discussed in an Editorial by Joel Perlmutter. Perlmutter resigned in protest from the committee that advised the FDA on the drug. Continued inaction on climate has rendered unlikely the 1.5°C target as a limit on the increase in global temperature. And scientists have been attacked in new ways and with new methods by politicians who are exploiting social media and long-tested methods of indoctrination to undermine scientific authority on issues ranging from vaccines in the midst of a pandemic to the impending devastation of climate change.

The contrasts of the breakthrough and breakdowns are stunning. The breakthrough in protein folding is one of the greatest ever in terms of both the scientific achievement and the enabling of future research. Meanwhile, the tragic loss of respect for scientific authority around the world is demoralizing, and it's hard to see the trend reversing any time soon.

It is truly the best of times and the worst of times in 2021—an age of wisdom and an age of foolishness. The scientific tale of two cities will be with us for a while to come.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org;
@hholdenthorp

10.1126/science.abn5795

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, Priscilla N. Kelly, 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, Yuen Yiu, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Bianca Lopez, 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 Haber-Katris, Nida Masiulis, Abigail Shashikanth, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Alexander Kief, Ronmel Navas, Isabel Schnaidt, 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, Kelly Servick, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERN** Rachel Fritts **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, Gabriel Popkin, Michael Price, Joshua Sokol, Richard Stone, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **INTERIM GRAPHICS MANAGING EDITOR** Chris Bickel **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB 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 GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR SCIENTIFIC ILLUSTRATOR** Valerie Altounian **SCIENTIFIC ILLUSTRATOR** Ashley Mastin **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski **SENIOR PODCAST PRODUCER** Sarah Crespi **VIDEO PRODUCER** Meagan Cantwell

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 **MANAGER, BUSINESS ANALYSIS** Cory Lipman **BUSINESS ANALYST** Kurt Ennis **FINANCIAL ANALYST** Isacco Fusi **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT AND PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR CONTENT SPECIALISTS** Steve Forrester, 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** Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris, Justin Wood **SENIOR DESIGNER** Kim Huynh

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

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

MARKETING MANAGER Kess Knight **BUSINESS DEVELOPMENT MANAGER** Rasmus Andersen **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan, Claudia Paulsen-Young **SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES** Chris Hoag **US WEST COAST SALES** Lynne Stickrod **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **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** Yoshimi Toda (Tokyo), Miyuki Tani (Osaka)

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

---

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

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

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

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

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

MEDIA CONTACTS
scipak@aaas.org

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

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

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

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

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, Janine 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)**)

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena, *Inst. Curie*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
David Awschalom, *U. of Chicago*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Nandita Basu, *U. of Waterloo*
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.*
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*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
M. Keith Chen, *UCLA*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
Amander Clark, *UCLA*
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*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *UPenn*
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*
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*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Christoph Hess, *U. of Basel & U. of Cambridge*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Janneke Hille Ris Lambers, *ETH Zürich*
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*
Kisuk Kang, *Seoul Nat. U.*
Sabine Kastner, *Princeton U.*
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*
Paul Kubes, *U. of Calgary*
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*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Suresh Naidu, *Columbia U.*
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci. & 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*
Ana Pêgo, *U. do Porto*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. 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 Ctr.*
Richard Smith, *UNC* **(S)**
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Northwestern U.*
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.*
Maria-Elena Torres Padilla, *Helmholtz Zentrum München*
Kimani Toussaint, *Brown U.*
Barbara Treutlein, *ETH Zürich*
Jason Tylianakis, *U. of Canterbury*
Wim van der Putten, *Netherlands Inst. of Ecology*
Ivo Vankelecom, *KU Leuven*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Julia Von Blume, *Yale School of Med.*
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)**
Hao Wu, *Harvard U.*
Wei Xie, *Tsinghua U.*
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*

FLEXIBLE.
SENSITIVE.
DEPENDABLE.

CLARIOstar® *Plus*

The CLARIOstar® *Plus* multi-mode microplate reader streamlines assay development and validation by combining monochromator flexibility with best-in-class sensitivity.

- LVF Monochromators™ for highest sensitivity
- EDR technology makes gain adjustment no longer required
- Dedicated detectors for luminescence and red fluorescence
- Best performance in TRF, TR-FRET, FP and AlphaScreen®
- Atmospheric control with gas ramping function
- Made-in-Germany dependability

www.bmglabtech.com

BMG LABTECH

The Microplate Reader Company

©2021 Allrightsreserved.Alllogosandtrademarksarethepropertyof BMGLABTECH

NEWS

> “This is like a teenager promising to clean their room in 30 years. We need action now.”
>
> **Bill Snape,** senior counsel at the Center for Biological Diversity, on President Joe Biden’s 8 December announcement of a plan for the U.S. government to achieve net-zero greenhouse emissions by 2050.

IN BRIEF *Edited by* **Kelly Servick**

Wreckage of the main building at the University of Kentucky’s Research and Education Center.

RESEARCH FACILITIES

Tornado razes agricultural center

A hub of agricultural research in the small city of Princeton, Kentucky, was among the many places left in ruins last week as a series of tornadoes ripped through the region, killing dozens. No employees of the University of Kentucky’s Research and Education Center were killed, and just one person there suffered minor injuries. But nearly all of the 60 buildings on-site—including student housing, offices, and research facilities—have been condemned. Director Carrie Knott estimates it will be at least a year before the center’s 15 principal investigators can resume their research programs, which include agronomy, plant pathology, and horticulture. Center leadership will spend the upcoming days removing debris and setting up temporary offices and storage facilities for employees.

Vaccine mimics virus particle

COVID-19 | A vaccine with a unique composition and production method worked in a large COVID-19 efficacy trial, its sponsor, Canada-based biotech company Medicago, announced in a press release last week. The vaccine consists of SARS-CoV-2 spike proteins that self-assemble into viruslike particles (VLPs). Already used in vaccines against human papillomavirus and hepatitis B, VLPs in theory stimulate robust immune responses because of the orderly way they pack many copies of the viral proteins into a particle that resembles a virus. Medicago produces the spike proteins in a genetically engineered plant, a tobacco cousin called *Nicotiana benthamiana*, rather than in lab cell cultures. The vaccine had 71% efficacy against symptomatic disease in a 24,000-person trial in six countries, where many variants were circulating—although not Omicron. The company is now seeking authorization in Canada and, pending success there, plans to submit data to U.S. and European regulators.

Tabak named acting NIH chief

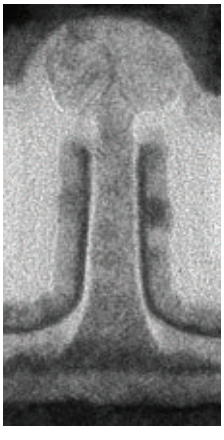
SCIENCE POLICY | Lawrence Tabak, principal deputy director of the National Institutes of Health (NIH), will become the agency’s acting director on 20 December—the day after current Director Francis Collins leaves his post. Collins, the physician-geneticist who has led the $43 billion agency for 12 years, announced his resignation in October. Before assuming his current role at NIH in 2010, Tabak, a dentist and biochemist, spent 10 years directing the National Institute of Dental and Craniofacial Research. Onlookers say he will bring stability as the White House searches for the next director. The permanent position requires approval by the Senate health committee and full Senate.

New Zealand to phase out tobacco

PUBLIC HEALTH | The government of New Zealand last week released an unprecedented proposal to outlaw sales of tobacco products to those turning 14 years old, a prohibition that would last their lifetimes and gradually end sales entirely. The country’s Smokefree 2025 Action Plan would also make it the first in the world to only allow sales of low-nicotine smoked tobacco products. Authorities decided on bold action after modeling showed a diminishing effect of existing strategies, including high taxes and bans on public smoking. The plan also calls for greater investment in smoking cessation support for the Māori community, which has a smoking rate of 24.8%, compared with 10.1% among New Zealanders of European heritage. Parliamentary approval is expected next year, with the measures likely taking effect in 2023.

PHOTO: MATT BARTON/UK COLLEGE OF AGRICULTURE

A new semiconductor chip contains rows of vertically stacked transistors (micrograph below).



PHOTOS: (TOP TO BOTTOM) CONNIE ZHOU FOR IBM; IBM

ELECTRICAL ENGINEERING

Vertical leap for chips

Microscopic transistors at the heart of computers and phones may soon pack a bigger punch. Researchers have stood these thin transistors vertically on their ends, making it possible to pack them in tighter on silicon chips to enable faster or more energy-efficient devices, IBM and Samsung reported last week at the International Electron Devices Meeting. Traditional transistors lie flat with electrical current moving through them laterally. But as they've shrunk to the nanometer scale, engineers have struggled to get sufficient current through their tiny electrical channels. The vertical transistors offer room for larger electrical channels and other components, which will enable future devices to run either twice as fast, or with 85% less energy use, researchers reported.

Arctic warming understated

CLIMATE SCIENCE | It's a common statistic, found even in this year's U.N. climate assessment: The Arctic is warming about twice as fast as the rest of the world. But that figure is misleading, scientists reported at a meeting of the American Geophysical Union this week. Since 1990, the Arctic has actually warmed four times faster, an amplification caused by melting sea ice and other factors. One reason for the discrepancy: Climate scientists often use 60°N to define "Arctic" rather than the more technically correct 66.6°N, lumping in lower latitudes where there's less amplification. Another reason: Many analyses use data from earlier time periods when light-reflecting pollution blocked Arctic warming. As a result, the researchers say, the true toll of global warming on the Arctic has been underestimated.

Researcher gets COVID-19 in lab

LAB SAFETY | A Taiwanese researcher contracted SARS-CoV-2 while working with infected mice in a biosafety level 3 laboratory at the Genomics Research Center of Academia Sinica in Taipei, in the first known case of laboratory transmission of the virus. Authorities suspect the vaccinated researcher was infected via a mouse bite. The genomic sequence of the variant infecting the researcher matched that of the lab's mice and not strains in the surrounding community, Taiwan's Central Epidemic Command Center reported on 11 December. More than 800 potential contacts have tested negative, and Academia Sinica has promised an investigation and a review of the lab's safety procedures. The incident is likely to bolster claims that a lab leak could have sparked the pandemic.

#METOO UPDATES

Prominent male scientists faced new allegations this week, while another's accuser fought his defamation lawsuit.

Paleontologist **Leonardo dos Santos Avilla** was placed on administrative leave from the Federal University of the State of Rio de Janeiro (UNIRIO) this week following an exposé on the Brazilian TV channel Fantástico in which former students accused him of bullying and sexual misconduct including unwanted kissing, touching, and coercive sex. At least 30 women have raised allegations dating back as far as 2007, the segment revealed. UNIRIO says Avilla will remain on leave until an investigation, begun on 7 December, is finalized. Avilla "vehemently denies the allegations," his lawyer said in the exposé.

More than a dozen women who worked or studied at the **Smithsonian Tropical Research Institute** in Panama are alleging harassment and exploitation by multiple STRI scientists, a *BuzzFeed News* investigation revealed last week. Allegations against soil biochemist **Benjamin Turner** include rape, unwanted touching, off-color remarks, and abuse of power. Turner, who *BuzzFeed* reports was removed in 2020 following an STRI investigation, has denied the allegations. Although STRI has implemented new policies to guard against abuse by its staff, lawyers representing 14 women scientists have written President Joe Biden's Gender Policy Council urging it to address these issues at STRI and other science institutions.

A woman last week lodged a counterclaim in a defamation suit filed in October by former Whitehead Institute biologist **David Sabatini**. He resigned in August after a probe concluded that he had violated the institute's sexual harassment policies. Sabatini's suit names Whitehead, its director Ruth Lehmann, and one of his accusers. That accuser's counterclaim, entered in Massachusetts Superior Court, says Sabatini's "frivolous" suit retaliates against her for speaking frankly with Whitehead investigators. It also alleges that he coerced her into sex when she was a graduate student and fostered a "toxic" and "sexualized" lab environment. Sabatini is on administrative leave at the Massachusetts Institute of Technology, where he is a professor. It is considering revoking his tenure.

At the foot of the Thwaites Glacier is a 45-kilometer-wide ice shelf. It has begun to fall apart.

CLIMATE CHANGE

Key Antarctic ice shelf is within years of failure

Breakup of shelf holding back Thwaites Glacier will ramp up sea level rise

By **Paul Voosen**

An alarming crackup has begun at the foot of Antarctica's vulnerable Thwaites Glacier, whose meltwater is already responsible for about 4% of global sea level rise. An ice sheet the size of Florida, Thwaites ends its slide into the ocean as a floating ledge of ice 45 kilometers wide. But now, this ice shelf, riven by newly detected fissures on its surface and underside, is likely to break apart in the next 5 years or so, scientists reported this week at a meeting of the American Geophysical Union.

The most dramatic sign of impending failure is a set of diagonal fractures that nearly span the entire shelf. Last month, satellites spotted accelerating movement of ice along the fractures, says Erin Pettit, a glaciologist at Oregon State University, Corvallis, who is part of a multiyear campaign to study the glacier. The shelf is like a windshield with a series of slowly opening cracks, she says. "You're like, I should get a new windshield. And one day, bang—there are a million other cracks there."

Once the ice shelf shatters, large sections of the glacier now restrained by it are likely to speed up, says Ted Scambos, a glaciologist at the University of Colorado, Boulder, and a leader of the Thwaites expedition. In a worst case, this part of Thwaites could triple in speed, raising the glacier's contribution to global sea level to 5%, Pettit says.

Even more worrisome is the process that has weakened the ice shelf: incursions of warm ocean water beneath the shelf, which expedition scientists detected with a robotic submersible. Because Thwaites sits below sea level on ground that dips away from the coast, the warm water is likely to melt its way inland, beneath the glacier itself, freeing its underbelly from bedrock. A collapse of the entire glacier, which some researchers think is only centuries away, would raise global sea level by 65 centimeters. And because Thwaites occupies a deep basin into which neighboring glaciers would flow, its demise could eventually lead to the loss of the entire West Antarctic Ice Sheet, which locks up 3.3 meters of global sea level rise. "That would be a global change," says Robert DeConto, a glaciologist at the University of Massachusetts, Amherst. "Our coastlines will look different from space."

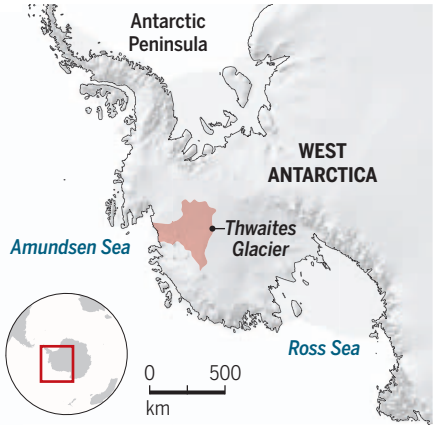
Regardless of whether the shelf breaks up in 1 year or 10 years, Pettit and her colleagues are doing important work, DeConto adds. The oceans are simply getting too warm for these marine ice sheets, which formed in much cooler conditions, he says. "This marine-based ice is not going to come back."

Exploring the future of this keystone of the West Antarctic Ice Sheet is the aim of the International Thwaites Glacier Collaboration (ITGC), a multiyear expedition costing more than $50 million and funded by the United States and United Kingdom (*Science*, 14 October 2016, p. 159). The glacier, far from any research stations, is tough to reach under the best circumstances, and ITGC's first scientific campaign on the ice, in the Antarctic summer of 2019–20, contended with severe storms. But the team managed to erect several temporary camps, including one in the middle of the ice shelf and another farther upstream, near the "grounding line" where the glacier detaches from the continent.

On top of the 300-meter-thick shelf, the researchers used ground-penetrating radar to image the underside of the ice. They were surprised to find it was not flat and smooth,

Keystone crackup

An ice shelf buttressing the Thwaites Glacier is breaking up. That could speed up the glacier's collapse, which would raise sea levels by 65 centimeters.



CREDITS: (PHOTO) NASA ICEBRIDGE/JAMES YUNGEL; (MAP) N. DESAI/*SCIENCE*

but carved into a series of upside-down valleys, some 50 meters deep. These undulations stress the ice shelf, and the team saw signs of that stress: Fractures had formed at the apex of each valley, Pettit says. "They're just waiting to be activated in a new way."

At the upstream camp, researchers led by Britney Schmidt, a planetary scientist at Cornell University, dug a borehole and sent a robot called Icefin plunging through it to the ocean hidden below. Schmidt navigated Icefin to the point where the ice and rock met. Nearly everywhere—even at the grounding line itself—the water was 1°C or 2°C above freezing. Although not unexpected, given the 10-kilometer retreat of the grounding line in the past decade, the readings were a sure sign of the extended reach of ocean warming ushered in by climate change.

During its surveys, Icefin also scanned the underside of the ice with a laser and found valleys similar to the downstream ones. Local variations in water temperature suggested the valleys create turbulence that draws in warmer waters, which deepen them, says Peter Washam, an oceanographer at Cornell. "They're really hot spots of melting."

As detailed in one of three papers about Thwaites in the journal *Cryosphere*, computer models of the ice shelf suggest the extensive surface cracks seen in the past 5 years have opened as ice thinned by melting grinds into an offshore, undersea mountain, which had long helped hold the ice shelf back. Several of these cracks, including one nicknamed "the dagger," are now extending toward the middle of the shelf. Once there, they may trigger the incipient cracks in the valleys underneath to grow and weaken the shelf further, Pettit says.

The newest wrinkle is the growth of the diagonal fractures, which stretch more than 40 kilometers from the grounding line to the offshore mountain. Although the ice directly behind the mountain still seems stuck, GPS stations placed during the first field season show slippage along the fracture zone is allowing other ice to maneuver around the mountain, which is likely to speed the shelf's collapse. "It's got enough freedom now that it can reroute itself around," Pettit says.

With several seasons left in the ITGC campaign, researchers will be able to watch the shelf disintegrate—although they'll have to retrieve their instruments before its demise, with several fissures only 3 kilometers away from one former campsite. The ice shelf failure will be a warning that Thwaites, and the rest of the West Antarctic Ice Sheet, could begin to see significant losses within decades, especially if carbon emissions don't start to come down, Pettit says. "We'll start to see some of that before I leave this Earth." ■

COVID-19

Scientists see a 'really, really tough winter' with Omicron

Another major pandemic wave seems inevitable. The big question is how much severe disease it will bring

By **Kai Kupferschmidt**

When South African scientists first alerted the world to the rapid spread of a new SARS-CoV-2 variant, some speculated it might not take off in other countries. After all, an earlier variant named Beta, which dominated in South Africa between November 2020 and May, did not spread much beyond its borders and has since petered out.

Today it's clear that the world will not be so lucky this time. Although many questions remain, scientists feel increasingly confident that the new arrival, Omicron, is likely to dramatically alter the trajectory of the pandemic—and not for the better.

Omicron has now been found in more than 70 countries and is rapidly gaining ground. As *Science* went to press, for example, Danish scientists estimated Omicron was just days away from replacing Delta as the most common variant. "What we see is an extraordinary, rapid spread," says Troels Lillebæk, an infectious disease researcher at the University of Copenhagen. Despite very high vaccination rates, the country of 5 million is now seeing more than 6000 cases a day, roughly twice the number seen during the highest previous peak. (The growth seemed to show signs of slowing down early this week, but that may be in part because the country is reaching the limits of its testing capability.) Neighboring Norway, which has about the same population, is now projecting more than 100,000 cases a day in a matter of weeks, unless people drastically reduce social contacts.

Even if Omicron causes milder disease, as some scientists hope, the astronomical case projections mean the outlook is grim, warns Emma Hodcroft, a virologist at the University of Bern. "A lot of scientists thought Delta was already going to make this a really, really tough winter," she says. "I'm not sure the message has gotten across to the people who make decisions, how much tougher Omicron is going to make this."

For Hodcroft and other virologists, immunologists, and epidemiologists, Omicron is another dizzying plunge on the pandemic roller coaster, right before the holidays—a time of frenzied phone calls,

PHOTO: ROB PINNEY/STRINGER/GETTY IMAGES

People line up outside a vaccination center in London on 13 December. The United Kingdom is accelerating its booster program in response to the rise of the Omicron variant.

late night work, and little sleep. "We have people working the whole weekend again," says Florian Krammer, a vaccine researcher at the Icahn School of Medicine at Mount Sinai. "It doesn't matter if something needs to be done at 10 p.m., it's getting done."

As with earlier variants, a handful of countries are providing the world with most of the early data. South Africa, where scientists first observed the spread of Omicron, has sequenced a wealth of genomes and produced data about early cases. Denmark, with one of the best genomic surveillance systems in the world, has provided an in-depth view of how Omicron can explode on top of a Delta surge. And scientists in the United Kingdom are conducting a host of studies to nail down how well Omicron transmits in households and elsewhere, and how vaccines are doing against the variant. "We should be pretty grateful" to these countries, Hodcroft says.

As a result, Omicron's key properties are becoming clearer by the day. It's beyond doubt that the variant has a substantial growth advantage. What's less clear is whether that's mostly because it can evade the human immune response or also because it is inherently more transmissible than its predecessors. That may not matter in the short run, but it does for the long-term outlook. If it's all about immune evasion, Omicron's advantage over Delta could wane as immunity to it builds up, and the two could end up cocirculating. If Omicron is also more infectious, it may replace Delta, just as Delta displaced earlier variants.

How good Omicron is at immune escape is also becoming clearer. Preliminary data from South Africa showed its rise coincided with an unexpected surge in reinfections. This past week, laboratory assays by several groups have shown antibodies, whether elicited by vaccines or a previous infection, are significantly less effective at neutralizing Omicron than other variants. And based on the first cases, scientists in the United Kingdom have estimated that protection from symptomatic illness is much lower in people who have received two doses of the AstraZeneca or messenger RNA vaccines. The good news is that boosters appear to bring protection against disease back to about 75%, and probably even higher against hospitalization. "I think it all boils down in the end to protection from severe disease," Krammer says.

Early data from Discovery, South Africa's largest health insurer, presented on 14 December, offered some additional reassurance that Omicron's immune escape isn't complete. The data showed hospital admissions in the country are growing more slowly than in previous waves. That could mean the protection from severe disease is still robust in vaccinated and recovered people—or the virus is inherently a bit milder than Delta.

But Harvard University epidemiologist William Hanage says the question of severity is still impossible to answer. Recent genomic comparisons suggest Omicron only began to spread in mid-October—earlier work had estimated late September—so the variant hasn't infected enough people to conclude much of anything, he says. By chance, many of the early cases in South Africa happened to be in younger people, who are less likely to develop severe disease. And even if the variant turns out to be inherently milder, the volume of cases will likely overwhelm health systems. "A colleague put it really well in one of our little depressing Slack channels," Hanage says: "There's not much that can spread this fast and be benign to a society that's already got full hospitals without it."

Scientists also worry Omicron—which represented a massive leap from known variants in genomic terms—may bring other, unpleasant evolutionary surprises. For instance, roughly one-tenth of Omicron genomes sequenced so far have an additional mutation in the spike protein called 346K that is predicted to make it even better at evading the immune system. "Omicron has most of the greatest hits for antibody escape already, so there aren't a ton of additions that it could make, but 346K is one of them," says Stephen Goldstein, a virologist at the University of Utah. "We have to keep an eye on it."

Given its divergence from earlier variants, Krammer thinks vaccine manufacturers should develop booster shots tailored to Omicron. Obtaining regulatory approval and making such boosters available in large numbers would take months, however—too long to address the crisis many scientists expect. And if the past year is any indicator, they are unlikely to be available to low- and middle-income countries in any meaningful quantities.

For now, most European countries are hoping that providing existing boosters widely, in tandem with added control measures such as a ban on large gatherings, mask mandates, better ventilation, and working from home, will help lower the wave of Omicron infections and prevent hospitals from buckling. Maria Van Kerkhove, an epidemiologist at the World Health Organization, says vaccinating those who have not received any shots at all is still very important—even though it may be too late to get the numbers up substantially. "Get the vaccine into the arms of people who are most at risk," Van Kerkhove says. "Look to see who you're missing and focus on these." ■

POPULATION

India defuses its population bomb

Sterilization, contraceptives push fertility rate down to two children per woman

By **Fred Pearce**

Back in the 1960s, India faced an exploding population, with a fertility rate of nearly six children per woman. When famine struck, U.S. President Lyndon B. Johnson initially refused to deliver food aid, citing the country's high birth rate. In response, India's Prime Minister Indira Gandhi dramatically expanded the first national family planning program in a major developing country, offering cash incentives for both men and women to be sterilized. The city of Madras, now called Chennai, paid men $6 a snip.

For the next 60 years, India continued to focus on sterilization as well as contraceptives and education for girls. Now, Indian health officials say the task of defusing their population bomb is finally done. Late last month, the National Family Health Survey (NFHS), a periodic investigation of half a million households, announced a milestone: The country's fertility rate had for the first time fallen below the widely accepted "replacement level" of 2.1 children per woman. (The U.S. rate is 1.8.) "Women are seeing the wisdom in having fewer children," says Poonam Muttreja, director of the nonprofit Population Foundation of India.

India's population growth is not over yet, however. Thanks to past high fertility rates, two-thirds of the population is under 35 years old, and a large cohort of people is now entering childbearing age. Even at replacement fertility rates, the children of these young people will continue to push up numbers, and India may exceed China as the world's most populous nation as early as next year.

Still, India's population is set to decline in about 3 decades, putting the country on the same track as a growing number of developing nations, such as its neighbor Bangladesh and Indonesia. India remains well behind China in falling fertility. In

A mother and her newborn in Chennai, India. With better health care in the nation, more babies today grow to adulthood.

China, where the population may be at its peak, official figures put the fertility rate at 1.7 children per woman.

State-sponsored family planning remains "the single most important driver" of India's drop in fertility, says Srinivas Goli, a demographer at Jawaharlal Nehru University. More than 55% of couples use modern contraceptives, the latest NFHS survey found. Of these, one-fifth use condoms and one-tenth the pill. But sterilization of women, generally in government-run clinics, accounts for two-thirds of all contraception.

Sterilization has a checkered past in India. During the mid-1970s, Gandhi allowed states to operate compulsory sterilization camps. An estimated 19 million people were sterilized, three-quarters of them men. The program's unpopularity helped bring down Gandhi's government in 1977, says Monica Das Gupta of the Maryland Population Research Center.

Das Gupta says sterilization was compulsory for only 2 years, after which the government relented and "never went back." Today, government sterilization clinics are chiefly aimed at women; just 0.5% of contraceptive use is male sterilization. On average, sterilization is performed on women in their mid-20s who have already borne children, says epidemiologist Tim Dyson of the London School of Economics.

Some states offer payments of up to $15 for women to be sterilized, and Das Gupta says "mindless allegations" of coercion still surround sterilization. But she argues financial incentives are not coercion, noting that such incentives "are routinely used across the world for innumerable purposes."

Dyson says declines in death rates in children under age 5—from 241 per thousand in 1960 to 34 per thousand today—have made women more receptive to family planning. That allows all families, including those who are poor and uneducated, to assume that every child is likely to grow up.

And education is playing an ever bigger role in encouraging smaller families. In the 1960s, about 90% of Indian women were illiterate, but by the 2011 census illiteracy had fallen to 35%, concentrated among older women. In the decade after 2005, better education contributed 47% of the fertility decline, according to an analysis by Milan Das, a research scholar at the International Institute for Population Sciences. "Women's aspirations have changed," Muttreja says. They look beyond the home for job opportunities and delay marriage and childbearing. "Education is the best contraceptive."

The strikingly different fertility rates in different regions of India reflect the role of education, says EM Sreejit of the International Planned Parenthood Federation in South Asia. Kerala in the south, which has the country's highest literacy rate, achieved replacement fertility back in 1988. Bihar in the east, with the lowest literacy rate, won't get there until 2039, the government's National Commission on Population predicted last year.

Some Indian politicians still talk of a population explosion and have proposed banning people with more than two children from government employment or, in Uttar Pradesh, even withholding welfare benefits. Critics say such rhetoric is often subtly aimed at the country's Muslim minority. Muslim women on average had 0.5 more children than Hindus, according to the 2015–16 NFHS survey.

But religion is a small factor in fertility today, Muttreja says. "Hindus in Uttar Pradesh, for example, have a much higher fertility than Muslims in Kerala. There is no Hindu fertility or Muslim fertility."

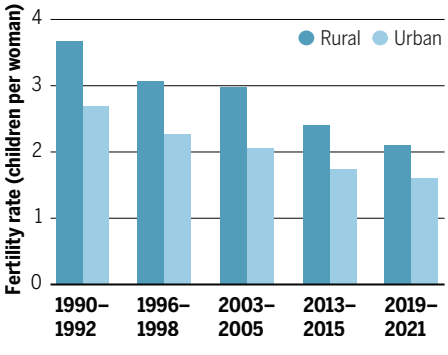
How low could India's fertility go? Goli says the Indian states that reached low fertility first have "stalled at 1.6 to 1.9 children per women." In contrast, "Highly educated couples in urban areas typically average 1.2 or 1.3," on par with the lowest rates seen in Europe and East Asia. "If the country continues to pursue women and girls' empowerment policies, this can push the country to … 1.4 or below," he says.

This would have a massive impact on the country's future size. The United Nations's 2019 population projections for India suggested it will rise to 1.64 billion by 2050 before falling to about 1.45 billion by century's end. But some demographers say a steep drop in fertility could drive a much faster decline and lead to an imploding economy.

That idea is controversial. Still, as India dips below replacement fertility, demographers are now arguing less about how scarily high its population might get and more about how scarily low it could go. ■

Steadily shrinking families

Rural women in India tend to have more children than urban women, but both groups have steadily lowered fertility rates.



Fred Pearce is a journalist in London.

CREDITS: (PHOTO) ARUN SANKAR/AFP/GETTY IMAGES; (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) NATIONAL FAMILY HEALTH SURVEY 4 2015–16/INTERNATIONAL INSTITUTE FOR POPULATION SCIENCES; NATIONAL FAMILY HEALTH SURVEY 5 2019–21/INTERNATIONAL INSTITUTE FOR POPULATION SCIENCES

In 2018, phytoplankton swirled in the Gulf of Finland. Some researchers want to trigger blooms with iron.

PHOTO: JOSHUA STEVENS AND LAUREN DAUPHIN/NASA EARTH OBSERVATORY

GEOENGINEERING

To draw down carbon, ocean fertilization gets another look

Panel urges large tests of iron-triggered plankton blooms

By **Warren Cornwall**

In January 2009, a German research ship set out for the Southern Ocean carrying 6 tons of iron and a boatload of controversy. The iron was meant to trigger a massive phytoplankton bloom that would suck carbon dioxide (CO_2) from the air, but environmentalists objected, viewing the trial as a reckless form of geoengineering. The German government briefly suspended the work, before letting it go ahead. It would be the last iron fertilization experiment for more than a decade.

But that could soon change, after a panel of leading ocean scientists last week said such experiments were a priority and called for the United States to spend up to $290 million on even larger ones that would spread 100 tons of iron across 1000 square kilometers of ocean. Already, researchers next year plan to pour iron across a patch of the Arabian Sea.

Rigorous tests of the strategy are critical, says Ken Buesseler, a biogeochemist at the Woods Hole Oceanographic Institution and a co-author of the National Academies of Sciences, Engineering, and Medicine (NASEM) panel report. "I think it is going to happen with or without the science," Buesseler says. "My fear is we see this commercialized before we know some of the fundamentals about the ocean response."

Even if nations make steep cuts to greenhouse gas emissions, many scientists believe that to prevent severe climate change, the world also needs to pursue "negative emissions technologies" that would pull CO_2 and other warming gases from the air. Billions of dollars have gone into land-based schemes that, for instance, promote reforestation or agricultural practices that store more carbon in the soil. But Scott Doney, a University of Virginia oceanographer who chaired the NASEM report, says when it comes to carbon sequestration research, "The ocean is a relatively new space."

The ocean has already absorbed nearly one-third of the carbon emissions from human activities, and scientists hope it can shoulder even more of the burden. Besides iron fertilization, the panel looked at rehabilitating coastal ecosystems; growing vast plantations of seaweed; and spurring plankton production by forcing nutrients up from deep in the ocean. Higher cost options included using electricity to strip CO_2 from seawater and inject it underground; and spreading pulverized rocks across the ocean to make it more alkaline, increasing the amount of CO_2 it can absorb.

Iron fertilization is among the cheapest options. Photosynthetic plankton act like tropical rainforests, sucking CO_2 from the atmosphere. Their populations are often limited by a scarcity of iron, which sifts into the ocean in windblown dust from deserts, in volcanic ash, and even from underwater hydrothermal vents. Extra iron would stimulate a bloom, the thinking goes, causing plankton to take up extra carbon. The carbon would sink into the depths in the form of dead plankton, or the feces or bodies of organisms that eat them. In theory, the carbon would be entombed for centuries.

Tests have shown the iron does stimulate plankton growth. But key questions remain, says Dave Siegel, a marine scientist at the University of California, Santa Barbara, who served on the NASEM panel. How much of the absorbed carbon makes it to the deep ocean is uncertain, he says: Other organisms might consume the sinking material and re-emit the carbon as CO_2. Another question on Siegel's mind: How would companies or governments track these carbon flows well enough to claim they are countering greenhouse gas pollution?

Buesseler is encouraged by recent computer modeling, published by Doney, Siegel, and colleagues in *Environmental Research Letters*, showing nearly one-third of the carbon captured near the ocean surface by events such as plankton blooms should sink to the deep ocean. Ocean-fertilization strategies could be viable "if we can get even 10% down deep enough," he says.

But skeptics note that a recent survey of 13 past fertilization experiments found only one that increased carbon levels deep in the ocean. That track record is one reason why making iron fertilization a research priority is "barking mad," says Wil Burns, an ocean law expert at Northwestern University.

Stephanie Henson, a marine biogeochemist at the United Kingdom's National Oceanography Centre, also worries about surprise consequences of the approach, likening it to the catastrophic introduction of rabbits to Australia ecology. "You could just imagine something like that happening in the oceans completely by accident." But Buesseler thinks gauging the potential risks is one reason to go ahead with the research.

David King, head of the Centre for Climate Repair at the University of Cambridge, is ready to test these politically charged waters. Next summer, working with scientists at India's Institute of Maritime Studies in Goa, he plans to spread iron-coated rice husks across a swath of the Arabian Sea, to learn whether suspending the nutrient for longer can spark a bloom with less iron.

To head off environmental concerns, King plans to confine the work within a giant plastic bag running from the surface to the sea floor several kilometers below. "There's an enormous amount of naysaying going on," King says. "There are many, many people saying let's leave the oceans alone, as if we haven't already interfered with them." ■

BIOMEDICINE

Blood condition may guard against Alzheimer's

Unexpected protection may result from wayward blood cells that enter brain

By **Mitch Leslie**

Medicine so far has nothing to offer that clearly prevents Alzheimer's disease, although keeping your weight down, exercising regularly, and inheriting certain beneficial genes can lower your risk. Now, a study has identified another, unexpected source of protection: clonal hematopoiesis, a blood cell imbalance best known as a risk factor for cancer and heart disease.

"Clonal hematopoiesis has been associated with so many bad outcomes that it is surprising that it is protective in this situation," says cardiovascular biologist Kenneth Walsh of the University of Virginia, who wasn't connected to the study, reported on 12 December at the American Society of Hematology meeting in Atlanta. But Walsh says the work is convincing and "will have to be reckoned with and explained."

He and other researchers caution that the discovery doesn't offer any immediate opportunities for treating or preventing Alzheimer's disease. Given the negative health effects of clonal hematopoiesis, inducing it in healthy people is a nonstarter. Still, the finding has a provocative implication: Cells from the bloodstream are restocking the brain's immune cells, perhaps bolstering its ability to clear out toxic debris.

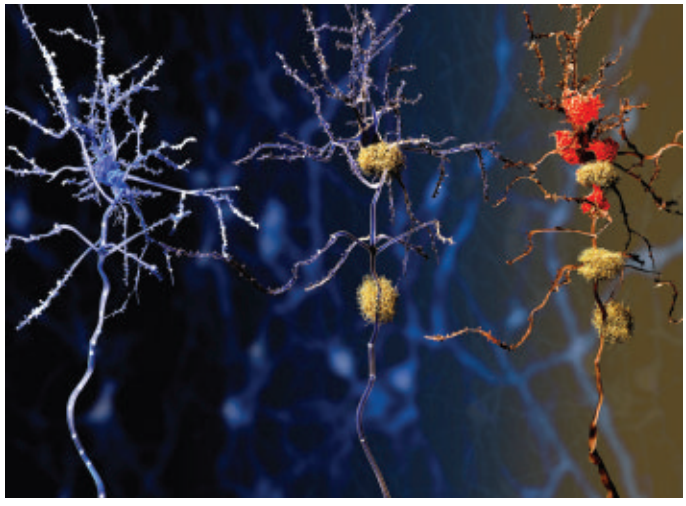
Charles Darwin probably never imagined that natural selection unfolds in our bone marrow. But clonal hematopoiesis results from competition among the 50,000 to 200,000 stem cells that dwell there and divide to produce all our red and white blood cells. Over the years these stem cells accrue mutations, some of which result in a "fitter" cell whose progeny, known collectively as a clone, can soon outnumber their counterparts. In some people with clonal hematopoiesis, the offspring of a single mutated stem cell account for more than half of the blood cells in the body.

This blood cell imbalance is rare in young people, but becomes more common with age, occurring in up to 30% of people over age 70. Clonal hematopoiesis is not cancer—people with the condition have a normal number of blood cells—but it can be a prelude to blood cancers such as leukemia and lymphoma. Affected people are also twice as likely to develop cardiovascular disease, and researchers have linked the condition to stroke, osteoporosis, type 2 diabetes, heart failure, and chronic obstructive pulmonary disorder. Overall, clonal hematopoiesis can boost a person's odds of dying within the next decade or so by up to 50%.

Pathologist Siddhartha Jaiswal of the Stanford University School of Medicine and colleagues wondered whether the skewed populations of immune cells in clonal hematopoiesis might also increase the risk of Alzheimer's disease. Other investigators, after all, had shown the brain's own immune cells, the microglia, may drive Alzheimer's disease. But any connection seemed a long shot. Cells from the blood clones might have to infiltrate the brain to have an effect—and most investigators think the microglia typically settle down in the brain early in life and are not born in the bone marrow.

When the scientists analyzed DNA sequences from the blood cells of more than 5700 people and compared the data with Alzheimer's incidence, they were stunned to see that clonal hematopoiesis has a substantial effect on developing the disease. But rather than increasing the risk, it appeared to make people 30% to 40% less likely to suffer from the illness. That's a larger protective effect than any other factor identified so far, Jaiswal says.

More evidence of a benefit came from postmortem brain samples of people in the DNA sequencing study who didn't have Alzheimer's disease when they died. Two molecular hallmarks of the illness—plaques of the protein beta amyloid and snarls of the protein tau—were less abundant in the brains of people with clonal hematopoiesis than in those without it, lead author Hind Bouzid, also at Stanford, reported at the meeting. She, Jaiswal, and their colleagues also revealed their findings in a preprint posted on medRxiv.



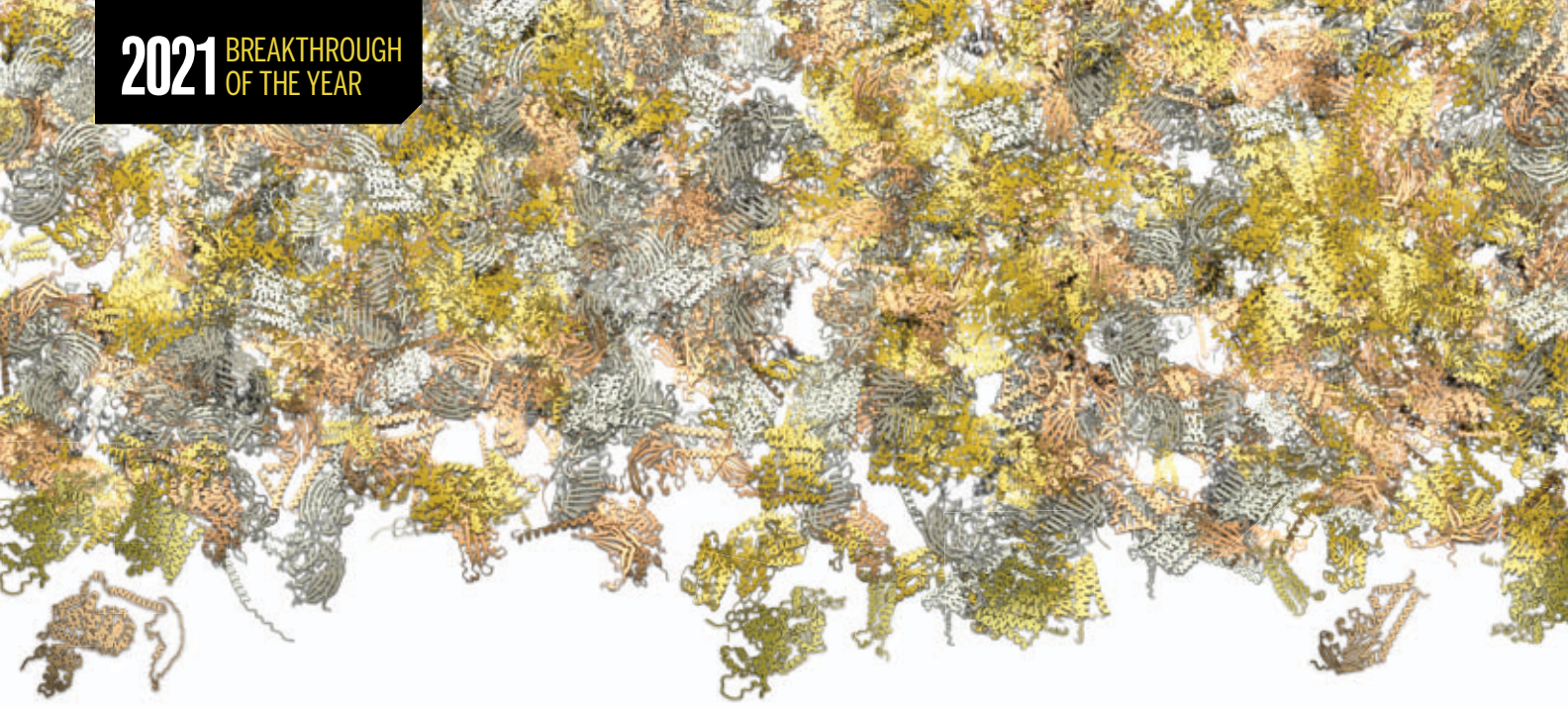
An illustration shows how microglia cells (red) can protect a healthy neuron (blue) by clearing out dead or dying neurons surrounded by clumps of amyloid (yellow). A blood disorder, clonal hematopoiesis, may boost the process.

IMAGE: JUAN GAERTNER/SCIENCE SOURCE

The team implicated microglia in the protective effect. They analyzed brain tissue samples from eight people with clonal hematopoiesis and in six found that unidentified cells carried the same mutations as the blood cell clones. In two other people with clonal hematopoiesis, they detected the blood mutations in the microglia themselves, finding the telltale changes in 40% to 80% of the cells. The researchers inferred that altered blood cells produced in the bone marrow had slipped into the subjects' brains. There, the cells may have morphed into microglia.

Immunologist Cameron McAlpine of the Icahn School of Medicine at Mount Sinai says the portion of the study linking clonal hematopoiesis to Alzheimer's disease is "beautiful." But he isn't convinced that bone marrow–derived blood cells regularly enter the brain. "They show it can occur, but I wonder to what degree it does occur."

If the work is confirmed, pinning down the protective mechanism is the next step in transforming the curious observation into something physicians might exploit. It's possible, for example, that mutant microglia are better at clearing away the beta-amyloid plaques and tau tangles that build up in Alzheimer's disease. But for now, Jaiswal says, how clonal hematopoiesis safeguards the brain remains a "million-dollar question." ■

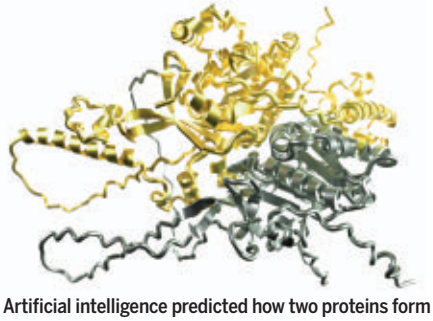


PROTEIN STRUCTURES

By **Robert F. Service**

In his 1972 Nobel Prize acceptance speech, American biochemist Christian Anfinsen laid out a vision: One day it would be possible, he said, to predict the 3D structure of any protein merely from its sequence of amino acid building blocks. With hundreds of thousands of proteins in the human body alone, such an advance would have vast applications, offering insights into basic biology and revealing promising new drug targets. Now, after nearly 50 years, researchers have shown that artificial intelligence (AI)-driven software can churn out accurate protein structures by the thousands—an advance that realizes Anfinsen's dream and is *Science*'s 2021 Breakthrough of the Year.

Protein structures could once be determined only through painstaking lab analyses. But they can now be calculated, quickly, for tens of thousands of proteins, and for complexes of interacting proteins. "This is a sea change for structural biology," says Gaetano Montelione, a structural biologist at Rensselaer Polytechnic Institute. David Baker, a University of Washington, Seattle, computational biochemist who led one of the prediction projects, adds that with the bounty of readily available structures, "All areas of computational and molecular biology will be transformed."



Artificial intelligence predicted how two proteins form a complex involved in DNA repair in yeast.

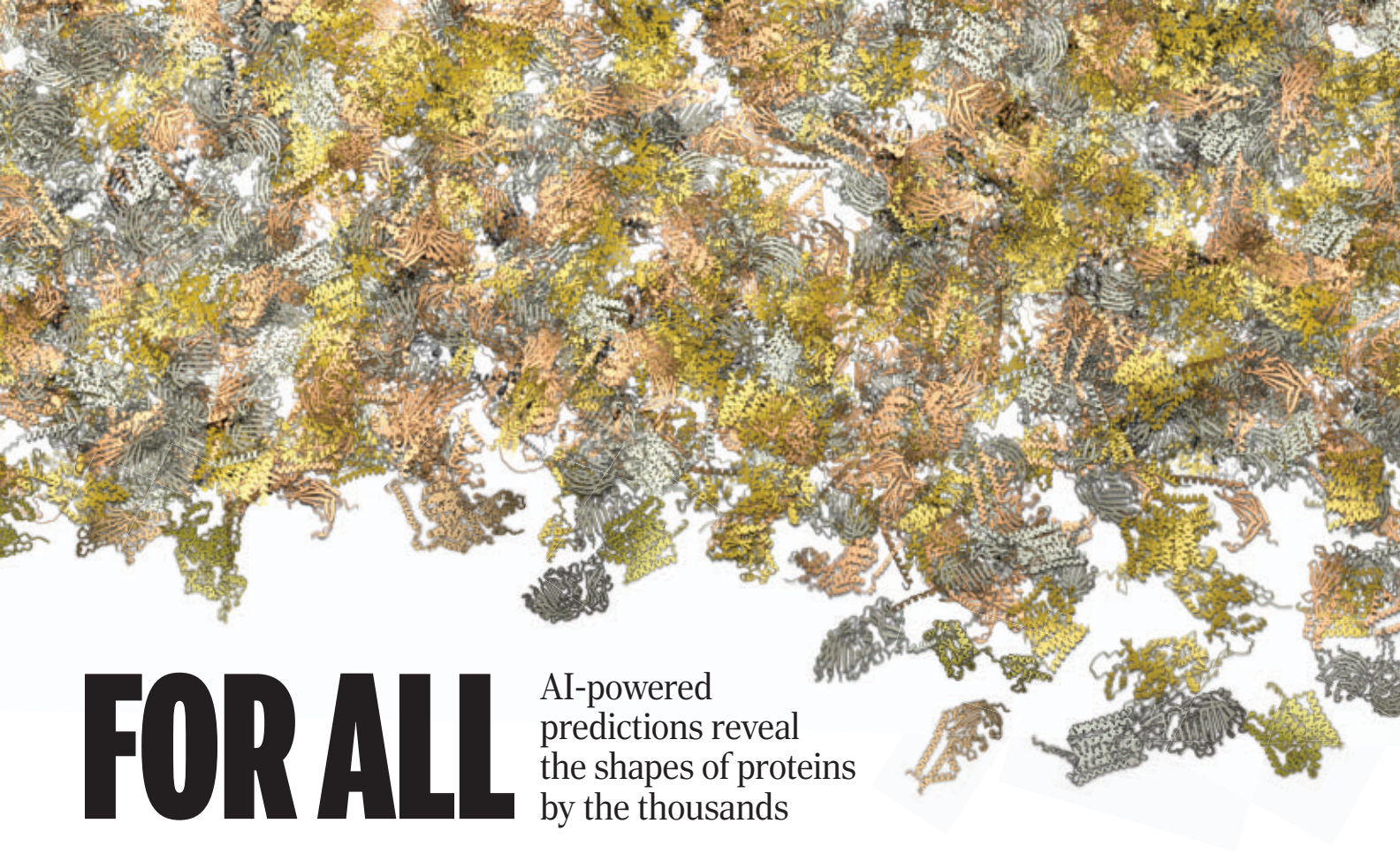
Proteins are biology's workhorses. They contract our muscles, convert food into cellular energy, ferry oxygen in our blood, and fight microbial invaders. Yet despite their varied talents, all proteins start out with the same basic form: a linear chain of up to 20 different kinds of amino acids, strung together in a sequence encoded in our DNA. After being assembled in cellular factories called ribosomes, each chain folds into a unique, exquisitely complex 3D shape. Those shapes, which determine how proteins interact with other molecules, define their roles in the cell.

Work by Anfinsen and others suggested interactions between amino acids pull proteins into their final shapes. But given the sheer number of possible interactions between each individual link in the chain and all the others, even modest-size proteins could assume an astronomical number of possible shapes. In 1969, American molecular biologist Cyrus Levinthal calculated that it would take longer than the age of a universe for a protein chain to cycle through them one by one—even at a furious pace. But in nature, each protein reliably folds up into just one distinctive shape, usually in the blink of an eye.

In the 1950s, researchers started to map proteins' 3D structures by analyzing how x-rays ricocheted off the molecules' atoms. This technique, known as x-ray crystallography, soon became the leading approach; today, the field's central repository, the Protein Data Bank, contains some 185,000 experimentally solved structures. But mapping structures can take years—and cost hundreds of thousands of dollars per protein. To speed the process, scientists started to create computer models in the 1970s to predict how a given protein would fold.

At first, that was possible only for small proteins or short segments of larger ones. By 1994, however, computer models had grown sophisticated enough to launch the biennial Critical Assessment of protein Structure Prediction (CASP) competition. Organizers gave modelers the amino acid sequences of dozens of proteins. At the end of the event, the modelers' results were compared with the latest experimental data from x-ray crystallography and emerging techniques such as nuclear magnetic resonance spectroscopy and cryo–electron microscopy (cryo-EM). Scores above 90 were considered on par with experimentally solved structures.

CREDITS: (ILLUSTRATION) V. ALTOUNIAN/*SCIENCE*; (DATA) I. R. HUMPHREYS *ET AL.*, *SCIENCE* **374**, 6573, EABM4805 (2021) DOI: 10.1126/SCIENCE.ABM4805



FOR ALL

AI-powered predictions reveal the shapes of proteins by the thousands

Early results were humbling, with median scores below 60. But over time, the modelers learned tricks to improve their calculations. For example, stretches of amino acids shared by two proteins often fold similarly. If a protein with an unknown structure shares, say, 50% of its amino acid sequence with a protein that does have a known structure, the latter can serve as a "template" to guide the computer models.

Another major insight came from evolution. Investigators realized that if one amino acid changed in a protein shared by closely related organisms, like chimpanzees and humans, amino acids located nearby in the folded molecule would have to change, too, to preserve the protein's shape and function. That means investigators can narrow down a protein's shape by looking for amino acids that coevolve: Even if they are far apart on the unfolded chain, they are likely neighbors in the final 3D structure.

By 2018, the modelers were often scoring in the mid-70s. Then, AlphaFold, an AI-driven software program, entered the scene. The program, developed by Google sister company DeepMind, trains itself on databases of experimentally solved structures. In its first competition, its median score was close to 80, and it won 43 of 90 matches against other algorithms. In 2020, its successor, AlphaFold2, shone even brighter. Powered by a network of 182 processors optimized for machine learning, AlphaFold2 rang up a median score of 92.4—on par with experimental techniques.

"I never thought I'd see this in my lifetime," John Moult, a structural biologist at the University of Maryland, Shady Grove, and CASP co-founder, said at the time.

This year, AI predictions shifted into overdrive. In mid-July, Baker and his colleagues reported that their AI program RoseTTAFold had solved the structures of hundreds of proteins, all from a class of common drug targets. A week later, DeepMind scientists reported they had done the same for 350,000 proteins found in the human body—44% of all known human proteins. In coming months, they expect their database will grow to 100 million proteins across all species, nearly half the total number believed to exist.

The next step is to predict which of those proteins work together and how they interact. DeepMind is already doing just that. In an October preprint, its scientists unveiled 4433 protein-protein complexes, revealing which proteins bind to one another—and how. In November, RoseTTAFold added another 912 complexes to the tally.

Code for AlphaFold2 and RoseTTAFold is now publicly available, helping other scientists jump into the game. In November, researchers in Germany and the United States used AlphaFold2 and cryo-EM to map the structure of the nuclear pore complex, an assembly of 30 different proteins that controls access to the cell nucleus. In August, Chinese researchers used AlphaFold2 to map the structures for nearly 200 proteins that bind to DNA, which could be involved in everything from DNA repair to gene expression. Last month, Google's parent company, Alphabet, launched a new venture that will use predicted protein structures to design new drug candidates. And Baker's team is using its software to dream up novel protein sequences that will fold into stable structures, an advance that could lead to new antivirals and catalysts.

Even now, scientists studying SARS-CoV-2 are using AlphaFold2 to model the effect of mutations in the Omicron variant's spike protein. By inserting larger amino acids into the protein, the mutations have changed its shape—perhaps enough to keep antibodies from binding to it and neutralizing the virus.

Much work remains. Protein structures aren't static; they bend and twist as they do their jobs, and modeling those changes remains a challenge. And it's still a daunting task to visualize most of the large, multiprotein complexes that carry out myriad jobs in cells. But this year's explosion of AI-driven advances offers a view of the dance of life as never seen before, a panorama that will forever change biology and medicine. ■

VIDEO AND PODCAST
The year in *Science* and our People's Choice award: science.org/boty2021

Ancient soil DNA comes of age

DNA from fossils has transformed the study of human and animal evolution, revealing unknown relationships, tracing early migrations, and exposing ancient interspecies mating. Yet for humans, the entire field depends on just 23 archaic genomes, 18 of them from Neanderthals. Recently, scientists unlocked a much larger trove of ancient DNA: from the soil of cave floors. This year, for the first time, cave dirt yielded DNA once housed in the nucleus of human cells, and researchers used such "dirt DNA" to reconstruct the identity of cave dwellers around the world.

The new work borrows from the study of environmental DNA from living species. To find out which organisms inhabit lakes, forests, and other places, scientists collect the free-floating DNA they shed into air, water, and soil. By 2003, evolutionary geneticists showed discarded DNA could persist for thousands of years. It was used by researchers in 2015 to help reconstruct entire ancient ecosystems, even in the absence of fossils.

But much of that DNA comes from mitochondria, the cell's power plants, which store tiny snippets of genetic material. Thanks to new techniques, scientists can now comb ancient soils for nuclear DNA, which carries the full instructions for life.

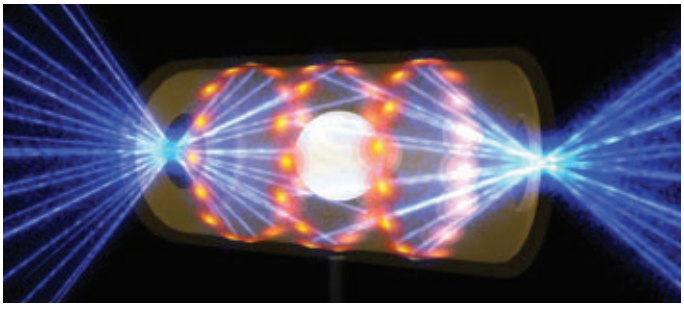
This year, scientists used nuclear DNA to chart the human and animal occupation of three caves. In Spain's Estatuas Cave, nuclear DNA revealed the genetic identity and sex of humans who lived there 80,000 to 113,000 years ago, and suggested one lineage of Neanderthals replaced several others after a glacial period that ended 100,000 years ago. In 25,000-year-old soil from Georgia's Satsurblia Cave, scientists found a female human genome from a previously unknown line of Neanderthals, along with the genetic traces of a bison and a now-extinct wolf. And by comparing 12,000-year-old black bear DNA from Mexico's Chiquihuite Cave with that of modern bears, scientists discovered that after the last ice age, the cave bears' descendants migrated as far north as Alaska.

Techniques for extracting and sequencing nuclear DNA from ancient soils are still improving. As they do, researchers hope to answer even more questions about the rise and fall of ancient species. *—Elizabeth Pennisi*

Fusion's day in the Sun?

Is fusion energy about to overcome its reputation as a field that promises the stars but never delivers? In an August result that surprised the researchers themselves, the U.S. National Ignition Facility (NIF) produced a fusion reaction that came tantalizingly close to reaching official "breakeven," the point at which a reaction produces more energy than the laser energy needed to kindle it.

Fusion, which powers the Sun and other stars, has long been seen as a solution to Earth's energy problems. But achieving the pressures and temperatures required—10 times as hot as the Sun's core—is notoriously difficult. Many efforts cage a superhot plasma in a magnetic field; NIF uses a pulse from the world's highest energy laser to compress a peppercorn-size capsule of the hydrogen isotopes deuterium and tritium. Earlier this year, that method generated 170 kilojoules of fusion energy per shot—far short of the laser input of 1.9 megajoules.



To produce NIF's fusion shot, 192 laser beams converged around a tiny fuel pellet.

But in an 8 August shot, that yield surged to 1.35 megajoules. Researchers think it's the result of a burning plasma, meaning the fusion reaction generated enough heat to spread through the compressed fuel like a flame. The result has not been scrutinized by a peer-reviewed journal, but it was presented at the American Physical Society's Division of Plasma Physics conference in November.

Now, the team is trying to understand the shot's high yield and figure out how to tweak starting conditions to do even better, by using larger or smoother fuel capsules, more even layers of frozen fuel, or higher quality laser pulses. They're also making efforts to replicate the shot: One attempt in October reached 430 kilojoules, and a November shot hit 700 kilojoules. Attempts will continue into 2022.

As NIF edges toward breakeven, private fusion projects are upping the pace. Several predict they will generate energy long before the ITER reactor, a $25 billion publicly funded magnetic fusion effort. This year, Commonwealth Fusion Systems and Tokamak Energy claimed progress with high-temperature superconducting magnets. And General Fusion and TAE Technologies, which use pistons and particle beams, respectively, are planning energy-producing demonstration power plants they say will switch on in 2025.

Whichever approach reaches energy gain first, formidable challenges in materials science and engineering remain before fusion can become a practical power source. *—Daniel Clery*

IMAGES: (TOP TO BOTTOM) DEVLIN A. GANDY; LAWRENCE LIVERMORE NATIONAL LABORATORY

A researcher records the locations of sediment samples in Mexico's Chiquihuite Cave.

PEOPLE'S CHOICE

Our readers weigh in with their picks for the top breakthrough this year

For the first time since 2018, your winner was our winner! This year's race for the People's Choice award—an annual honor chosen by *Science* readers—was tight, with three finalists running neck and neck in the last week of voting on Twitter: ancient soil DNA, CRISPR gene editing in the body, and artificial intelligence–powered protein structure predictions.

But after a final surge of support for ancient soil DNA and more than 2100 votes cast, AI-powered protein structure predictions—*Science*'s 2021 Breakthrough of the Year—kicked it in to capture the gold.

This year's breakthrough was itself predicted in 2020, when it appeared as a runner-up. Since then, the field has exploded, with AI able to predict not only protein structures, but also how they form complexes and interact.

1. Protein structure predictions **39%**
2. Ancient soil DNA **34%**
3. In vivo CRISPR **27%**

Potent pills boost COVID-19 arsenal

Vaccines have had a starring role in the fight against COVID-19, but a new player is joining them on stage: antiviral pills that prevent symptoms and death if taken early in infection.

In the fall, drugmakers Pfizer and Merck & Co. touted positive clinical trial results via press release. More antivirals are in trials, and existing generic drugs—including the obsessive-compulsive disorder treatment fluvoxamine—may also prove useful. Merck's antiviral, molnupiravir, reduces the risk of hospitalization or death by 30% in high-risk, unvaccinated individuals, according to final data submitted to regulators. (That figure is lower than the 50% from an interim analysis.) Pfizer's antiviral, PF-07321332, reduces hospitalization by 89% if started within 3 days of symptoms. The United Kingdom approved molnupiravir in November, and an advisory body to the U.S. Food and Drug Administration has narrowly endorsed it; regulators are also considering Pfizer's treatment.

Scientists are quick to stress that antivirals can't replace vaccination. But they are still vital, and may become even more so if the new Omicron variant causes a surge in breakthrough infections. Pfizer is running another PF-07321332 trial in a group that includes vaccinated people, and both Merck and Pfizer are testing whether the drugs can head off illness in people who have recently been exposed.

Still, questions abound. Will the antivirals reduce transmission from infected people? Can lower income countries access the supply they need? (Both companies have pledged to sell the drugs in these places at a steep discount.) Will there be side effects not seen in clinical trials? Despite uncertainties, scientists and doctors are buoyed by the results—and pleased to have multiple therapies, which may help keep the virus from becoming resistant to any one. —*Jennifer Couzin-Frankel*

Merck's molnupiravir cut the risk of hospitalization or death from COVID-19 by 30% in high-risk, unvaccinated individuals.

PHOTO: © 2009-2021 MERCK SHARP & DOHME CORP., A SUBSIDIARY OF MERCK & CO., INC., KENILWORTH, N.J., U.S.A. ALL RIGHTS RESERVED

A psychedelic PTSD remedy

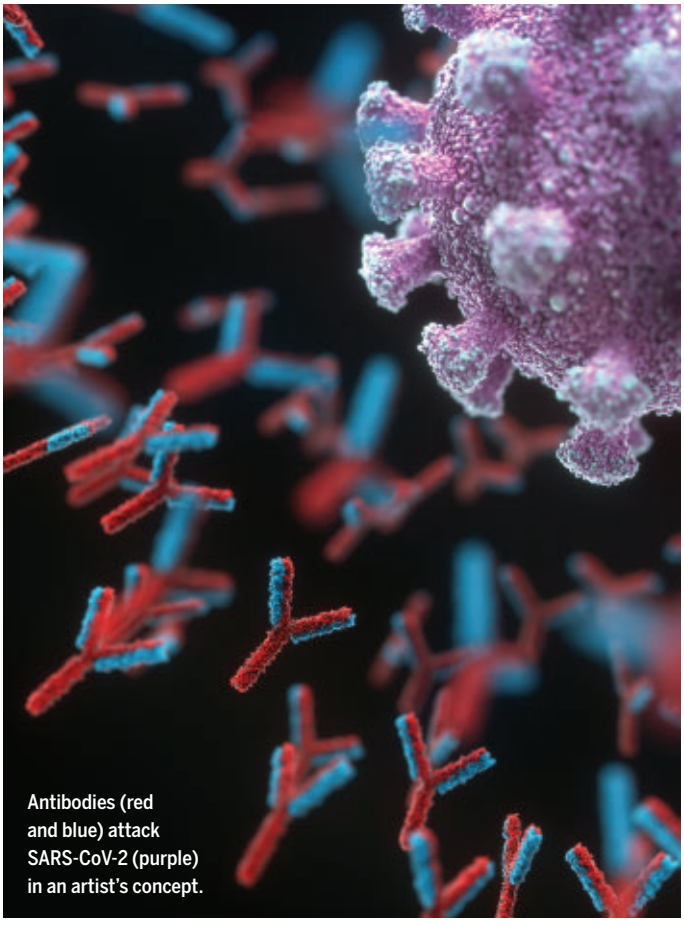
The mind-altering power of psychedelic drugs has raised hopes that they can ease psychiatric disease, but few large, rigorous trials have shown they're effective. This year brought a big win for the field: A multicenter, randomized, controlled trial found that 3,4-methylenedioxymethamphetamine (MDMA), popularly called ecstasy, significantly reduced symptoms in patients with post-traumatic stress disorder (PTSD).

The study, published in *Nature Medicine* in May, tested an intensive combination of talk therapy and MDMA, which can create a sense of well-being and empathy that may help people process traumatic experiences. The trial's 76 participants had three 8-hour guided therapy sessions with either the drug or a placebo, plus a set of shorter "preparatory" and "integration" sessions with therapists before and after treatment. After 2 months, 67% of those who got MDMA no longer met the diagnostic criteria for PTSD, versus 32% in the placebo group.

The results prompted enthusiasm, but also caution. Such trials face a "Gordian knot of blinding and placebo effects," neurologist Matthew Burke and psychiatrist Daniel Blumberger of the University of Toronto warned in *Nature Medicine* in October. That's because MDMA's psychoactive effects are obvious to participants and could influence their expectations—and even the odds they will improve. (Follow-ups with participants after the study suggested as many as 90% correctly guessed which group they were in.) And simply accepting that such expectations are part of the treatment "would require a complete overhaul of how we measure efficacy in psychiatry," Burke and Blumberger write.

MDMA treatment for post-traumatic stress disorder includes trained supervision.

Still, psychedelic research is booming as academic labs and companies explore the potential of MDMA and other psychedelics to treat conditions like depression, anxiety, and addiction. In November, the London-based mental health care company COMPASS Pathways announced positive results from a 233-participant randomized trial of psilocybin, the substance in so-called magic mushrooms, in people with treatment-resistant depression. The company is now planning a larger trial. And if an ongoing follow-up study can confirm the initial MDMA results, its sponsor—the nonprofit Multidisciplinary Association for Psychedelic Studies—plans to seek approval from the U.S. Food and Drug Administration as early as 2023. —*Kelly Servick*



Antibodies (red and blue) attack SARS-CoV-2 (purple) in an artist's concept.

Artificial antibodies tame infectious diseases

Labmade antibodies called monoclonals have revolutionized the treatment of some cancers and autoimmune diseases, but they've had limited success against infectious diseases. That changed this year, as monoclonal antibodies (mAbs) made inroads against SARS-CoV-2 and other life-threatening pathogens, including respiratory syncytial virus (RSV), HIV, and malaria parasites.

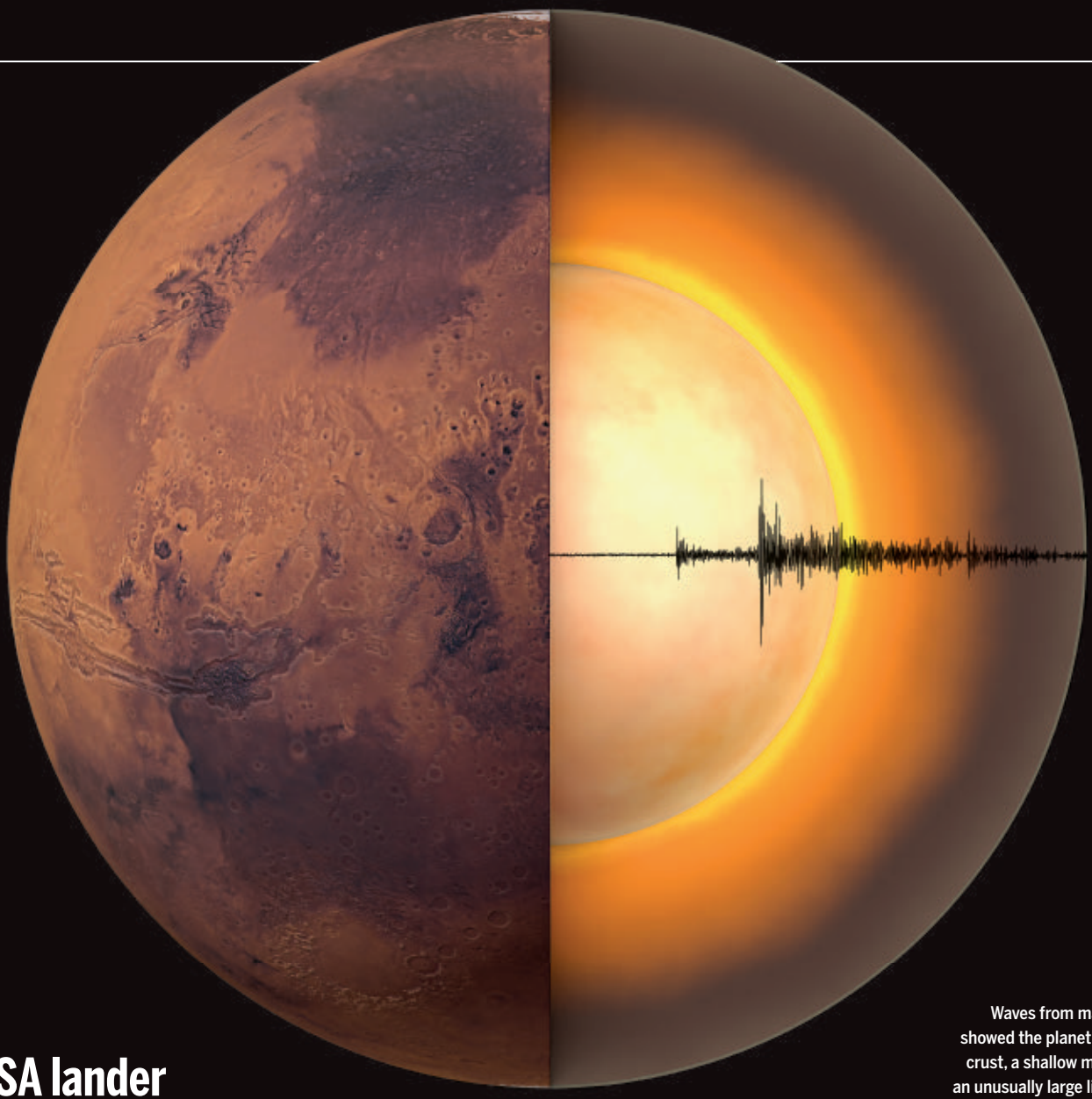
To make mAbs, scientists isolate the most powerful antibodies from lab animals and humans and reproduce them in massive quantities. As medicines, they are mostly used to tamp down immune responses or mark tumor cells for destruction. The only mAbs approved for infectious diseases in the United States are limited to rare maladies: Ebola, inhalational anthrax, recurrent *Clostridium difficile*, RSV in high-risk infants, and HIV in people for whom all drugs have failed. India has approved a mAb for rabies.

With advances in cloning, animal models, and x-ray crystallography, researchers can now make and screen more mAbs than ever before, simplifying their search. SARS-CoV-2 mAbs showed promising results in clinical trials in 2020, and by late this year, the U.S. Food and Drug Administration had granted emergency use authorization to three to treat COVID-19 and, in some cases, prevent infection.

Monoclonals are also being developed against influenza, Zika, and cytomegalovirus. High hopes surround two candidates designed to prevent RSV in all infants. And in one HIV prevention study that failed overall, the candidate worked well against some strains.

High costs and the need to infuse mAbs in a clinic have put them out of reach for many. But as prices plummet, injections replace infusions, and more potent mAbs come to market, they may become standard weapons in the infectious disease arsenal. —*Jon Cohen*

IMAGES: (TOP TO BOTTOM) MULTIDISCIPLINARY ASSOCIATION FOR PSYCHEDELIC STUDIES; KTSDESIGN/SCIENCE SOURCE



Waves from marsquakes showed the planet has a thin crust, a shallow mantle, and an unusually large liquid core.

ILLUSTRATION: C. BICKEL/*SCIENCE*

NASA lander uncovers the Red Planet's core

The interior of a rocky planet is a kind of time machine: Its dense core, viscous mantle, and hardened crust can reveal how it coalesced, churned, and settled into what it is today. Until this year, scientists have had access to just two such time capsules: Earth and, briefly during the Apollo missions, the Moon. Now, for the first time, instruments aboard NASA's InSight lander are bringing Mars's planetary core into focus.

When InSight arrived on the Red Planet in 2018, Mars seemed reluctant to expose its buried secrets. InSight's heat probe failed to penetrate the planet's surprisingly sticky sediments, despite repeated attempts. The lander's hypersensitive seismic station, designed to monitor the underground rumblings that could help chart the planet's interior, never picked up a marsquake powerful enough to do the job. And dust piled on the lander's solar panels, slowly eroding their power output.

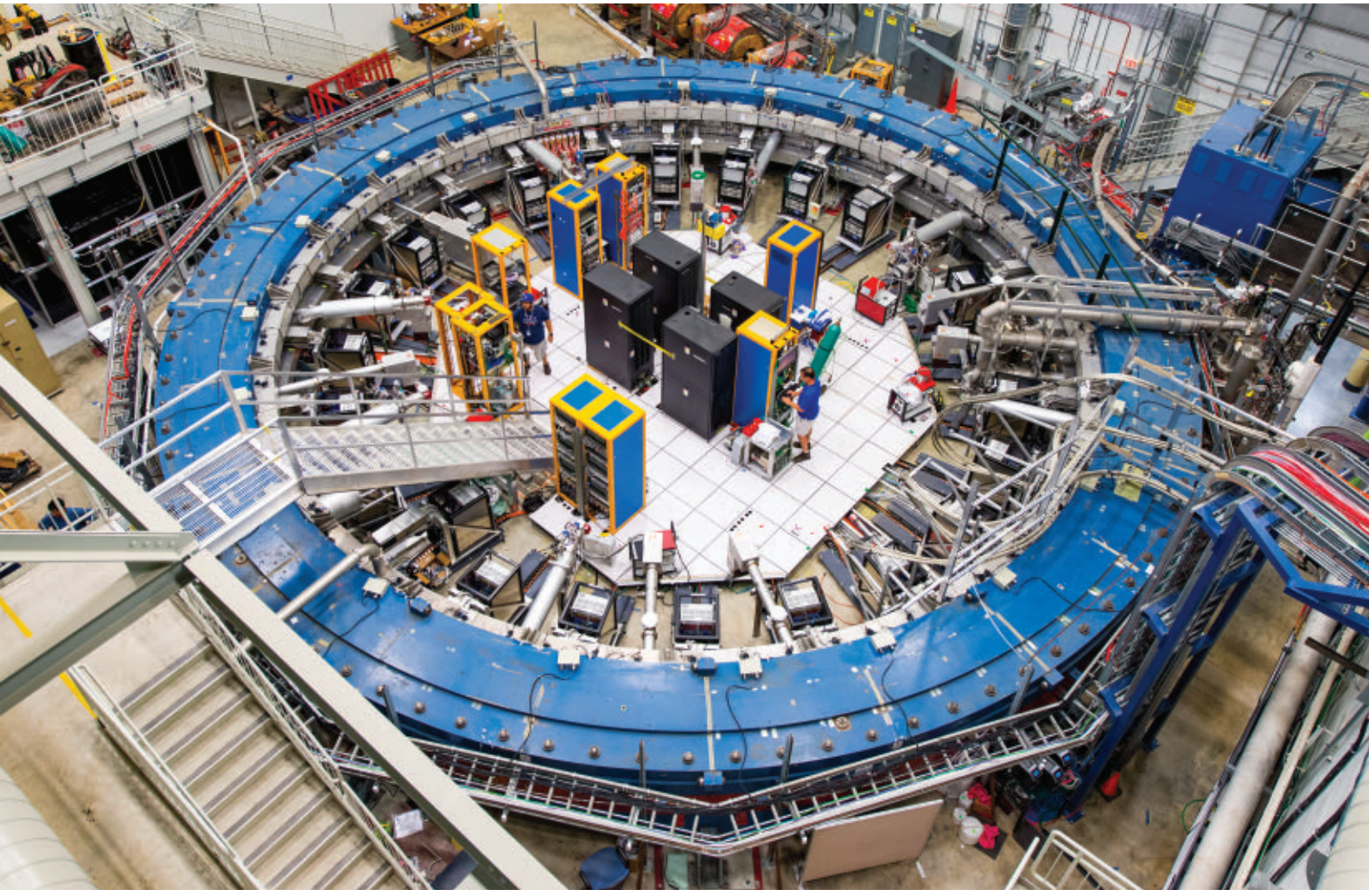
But within 1 year, InSight picked up a handful of moderate quakes, including several stemming from Cerberus Fossae, a fissured region 1600 kilometers away. When combined with estimates of the interior's composition, those readings helped chart the planet's depths. Offsets in the quakes' seismic waves revealed that the martian crust is layered and less than 40 kilometers thick—thinner than Earth's continental crust. That thin shell would have let Mars quickly shed its early internal heat.

Looking deeper, InSight found the martian mantle lacked the insulating lower layer seen in Earth's. The mantle was also shallow, squeezed between the crust and an unusually large, liquid core that occupies more than half of Mars's width. Given the planet's mass, scientists concluded that the core's density is low, and that a mixture of light elements such as sulfur likely keep its iron and nickel liquid, despite the planet's rapid heat loss—much as salt prevents icing. The researchers published their findings in *Science* this year.

Armed with these new data, scientists will be puzzling over Mars's history for years to come. Questions include whether the planet once had something resembling plate tectonics and when its liquid core stopped churning, shutting off the magnetic field it once generated.

InSight may have still more stories to share: In August and September, the lander heard its largest marsquakes yet. But as red dust continues to build on its solar panels, its time is growing short. It is estimated to run out of power by the end of 2022.

Until then, InSight will wait, and listen. —*Paul Voosen*



Within this ring at the Fermi National Accelerator Laboratory, muons twirl like compass needles in a magnetic field mapped with an accuracy of 30 parts per billion.

PHOTO: REIDAR HAHN/FERMI NATIONAL ACCELERATOR LABORATORY

At last, a crack in particle physics' standard model?

It may be a sign of particle physicists' desperation for something new that the biggest result in years confirms an oddity first observed 2 decades ago. A particle called the muon—a heavier, unstable cousin of the electron—is slightly more magnetic than physicists' prevailing theory of fundamental particles and forces predicts. Reported in April, the 2.5-parts-per-billion discrepancy could signal new particles lurking just over the high energy horizon.

Developed in the 1960s and '70s, the current theory, known as the standard model, accounts for three forces—electromagnetism, the strong nuclear force, and the weak nuclear force—and two dozen fundamental particles. It cannot be the final description of nature, as it leaves out both gravity and dark matter, the mysterious stuff thought to outweigh the universe's ordinary matter. Yet, so far, the standard model accounts for every particle blasted into fleeting existence with high energy particle accelerators.

The muon's magnetism gives scientists an indirect way to search for additional, undiscovered particles. Thanks to quantum uncertainty, empty space around the muon roils with particle-antiparticle pairs popping in and out of "virtual" existence too fast to be directly observed. Those in the standard model increase the muon's magnetism by a precise amount. New particles could change that calculation in unpredicted ways.

To measure muon magnetism, scientists fire a beam of them into a magnetic field, where they twirl like compass needles at a rate that depends on how magnetic they are. Physicists ran just such an experiment from 1997 to 2001 at Brookhaven National Laboratory in New York, where they first detected the anomaly. In 2003, they hauled their 15-meter-wide magnet to Fermi National Accelerator Laboratory in Illinois, to obtain a purer muon beam. This year, they proved their previous result was not a fluke.

But a proper comparison depends on the precision of the standard model prediction. On the same day experimenters released their result, one team of theorists published a calculation that, they argued, increases the standard model prediction and closes the observed gap. Other physicists say the theoretical consensus still indicates that the muon is extra magnetic.

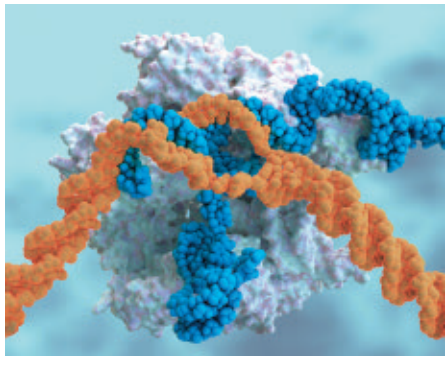
Now, the question is why. Other searches for tiny discrepancies from the standard model's predictions could yield more clues to the hoped-for new physics. Or, if physicists are lucky, the world's biggest atom smasher, Europe's Large Hadron Collider, will blast some new particle into plain view when it comes back online next spring after 3 years of upgrades. —*Adrian Cho*

CRISPR fixes genes inside the body

The gene-editing tool CRISPR had its first clinical victory in 2020, when it appeared to cure people with two inherited blood disorders, sickle cell disease and beta-thalassemia. Those treatments took place in a lab dish: Scientists removed defective blood stem cells from patients, edited them, and reinfused the cells into patients. This year, scientists took things one step further, deploying CRISPR directly in the body. In small studies, the strategy reduced a toxic liver protein and modestly improved vision in people with inherited blindness.

Gene editing could tackle many more diseases if the therapy could be injected into an organ or the bloodstream. But getting CRISPR to work inside a person, or in vivo, poses significant challenges. Before CRISPR's molecular components can correctly modify a specific gene, they must be ferried safely to the right cells in the right quantities.

To fight hereditary transthyretin (TTR) amyloidosis, a disease in which a misfolded TTR protein builds up and damages nerves and the heart, researchers at Intellia



Guide RNA (blue) from a CRISPR injection leads a DNA-cutting enzyme (white) to its target (orange).

Therapeutics and Regeneron Pharmaceuticals gave six patients an infusion of tiny fat balls encasing a guide RNA and the RNA instructions for CRISPR's genome-snipping enzyme. The team hoped the patients' own liver cells would take up the particles and make the CRISPR components, which would snip both strands of DNA at the TTR gene. The cell's repair system would mend the cuts imperfectly, leaving the gene disabled. It worked: After 4 weeks, average blood levels of TTR dropped 52% or 87% depending on the dose, researchers reported in June in *The New England Journal of Medicine*.

It will take many months to learn whether the drop in TTR eases symptoms. The hope is that the one-time treatment will work as well as, if not better than, an RNA-based drug that must be injected every 3 weeks.

In another study, researchers at Editas Medicine injected a harmless virus carrying CRISPR DNA into the eyes of six adults with an inherited vision disorder called Leber congenital amaurosis 10. The scientists hoped to snip out extra DNA that disrupted a mutated eye gene so cells would then make its missing protein. After 3 to 6 months, two patients—who had been almost completely blind—could sense more light, and one could navigate an obstacle course in dim light, the researchers reported at a September meeting. They hope for greater vision gains in adults receiving a higher CRISPR dose, and in young patients. —*Jocelyn Kaiser*

Embryo 'husbandry' opens windows into early development

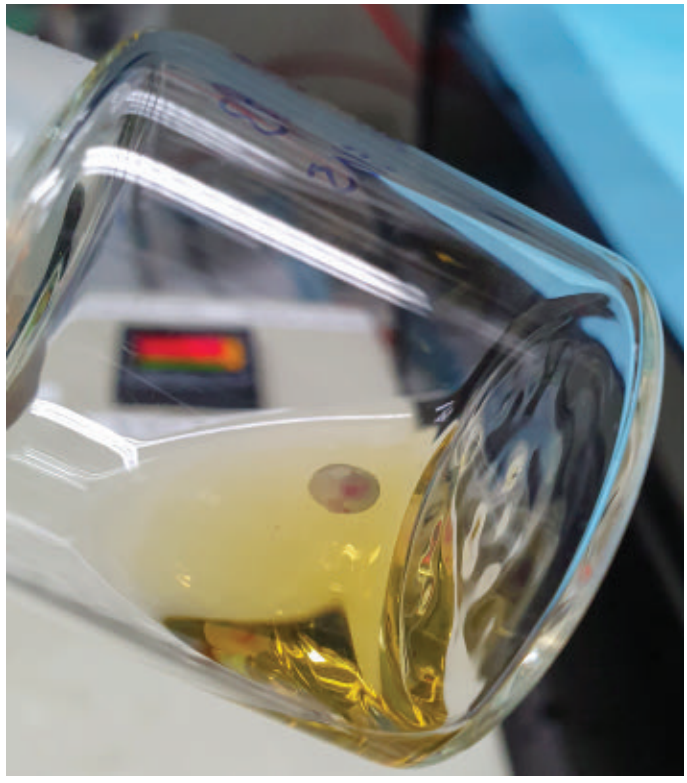
Insights into early embryonic development can help scientists understand miscarriages and birth defects—and hone in vitro fertilization (IVF) protocols. But legal, practical, and ethical limitations constrain studies with human embryos. This year, scientists unveiled potential stand-ins: mouse embryos reared far longer than before, and embryo replicas made from human stem cells or reprogrammed adult cells.

Scientists have struggled to grow mouse embryos outside a mother mouse's body for much longer than 3 or 4 days. But in March, one team reported a recipe for stretching that to 11 days. A key step, they found, was rotating the jars containing the embryos on a device that resembles a miniature Ferris wheel. It continually mixes the nutrient broth that bathes the embryos and ensures that oxygen levels and atmospheric pressure are congenial. The embryos underwent a key stage of cellular reorganization, grew organs, and sprouted hind legs.

Other scientists devised substitutes for a crucial embryonic stage known as the blastocyst. A hollow ball harboring only a few hundred cells, the blastocyst implants into the uterus and is the first embryonic stage to feature specialized cells. It's also inserted by many IVF clinics into prospective mothers.

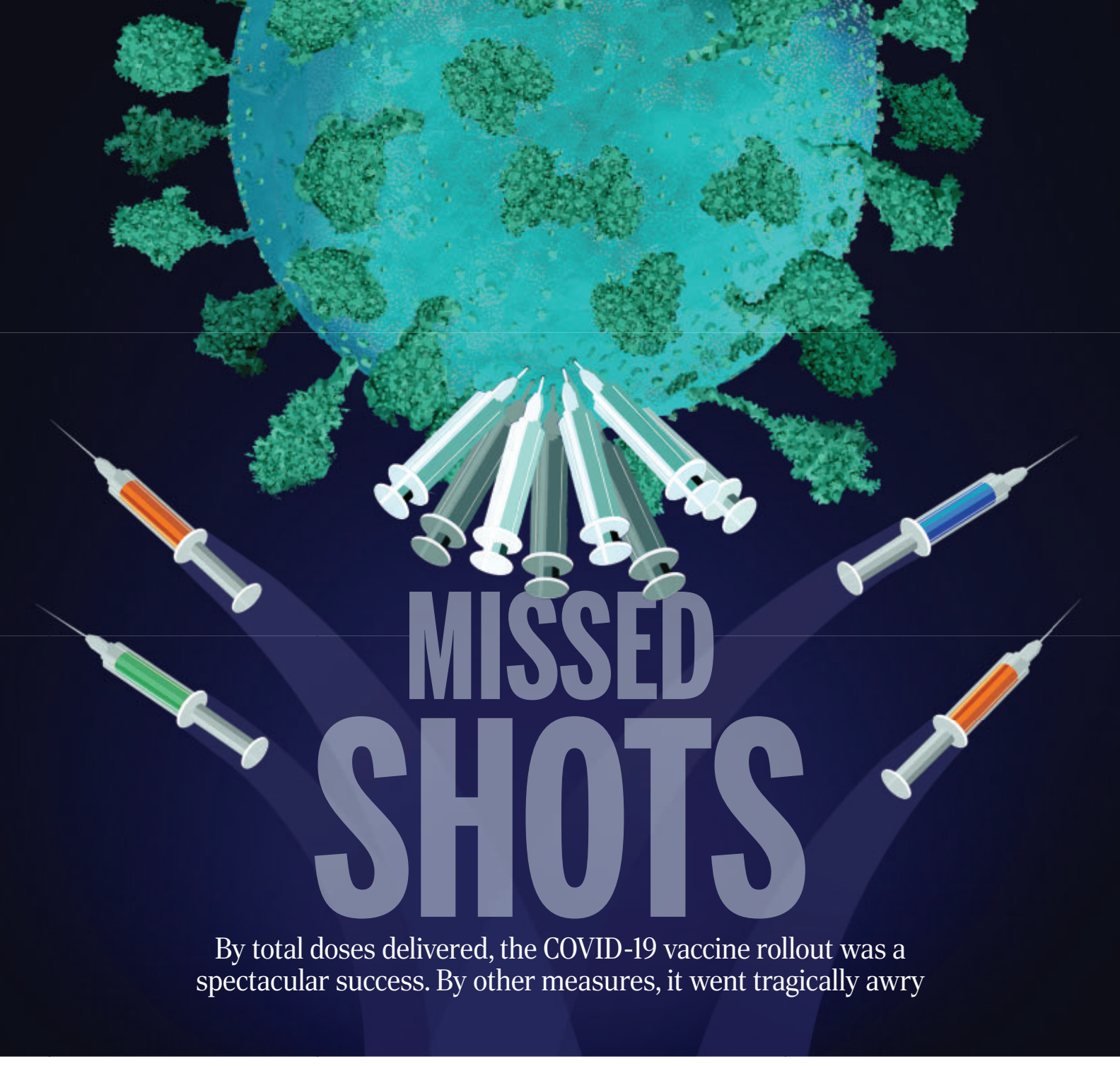
One team made blastocyst replicas from human embryonic stem cells and induced pluripotent stem (iPS) cells, stem cells reprogrammed from specialized adult cells. Another discovered that skin cells undergoing the transition to iPS cells produce blastocystlike structures. These ersatz blastocysts aren't real embryos, but some of them could offer an instructive—and less controversial—alternative.

The field received another boost in May. The international organization that sets stem cell research guidelines relaxed its long-standing prohibition on growing human embryos in the lab for more than 14 days, allowing scientists to probe embryonic events that occur after that time. —*Mitch Leslie*



A mouse embryo grows in a rotating jar. Such embryos can help researchers better understand the early stages of human development.

IMAGES: (TOP TO BOTTOM) ELLA MARU STUDIO/SCIENCE SOURCE; WEIZMANN INSTITUTE OF SCIENCE



MISSED SHOTS

By total doses delivered, the COVID-19 vaccine rollout was a spectacular success. By other measures, it went tragically awry

By **Kai Kupferschmidt**

When *Science* crowned the development of effective SARS-CoV-2 vaccines the scientific Breakthrough of the Year in December 2020, it was a moment of celebration. "This breakthrough is a triumph for all of science," Editor-in-Chief Holden Thorp wrote in an editorial. "There will be plenty of time for an exegesis of what went wrong. But for now, what went right is far more important."

One year later, what went right and what went wrong with COVID-19 vaccines cannot be so neatly divided.

The vaccines were developed at astonishing speed, and their efficacy in large-scale clinical trials—95% protection from symptomatic illness for the messenger RNA vaccines—surpassed most scientists' dreams. But once they entered the real world, things got messy. The production and distribution of billions of shots posed huge logistical obstacles. The stellar protection began to wane. The virus proved highly adaptable, with new variants appearing in quick succession. And the shining scientific triumph entered the murkier realm of politics and patent rules, commerce and conspiracy theories. The result has been a muddle—half triumph, half tragedy.

More than 8 billion shots have now been given worldwide—easily enough to fully vaccinate everyone on most priority lists: people who are over age 60, work in health care, or suffer from underlying conditions that can worsen COVID-19. That is not

ILLUSTRATION: ADAM SIMPSON/HEART AGENCY

what has happened. Teenagers in some countries have received three shots; people elsewhere with a much higher risk of dying are still waiting for their first.

Millions are alive today because they got vaccinated, but millions of others have declined the offer of a safe, free vaccine. Hundreds of thousands died needlessly. And any hopes that the vaccines might curb transmission enough to stop the virus from spreading have proved ill-founded. As this year draws to a close, evidence is mounting that the Omicron variant may further erode vaccines' protection. "It is a little depressing to be here in December 2021 and just still feeling like we have an uphill battle," says Natalie Dean, a biostatistician at Emory University.

FOR ANY DRUG or vaccine, expectations help shape whether it is seen as a success. Early on, scientists worried COVID-19 shots might only prevent 50% of cases. By that standard, vaccines vastly overperformed. But many in the public expected something like the measles or rubella vaccines, which offer complete, lifetime protection from infection—an unrealistic hope. "In our excitement of having a vaccine and having one quickly, we forgot that not everyone in the general public is a vaccinologist or virologist," says Boghuma Titanji, a virologist at Emory.

Early, hopeful speculation that widespread vaccination would create herd immunity—when so many people are protected that the virus has nowhere to go—compounded the disappointment later. Herd immunity, always an ambitious goal, slipped out of reach as the more infectious Delta variant spread around the world and the shots' protection waned.

Overall, the vaccines have proved very safe. But in the spring, a very rare but potentially deadly clotting disorder emerged in people who had received vaccines produced by AstraZeneca and Johnson & Johnson, which are based on an adenovirus vector. Although the risk-benefit ratio was still very good, many rich countries stopped using the shots once they had alternatives, which damaged confidence in vectored vaccines elsewhere, especially in poorer countries that received large shipments of the AstraZeneca vaccine. "The perception … has been that Africa was given subpar vaccines," says Titanji, who had to convince her own parents in Cameroon to get the shot.

Cynical peddlers of half-truths and lies about the risks of vaccines and the promise of unorthodox remedies exploited the confusion, aided by an information ecosystem that prioritizes "engagement" over veracity and politicians who would rather put their voters' lives at risk than risk acknowledging a complex truth. In the end, many people felt more comfortable swallowing an unproven drug used to deworm horses or taking their chance with a deadly virus than getting a vaccine that had been shown to protect the vast majority of people against severe disease and death.

Nonbiomedical research could have helped: studies of vaccine hesitancy, the ways people make medical decisions, and how misinformation spreads and can be fought. "That is science, too, and we do not pay enough attention to it," Dean says. Public health agencies neglected to approach those most vulnerable to misinformation before they were bombarded with lies, Titanji says.

> **"It is a little depressing to be here in December 2021 and just still feeling like we have an uphill battle."**
> **Natalie Dean**, Emory University

MANY PEOPLE have not had the luxury of choosing whether to get the shots. Only 8% of Africa's population is fully vaccinated. "You had this unseemly scramble with the rich countries paying whatever they could to get hold of what they wanted, and the rest pushed to the back of the queue," says Helen Clark, co-chair of the Independent Panel for Pandemic Preparedness and Response, which was established by the World Health Organization (WHO) to draw lessons from the pandemic.

WHO and other groups in 2020 formed the COVID-19 Vaccines Global Access (COVAX) Facility to ensure a more equitable distribution. But its approach—buying vaccines in bulk and providing them for free to poor countries—was flawed, Clark says. Rich countries were happy to donate money but cornered the market for vaccines, leaving little for COVAX. And the scheme's main supplier, the Serum Institute of India, provided far fewer doses than promised this year. "We would have been even worse off without COVAX," Clark says. But its modest track record—some 650 million doses shipped to low- and middle-income countries so far—"exposed the limit of optimism in multilateralism," says John Nkengasong, director of the Africa Centres for Disease Control and Prevention.

Targeting those most at risk across the globe would not only have been more just, but also would have averted more deaths, says Maria Van Kerkhove, a top WHO epidemiologist. And if rich countries had fewer doses, they might have followed WHO's advice not to relax public health measures such as mask wearing and limits on gatherings, which could have blunted recent surges. "Particularly in those countries with access to the vaccines, there was a sense that the worst is over," says Jeremy Farrar, head of the Wellcome Trust. That sense also eased the pressure to ramp up vaccine production and deliver more doses to more places fast. As WHO Director-General Tedros Adhanom Ghebreyesus recently put it: "In too many countries, the bright light of vaccines has also become a blinding light for the continued need to stop this virus from spreading."

THE PANDEMIC IS RAGING ON. It's back with a vengeance in the Northern Hemisphere, even in countries with high vaccination coverage. So far, vaccines have held up well against new variants, from Alpha to Delta. But the virus' continuing spread gives it every opportunity to evolve and find better ways around human immunity, as early data suggest Omicron is doing.

Booster shots have been shown to bolster waning immunity, and may be a way for rich countries to keep future waves in check, but low- and middle-income countries cannot adopt such a strategy "without destroying their health budgets," says Gagandeep Kang, a virologist at Christian Medical College, Vellore, in India. In the long run, we may need a new generation of vaccines, says Richard Hatchett, head of the Coalition for Epidemic Preparedness Innovations. Hatchett says he always viewed the current arsenal as "rapid response vaccines," fit for the pandemic emergency, but not for the long-term struggle with the virus.

One solution may be vaccines that ramp up the immune response in the mucosa of the respiratory tract, where the virus first takes hold. If the approach can curb transmission, Nkengasong says, "That will truly become the game changer." Others are holding out hope for a pancoronavirus vaccine that would not only thwart new variants, but any future cousins of SARS-CoV-2 as well.

Maybe *Science* will anoint a new vaccine as its breakthrough next year, or the year after. But whatever vaccines the world develops next, we will also have to find ways to use them better. So far in this pandemic, it has been the virus, not humanity, that has done most of the evolving. ■

This coal-fired power plant in Boxberg, Germany, won't be powered off until 2038.

PHOTO: FLORIAN GAERTNER/GETTY IMAGES

BREAKDOWNS OF THE YEAR

What went wrong in the world of science

Hope dims for climate target

The world's nations gathered in early November for U.N. climate talks in Glasgow, U.K., to keep alive the hope of limiting global warming to 1.5°C above preindustrial levels. It's an ambitious goal, with the world already at 1.2°C and the global economy still overwhelmingly reliant on fossil fuels. After 2 weeks of discussion, it was clear that—although rapid change has begun—the 1.5°C target is on life support.

The recent, rapid adoption of renewable energy has likely averted the worst-case warming scenarios of 4°C or more that seemed probable 10 years ago. And for the first time, there is a path toward limiting warming to just below 2°C, as long as nations keep their Glasgow pledges. But it would take far more ambitious action—halving present emissions within 10 years—to reach the 1.5°C goal.

Other progress in Glasgow included a deal on the rules for carbon markets and reporting emissions. But given that U.N. agreements are not binding, the true fate of carbon emissions will come down to action at the national level.

To keep the pressure on, the U.N. talks are now encouraging countries to offer revised emissions reduction pledges every year, including at the next meeting, in Sharm el-Sheikh, Egypt, in November 2022. *—Paul Voosen*

Alzheimer's drug prompts outrage

Under normal circumstances, U.S. approval of the first drug designed to interfere with the biology underlying Alzheimer's disease would be cause for celebration. But aducanumab, cleared by the Food and Drug Administration (FDA) in June, was not the drug many researchers were hoping for. Its bizarre path to market stoked divisions between scientists, sowed confusion among doctors and patients, and prompted many to question the integrity of the U.S. drug regulatory system.

The drug, an intravenously delivered antibody developed by Biogen and marketed as Aduhelm, clears sticky plaques of the protein beta amyloid, thought to cause damage in the brains of people with Alzheimer's disease. But only one of two large clinical trials showed the drug was better at slowing cognitive decline than a placebo. (One possible explanation is that removing amyloid plaques isn't sufficient to reverse a cascade of harmful brain changes already underway by the time Alzheimer's symptoms emerge.)

In November 2020, an independent advisory committee to FDA recommended overwhelmingly against approving the drug. But 7 months later, the agency stunned scientists by greenlighting aducanumab under an accelerated approval pathway. That process relies on a "surrogate endpoint"—here, beta amyloid reduction—rather than a demonstration of clinical benefit.

Reaction was swift. Three of 10 advisory committee members resigned in June. One of them, Aaron Kesselheim of Harvard Medical School, called the move "probably the worst drug approval decision in recent U.S. history." Journalists dug into the close relationship between Biogen executives and FDA officials, reporting back-channel communications and unofficial meetings. Investigations into the approval process by the inspector general of the Department of Health and Human Services and by two committees in the U.S. House of Representatives are ongoing.

Some physicians have said they will not prescribe the drug, and medical centers, including the Cleveland Clinic and Mount Sinai, have announced they won't administer it. Biogen told investors in September that sales of the $56,000-a-year drug had been slower than anticipated. The company announced in November that it was investigating the death of a 75-year-old woman who took aducanumab and developed brain swelling, a known side effect.

Even supporters of the drug say the turbulent rollout has harmed the field. "All of this is confusing and scary for people and families who are facing this devastating and fatal disease," Maria Carrillo, chief scientific officer of the Alzheimer's Association, which advocated for approval, said at a conference last month. *—Kelly Servick*

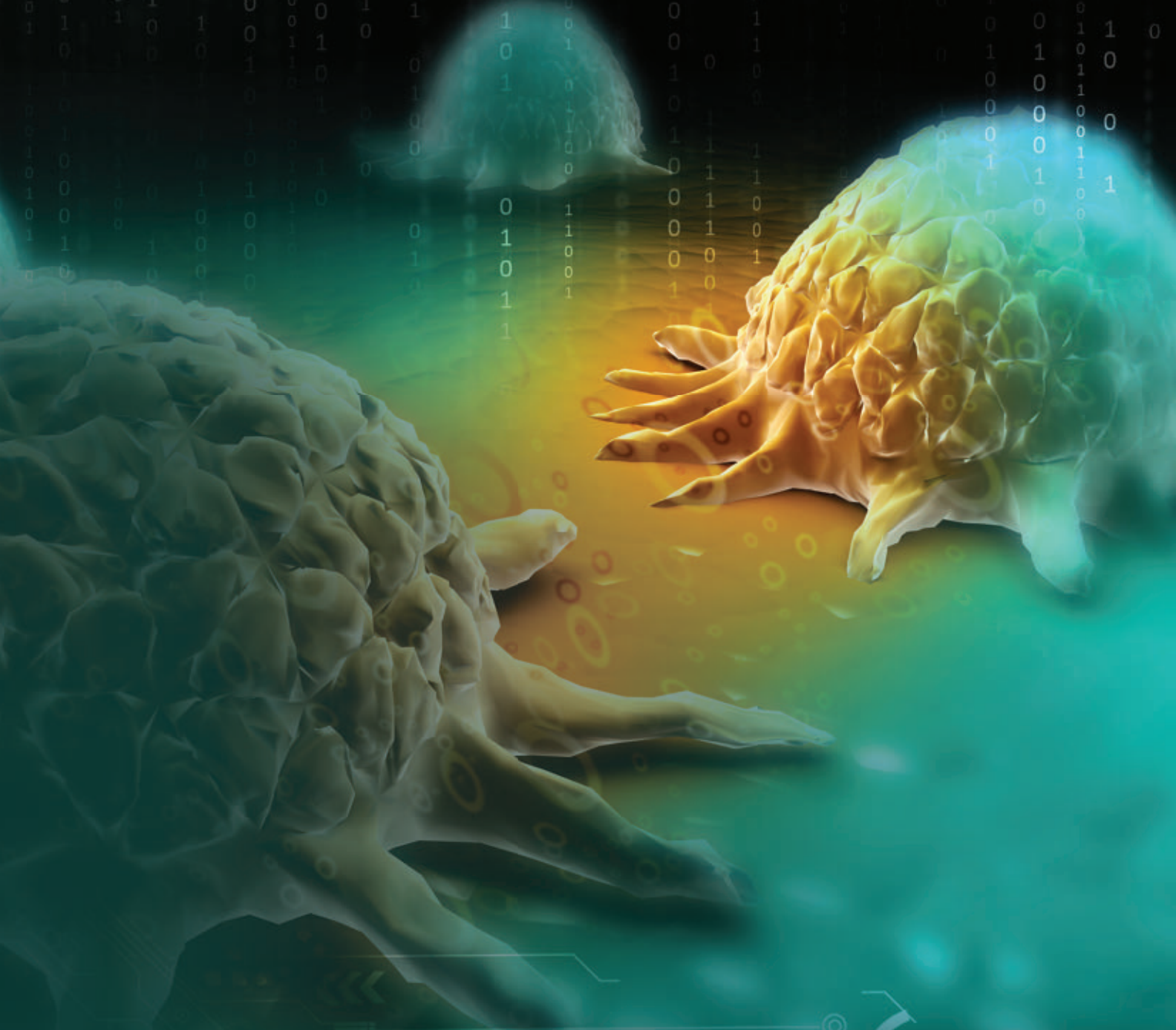
Scientists under fire

Scientists have long come under attack for their work. But this year, political rifts over the COVID-19 pandemic sparked unprecedented public hostility toward scientists, including online and offline intimidation, protests, and death threats.

Those involved in public health suffered the highest profile harassment. U.S. National Institute of Allergy and Infectious Diseases Director Anthony Fauci, for example, needed a full-time security detail as early as April 2020. Chris Whitty, England's chief medical officer, faced harassment in public and regular demonstrations outside his house before the U.K. government took protective measures. Health workers and officials around the world have reported physical and online attacks—and many have quit their jobs as a result.

Researchers who criticized unproven treatments for COVID-19 also faced abuse. Marcus Lacerda, a clinical researcher who led a large trial in Manaus, Brazil, showing no benefits from high-dose hydroxychloroquine, received death threats. Scientific integrity consultant Elisabeth Bik faced online intimidation—and legal threats—after raising concerns about ethical and methodological problems in French microbiologist Didier Raoult's hydroxychloroquine research.

Such threats have had a chilling effect on scientists: A *Nature* survey of 321 researchers who spoke to the media about the pandemic found that more than half have had their credibility attacked, and 15% had received death threats. Many said the experiences left them unwilling to give future interviews. *—Cathleen O'Grady*



2022

AAAS MARTIN AND ROSE WACHTEL CANCER RESEARCH AWARD

Recognize the work of an early career scientist who has performed outstanding work in the field of cancer research. Award nominees must have received their Ph.D. or M.D. within the last 10 years. The winner will deliver a public lecture on his or her research, receive a cash award of **$25,000**, and publish a Focus article in *Science Translational Medicine*.

For more information visit
www.aaas.org/aboutaaas/awards/wachtel
or e-mail wachtelprize@aaas.org.
Deadline for submission: **February 1, 2022**.

AAAS Science Translational Medicine

INSIGHTS

POLICY FORUM

GLOBAL HEALTH

COVID-19 impact on infant and adolescent vaccine supplies

Vaccine production is quadrupling rapidly, creating supply chain challenges

By **Tania Cernuschi[1], Stefano Malvolti[2], Matthew Downham[3], Dominique Maugeais[4], David Robinson[5], Alejandro Cravioto[6]**

Immunization prevents 4 to 5 million deaths annually, primarily among children, but each year 20 million infants do not receive a full course of the most essential basic vaccines. COVID-19 has underscored the importance of vaccines to public health, but immunization coverage dropped in 2020 as a result of the pandemic, leaving even more infants un- or underimmunized (*1*). The push to manufacture COVID-19 vaccines has raised concerns that supplies of other essential vaccines may be compromised, which could erode the gains achieved by immunization and delay access for underserved populations. Drawing on data assembled by the World Health Organization (WHO) and on the advice of technical experts (see supplementary materials), we describe how COVID-19 is affecting the global supply of key infant and adolescent vaccines (see the table). We assess the risks to those essential vaccines, identify mitigations, and explore how emerging innovations can help improve market health.

ISSUES AFFECTING VACCINE SUPPLIES

Prior to the pandemic, approximately 5 billion doses of vaccines were manufactured globally each year. In 2022, COVID-19 vaccine production is projected to reach at least 14 billion doses per year (*2*). This unprecedented fourfold increase in overall

PHOTO: ZAKIR HOSSAIN CHOWDHURY/NURPHOTO/GETTY IMAGES

A health assistant applies a vaccine to a child in a rural area of Bangladesh in September 2020.

manufacturing is affecting production of essential non-COVID vaccines in several ways.

Shortages of input materials

Manufacturers report that they are facing difficulties in timely access to consumable components such as bioprocessing bags, filters, tubing, and laboratory supplies. Culture media—an essential raw material for most vaccines—and vials, syringes, and ampoules have been backordered, with promised delivery times as long as 18 months. Several of those shortages are expected to persist for a considerable time.

These input material shortages are due to the rapid scale-up of COVID-19 vaccine manufacturing, exacerbated by distribution challenges. Because suppliers for certain critical components are limited in number and concentrated in a few countries, access was difficult early in the pandemic when goods distribution was greatly constrained. In addition, emergency measures may have contributed to prioritizing the manufacture of COVID-19 vaccines over the manufacture of other products (*3*).

Multiple factors are magnifying the impact of these shortages. Vaccine manufacturers have moved toward lean supply chains, limiting their reserve stocks of input materials and becoming more sensitive to supply disruptions. Manufacturers have also shifted away from stainless-steel bioreactors to single-use assemblies and bioprocessing bags. As a result, they are critically dependent on sophisticated consumable components that are typically not interchangeable.

Manufacturing capacity constraints

Some manufacturers report that they have repurposed facilities or production capacity built for other vaccines to manufacture COVID-19 vaccines. In some instances, capacity expansion for non–COVID-19 vaccines has been hampered by delays in equipment availability and installation, with lead times of up to 18 months. Contract manufacturing organizations (CMOs) have limited availability to address capacity constraints for other vaccines because they are being used extensively for COVID-19 vaccine production.

Human resources are also limited. Reassignment to COVID-19 vaccine production and testing, facility lockdowns, and absences due to quarantines and illnesses have caused shortages of experienced personnel, including laboratory staff and engineers. These shortages have delayed some manufacture, testing, and release of essential vaccines and continue to constrain capacity.

Postmanufacturing concerns

To monitor product quality, some regulators require submission of every vaccine lot for review and release. This process has been slowed owing to the increased workload for regulatory agencies and because COVID-19 vaccines have been prioritized for regulatory review, delaying the availability of other vaccines. Updates to the manufacturing or testing for vaccines that are already licensed also require regulatory review and may be delayed for similar reasons. Regulatory delays have also been a bottleneck for countries that import vaccines.

More broadly, manufacturers are reevaluating their product portfolios; this may result in reprioritization or even discontinuation of less profitable vaccines in favor of COVID-19 vaccine manufacture. Vaccines with lower prices and smaller market shares, such as polysaccharide vaccines for pneumococcus and meningococcus, are at greater risk.

CURRENT AND ANTICIPATED DISRUPTIONS

Supply disruptions have occurred for pneumococcal conjugate vaccines, measles-containing vaccines, diphtheria and tetanus-containing vaccines, and inactivated poliovirus vaccines. Disruptions have included delivery delays and reduced quantities of specific products (see the table). In most cases, manufacturers have been able to secure stocks of input materials and procurement agencies have been able to access alternative sources, minimizing the impact on vaccination programs.

In the latter part of 2021, consumables and single-use materials such as biopro-

Current and anticipated vaccine supply disruptions

Anticipated disruptions reflect expert judgment informed by aggregated data collected from vaccine manufacturers. See supplementary materials for details.

VACCINES	DISRUPTIONS AS OF NOVEMBER 2021	RISKS OF FUTURE DISRUPTIONS
Pneumococcal conjugate vaccines	Short-term disruptions	**Higher risk.** Owing to similarities between conjugation steps and messenger RNA duplication steps, COVID-19 has competed for laboratories and workforce in the past, and this risk has persisted. In this highly segmented market, delays in production, capacity scale-up, and availability of new products may trigger shortages in countries that prefer specific products.
Measles-containing vaccines	Short-term disruptions	**Higher risk.** If not rapidly addressed, gaps in vaccine coverage are predicted to lead to an increase in measles outbreaks. Owing to high reliance on one supplier that is also manufacturing COVID-19 vaccines, delays in production and constrained surge capacity may limit supplies available to respond to these outbreaks, especially for large countries.
Human papillomavirus vaccine	None observed	**Moderate risk.** Regulatory approval and clinical development of three products have been affected because of the pandemic and travel restrictions, leading to increased risk of delays in availability of those new products.
Diphtheria and tetanus-containing vaccines, including pertussis vaccines	Reduced availability or delays for tetanus-diphtheria (Td) and diphtheria-tetanus (DT)	**Moderate risk** for products containing acellular pertussis (aP) antigens. Owing to similarities in purification steps, aP may compete with COVID-19 vaccines for materials and equipment. Because the supply of aP-containing vaccines closely matches demand, any delays or decreases in production may lead to shortages. **Very low risk** to supplies of other diphtheria- and tetanus-containing vaccines.
Bacille Calmette-Guérin vaccine (for tuberculosis)	None observed	**Low to moderate risk** due to increasing market concentration among a small number of prequalified products. Variability in the manufacturing process may contribute to future shortages of prequalified products.
Rotavirus vaccines	None observed	**Low risk.** Because the supply of rotavirus vaccines is increasing, manufacturing delays are not expected to lead to shortages.
Oral poliovirus vaccines (OPV)	None observed	**Low risk.** Risk of future supply delays due to COVID-19 impact in countries where OPV vaccines are manufactured.
Inactivated poliovirus vaccines	Short-term disruptions	**Very low risk.** Some manufacturers are at risk for supply interruptions, but given the number of different suppliers, major impacts are not anticipated.

cessing bags remain difficult to obtain. If this leads to vaccine shortages, countries may find it more difficult to close immunity gaps, increasing the potential for outbreaks of measles and other vaccine-preventable diseases (*4*).

Products in clinical development, such as human immunodeficiency virus vaccines and next-generation human papillomavirus, tuberculosis, and diphtheria- and tetanus-containing vaccines, are currently delayed or at risk of delay owing to disruptions in the manufacture of clinical materials, the conduct of clinical studies, or regulatory review.

MITIGATION STRATEGIES

We propose these actions to reduce the impact of current disruptions and minimize the risks of future interruptions in vaccine supply.

Enabling access to input materials

Access to input materials can be improved by reducing trade barriers to accelerate the flow of goods. The COVID-19 Vaccines Global Access (COVAX) Manufacturing Taskforce has developed the first virtual marketplace to match suppliers of critical input materials with manufacturers of vaccines (*5*). Approaches to enabling the flow of these and other materials include using market information to identify specific materials that require expedited customs processing; removing import and export tariffs and other barriers to trade; implementing harmonized coding on critical consumables to expedite their cross-border processing; and eliminating export restrictions and prohibitions on critical input materials, as called for by leaders of the World Trade Organization and other international bodies. Such mechanisms are being used to facilitate access to COVID-19 vaccines. Extending them to the overall flow of input materials would help ensure supplies of other important vaccines.

Enhancing regulatory capacity and efficiency

Regulators must ensure vaccine safety and efficacy without loss of time. This will require a combination of greater resources for regulatory authorities and more efficient regulatory processes. As demonstrated in the response to COVID-19, harmonizing processes and expectations across different regulatory bodies, acceptance by one regulatory body of decisions made by another (referred to as "reliance"), and eliminating delays between successive phases of review can drastically accelerate timelines without sacrificing diligence and rigor. These practices can be applied more broadly for other essential vaccines.

In addition, streamlined regulatory processes for allowing manufacturers to use functionally equivalent input materials, such as stoppers of different colors, would enable more nimble procurement during shortages.

Compiling and disseminating global demand forecasts for essential non-COVID vaccines

Manufacturers rely on demand forecasts for planning and to optimize their use of manufacturing capacity. Through the efforts of WHO, the Pan-American Health Organization Revolving Fund, UNICEF, and Gavi, the Vaccine Alliance, this kind of market information has become more widely available in recent years (*6*–*8*). Now, as vaccine supplies and health systems are stretched thin and the context is evolving rapidly, it is even more important to keep abreast of sudden, large changes in demand. This is especially true for vaccines at risk for disruptions (see the table); vaccines at risk of shortages because supply and demand are too closely matched; and for vaccines with a potential for demand surges, such as those used in outbreak response. We suggest that comprehensive, global demand forecasts should be compiled more frequently

[1]Department of Immunization, Vaccines and Biologicals, World Health Organization, Geneva, Switzerland. [2]MMGH Consulting GmbH, Zurich, Switzerland. [3]Coalition for Epidemic Preparedness Innovations, London, UK. [4]Gavi, the Vaccine Alliance, Geneva, Switzerland. [5]Bill & Melinda Gates Foundation, Seattle, WA, USA. [6]Faculty of Medicine, Universidad Nacional Autonóma de México, Mexico City, Mexico. Email: cernuschit@who.int

and that more countries should contribute high-quality demand data for the sake of forecasting.

Enabling demand flexibility

Vaccine procurement specifications typically limit the range of products that can be purchased. They may require specific product presentations (such as multidose vials or prefilled syringes), suppliers of particular nationalities, or products and labeling that have been approved by the local regulatory authorities. Where global policy and regulatory recommendations support the use of similar vaccines made by different manufacturers, greater flexibility in product choice and procurement could enhance access (*9, 10*). Such flexibility may require acceptance of a product addressing fewer strains, or use of reliance mechanisms, including WHO prequalification (which enables procurement by United Nations agencies), that can expedite marketing authorization (*11*). Further regulatory convergence on simplified, harmonized labels acceptable across many jurisdictions would increase flexibility in procurement.

LONG-TERM PROSPECTS

Healthy markets for vaccines provide reliable, timely supplies of high-quality vaccines at sustainable prices and enable innovation to meet evolving needs (*12*). Although COVID-19 has posed new risks in the short term, it has also driven several changes that may create opportunities for long-term improvements in market health.

First, once demand for COVID-19 vaccines is met, manufacturers and CMOs will be left with a legacy of much higher capacity. Although some downsizing is anticipated, there will also be opportunities to repurpose this capacity to address existing shortages; support the development and manufacture of new vaccines or other biomedical products; or maintain existing manufacturing capacity in a state of readiness to respond to future emergencies. Leveraging this capacity will require a willingness to invest in being better prepared for the next pandemic.

Second, challenges in global access to COVID-19 vaccines have led to calls for the global diversification of vaccine manufacturing. Increasing manufacturing capacity in underserved regions would accelerate regional access to vaccines and improve preparedness and rapid response for future emergencies.

Third, newer vaccine platforms such as nucleic acid and viral-vectored vaccines are showing their flexibility and utility in the response to the COVID-19 pandemic. These new platforms can drastically accelerate adaptation to new pandemic and seasonal influenza strains; could be applied to additional pathogens, accelerating access to new vaccines; and could potentially be used in curative applications—for example, as cancer vaccines (*13, 14*).

Finally, the mitigation strategies discussed above will yield a more efficient and resilient vaccine manufacturing ecosystem. In future pandemics, this ecosystem will better enable rapid, equitable access to pandemic vaccines and ensure reliable supplies of other essential vaccines.

STRONG, PRE-EMPTIVE, AND COORDINATED

The global response to the COVID-19 pandemic has demonstrated the power of vaccines to fight disease and the power of global collaboration to achieve unprecedented speed and scale in the discovery, manufacture, and delivery of new vaccines. It has also shown the importance of equitable access to vaccines and of resilient health systems that can deliver routine care, including vaccination, while coping with health emergencies.

> **"Increasing manufacturing capacity in underserved regions would accelerate regional access to vaccines and improve preparedness and rapid response..."**

At the same time, COVID-19 is creating challenges for other vaccines. Disruptions in the supply of essential infant and adolescent vaccines will disproportionately affect poorer countries, which are at greater risk of disease.

Owing to complex manufacturing processes and long production times, manufacturing constraints are typically slow to show their full impact, with a delay sometimes up to 12 to 18 months. Similarly, actions taken may not result in increased supplies for over a year. It is therefore critical to identify and mitigate risks to vaccine supplies as early as possible to minimize their impact.

Mitigating those risks requires a strong, pre-emptive, and coordinated effort. The measures proposed in this paper build on innovative collaborations underway to accelerate access to COVID-19 vaccines. Expanding their scope will help ensure reliable supplies of all essential vaccines. Moreover, applying these mechanisms more broadly will contribute to healthier vaccine markets that are better prepared to respond to future emergencies.

The actions described here are needed to enable recovery, extend the benefits of immunization to underserved populations, and speed the introduction of new vaccines, contributing to a healthier, safer, and more prosperous future (*15*). ■

REFERENCES AND NOTES

1. World Health Organization, UNICEF, "Progress and Challenges with Sustaining and Advancing Immunization Coverage During the COVID-19 Pandemic" (2021); https://cdn.who.int/media/docs/default-source/immunization/wuenic—progress-and-challenges-15-july-2021.pdf.
2. World Health Organization, "Global COVID-19 Vaccination — Strategic Vision for 2022: Technical Document" (2021); https://cdn.who.int/media/docs/default-source/immunization/sage/covid/global-covid-19-vaccination-strategic-vision-for-2022_sage-yellow-book.pdf.
3. T. J. Bollyky, C. P. Brown, The real vaccine procurement problem: Why America should make its supply chain more transparent. *Foreign Aff.* (24 June 2021).
4. L. Roberts, *Nature* **580**, 446 (2020).
5. Coalition for Epidemic Preparedness Innovations, "The COVAX Marketplace" (2021); https://cepi.net/the-covax-marketplace/.
6. UNICEF, "Market notes and updates"; www.unicef.org/supply/market-notes-and-updates.
7. World Health Organization, "Immunization, Vaccines and Biologicals"; www.who.int/teams/immunization-vaccines-and-biologicals/vaccine-access/mi4a.
8. Gavi the Vaccine Alliance, "Supply and procurement roadmaps" (2020); www.gavi.org/our-alliance/market-shaping/supply-and-procurement-roadmaps.
9. World Health Organization, *Wkly. Epidemiol. Rec.* **21**, 421 (2012).
10. World Health Organization, *Wkly. Epidemiol. Rec.* **92**, 241 (2017).
11. World Health Organization, "WHO Expert Committee on Specifications for Pharmaceutical Preparations: Fifty-fifth report," WHO Technical Report Series no. 1033 (World Health Organization, Geneva, 2021); https://apps.who.int/iris/bitstream/handle/10665/340323/9789240020900-eng.pdf.
12. Immunization Agenda 2030, "SP6 Supply Security," (2021); www.immunizationagenda2030.org/images/documents/BLS20116_IA_Global_strategy_document_SP_6_supply_002.pdf.
13. N. Pardi, M. J. Hogan, F. W. Porter, D. Weissman, *Nat. Rev. Drug Discov.* **17**, 261 (2018).
14. R. A. Feldman *et al.*, *Vaccine* **37**, 3326 (2019).
15. Immunization Agenda 2030, "Call to Action" (2021);www.immunizationagenda2030.org/pledge-support.

ACKNOWLEDGMENTS

We thank all those who contributed data and perspectives to this analysis, the Market Information for Access to Vaccines Advisory Group, the participants of the WHO workshops on impact of COVID-19 on supply of other vaccines, and A. Hwang for editorial support. The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the agencies or organizations with which the authors are employed or affiliated. The authors receive funding from the Bill & Melinda Gates Foundation (grant 70988 to WHO), core funding from the WHO program budget, and funding from WHO (contract 202651859 to S.M.). The authors declare no interests beyond those shown on the title page.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7019

10.1126/science.abl7019

PERSPECTIVES

Biosphere 2 is a large-scale, fully enclosed Earth science facility in Arizona, USA, that harbors a well-structured tropical forest ecosystem with a large number of different woody and herbaceous plant species.

PHOTO: LAURA MEREDITH

ECOLOGY

Ecosystem effects of environmental extremes

A large-scale experimental facility reveals tropical rainforest responses to drought

By **Nico Eisenhauer**[1,2] *and* **Alexandra Weigelt**[1,3]

Climate and biodiversity change influence carbon, water, and greenhouse gas dynamics, thereby driving the delicate balance of ecosystems as carbon sinks or sources (*1*, *2*). On page 1514 of this issue, Werner *et al.* (*3*) show how highly controlled and large-scale Earth science facilities can reveal the plant physiological processes, the role of plant-soil-atmosphere interactions, and the dynamics of greenhouse gases and volatiles that underlie tropical forest responses to drought. Specifically, they used an enclosed experimental rainforest in the Biosphere 2 facility in Oracle, Arizona (*4*) to explore how drought influences carbon and water fluxes as well as soil-plant-atmosphere interactions by tracing isotopically labeled $^{13}CO_2$ and deep-water $^{2}H_2O$, and volatile organic compounds related to drought stress. The forest ecosystem displayed highly diverse drought responses within and among species throughout increasing drought and recovery phases, reflecting differences in functional plant traits, drought adaptations, and microclimatic conditions.

Werner *et al.* observed that drought-tolerant and drought-sensitive tree species differed substantially in how they contributed to carbon and water fluxes over the course of the drought and rewetting periods (see the figure). Drought-sensitive species were the largest contributors before and after drought, whereas the importance of drought-tolerant species increased under severe drought. These results suggest that different hydraulic strategies are used by distinct plant functional groups and that their changing proportional contributions determine overall ecosystem fluxes and plant-soil-atmosphere interactions (*5*). These findings further point to a key role of plant hydrological traits, which may determine forest responses to climate extremes (*6*), and to the effects of tree diversity on ecosystem functioning, as found in subtropical China (*7*). Given the importance of belowground processes and intimate plant-soil interactions, future work should consider above-belowground trait syndromes (i.e., associations between leaf and root traits) (*8*), because root traits are critical in carbon and water fluxes and provide a predictive framework for plants and their environment in a changing world (*8*).

Werner *et al.* also report that drought effects cascaded through all forest strata and changed over time. Atmospheric drought was followed by overstory responses that propagated to the understory and to the soil, from top layers to deeper layers during late and severe drought. The authors also observed decreases in evapotranspiration, ecosystem respiration, and gross primary productivity. At the same time, the concentrations, dynamics, and composition of atmospheric volatiles changed substantially during drought. During predrought conditions, the soil absorbed volatiles produced by the canopy, but it did not function as a strong volatile sink during severe drought. As drought progressed, several volatiles characteristic for drought stress of trees were released into the atmosphere. These findings stress that there are critical thresholds of soils to buffer the consequences of drought and underline climate change effects on plant-soil feedbacks (*9*). Moreover, these insights show that the composition of volatiles can be a powerful stress indicator as well as an important feedback mechanism to climate change. For example, isoprene production (a volatile liquid hydrocarbon synthesized by plants) indicated the onset of reductions in evapotranspiration and gross primary productivity, whereas production of the organic volatile hexanal denoted their final decline under severe drought and was related to leaf senescence. Drought-induced release of volatile organic compounds and particles may contribute to producing nuclei for cloud formation and precipitation, potentially influencing the hydrological cycle in tropical rainforests (*10*). Moreover, Werner *et al.* observed pronounced legacy effects even after several months of rewetting after the drought, which were caused by persistent structural changes in drought-sensitive species (such as a loss of hydraulic conductivity and leaves). This indicates the potential problem of recurrent droughts in consecutive years and may also be related to changes in plant-soil interactions (*11*).

The Werner *et al.* study exemplifies the potential of highly controlled enclosed facilities to explore the consequences of environmental change and unravel insights into the underlying mechanisms. Such facilities, including the so-called Ecotrons, enable the simulation of a wide range of environmental conditions in replicated experimental units and the comprehensive assessment of ecosystem processes (*12*). Mechanistic experiments on physiological responses of organisms can improve our process-based understanding of reciprocal effects of climate change and ecological communities (*1*). Moreover, they en-

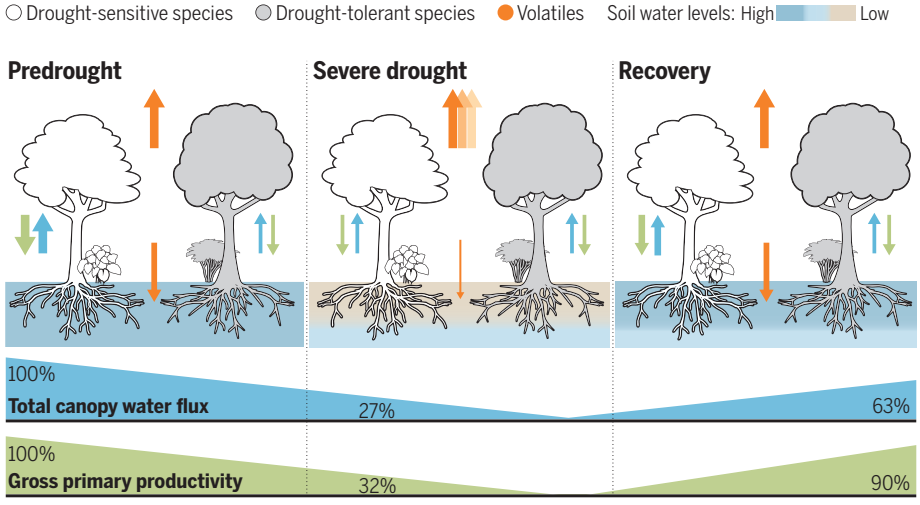
[1]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, 04103 Leipzig, Germany. [2]Institute of Biology, Leipzig University, Puschstraße 4, 04103 Leipzig, Germany. [3]Institute of Biology, Leipzig University, Johannisallee 21, 04103 Leipzig, Germany. Email: nico.eisenhauer@idiv.de

able the identification of proxy measures for larger-scale assessments and biogeochemical models. For instance, Werner *et al.* found characteristic blends of volatiles released by drought-stressed vegetation that can function as indicators of different levels of drought stress and plant senescence. Such information is urgently needed to improve the predictive capacity of biogeochemical models (*1*, *12*). At the same time, following carbon and water dynamics with targeted labeling experiments was mostly possible above the ground, whereas continuously measuring labels in belowground organs, such as fine roots, still poses major challenges that need to be addressed in future work. New facilities like the Deep Soil Ecotron (*13*) will help elucidate belowground processes, going beyond the topsoil and down to a depth of 3 m.

The results by Werner *et al.* indicate that climate extremes like drought can modulate the balance between carbon sink and source processes through changes in plant physiology, plant carbon allocation, and plant-soil-atmosphere interactions. Such insights are critical for linking local experimental studies and global monitoring and modeling approaches. Data availability and information gained from global networks such as FLUXNET (*14*), which measures surface gas exchange between terrestrial ecosystems and the atmosphere, or remote sensing (*15*) across major terrestrial biomes are continuously improving and will allow the assessment of changes in biodiversity, ecosystem functioning, and stress levels based on proxy variables provided by local experimental work. Given that evapotranspiration, ecosystem respiration, and gross primary productivity all responded substantially to drought in the Biosphere 2 study, future projects should explore whether the plant traits identified as related to drought resistance have similar predictive power across ecosystem types and biomes. Integrating information on plant functional types related to drought resistance and resilience with the role of soils in water and carbon dynamics will help develop a predictive framework for ecosystem effects of extreme events and help parametrize Earth system models by accounting for coupled plant-soil interactions. ■

Tropical forest responses to drought and rewetting

In Biosphere 2, drought effects revealed major changes to water and carbon fluxes, volatile compound composition by vegetation and uptake by soil, and gross primary productivity in a tropical forest ecosystem. After rewetting, water and carbon fluxes slowly recovered, but not to predrought levels.



GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. H. O. Pörtner *et al.*, IPBES-IPCC co-sponsored workshop report on biodiversity and climate change, Zenodo (2021); https://doi.org/10.5281/zenodo.5101133.
2. G. B. Bonan, *Science* **320**, 1444 (2008).
3. C. Werner *et al.*, *Science* **374**, 1514 (2021).
4. L. S. Leigh, T. Burgess, B. D. V. Marino, Y. D. Wei, *Ecol. Eng.* **13**, 65 (1999).
5. W. R. L. Anderegg, A. T. Trugman, D. R. Bowling, G. Salvucci, S. E. Tuttle, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14071 (2019).
6. F. Schnabel *et al.*, bioRxiv 10.1101/2021.01.06.425434 (2021).
7. F. J. Bongers *et al.*, *Nat. Ecol. Evol.* **5**, 1594 (2021).
8. A. Weigelt *et al.*, *New Phytol.* **232**, 42 (2021).
9. F. I. Pugnaire *et al.*, *Sci. Adv.* **5**, eaaz1834 (2019).
10. U. Pöschl *et al.*, *Science* **329**, 1513 (2010).
11. A. Canarini *et al.*, *Nat. Commun.* **12**, 5308 (2021).
12. J. Roy *et al.*, *Glob. Change Biol.* **27**, 1387 (2021).
13. E. Stokstad, *Science* **373**, 250 (2021).
14. M. Migliavacca *et al.*, *Nature* **598**, 468 (2021).
15. A. K. Skidmore *et al.*, *Nat. Ecol. Evol.* **5**, 896 (2021).

ACKNOWLEDGMENTS

We acknowledge the support of iDiv funded by the German Research Foundation (DFG) (DFG–FZT 118, 202548816). The Jena Experiment is funded by the DFG (FOR 5000). We appreciate helpful comments by C. Werner.

10.1126/science.abn1406

MOLECULAR BIOLOGY

Toward single-molecule proteomics

Nanopore rereading of single proteins opens a pathway to next-generation proteomics

By **Filip Bošković** *and* **Ulrich F. Keyser**

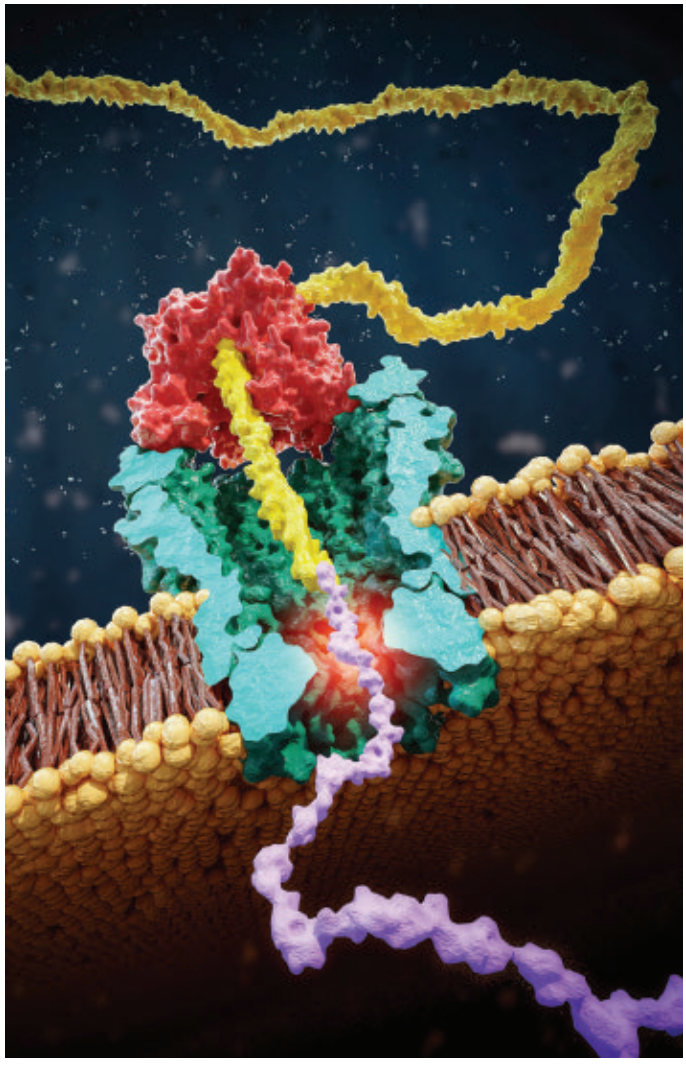
Most sequencing methods for nucleic acids require multiple copies of the target molecules. The copying of nucleic acids, through polymerase chain reaction (PCR), enabled next-generation sequencing technology that uses more finite targets. This has allowed the development of genomics and transcriptomics approaches to profile single cells. Proteins are arguably more important for the proper function of living systems, and so proteomics methods that provide the identity, quantity, and sequence of proteins in a single cell is a key goal. However, one obstacle in protein sequencing is our inability to make identical copies of proteins. On page 1509 of this issue, Brinkerhoff *et al.* (*1*) develop a method to reread a single polypeptide many times, which allows the identification of synthetic peptide variants. Their approach is based on nanopore DNA sequencing and overcomes one of the biggest obstacles toward single-molecule proteomics.

Protein sequencing currently relies on mass spectrometry–based methods that require large amounts of starting materials and miss rare proteins. The quantification of messenger RNA molecules generally does not reflect the proteome (*2*). Hence, for an accurate determination of cellular protein isoforms and variants, including their post-translational modifications (PTMs), new methods are needed. Often inspired by DNA sequencing approaches, several methods have been proposed, including single-molecule fluorescence, tunneling current analysis, and nanopore sequencing (*3*). Nanopore DNA sequencing involves ionic current sensing for the identification of single nucleotides in nanopores (*4*). This inspired

Cavendish Laboratory, University of Cambridge, Cambridge, UK. Email: ufk20@cam.ac.uk

the development of a similar approach for amino acids (AAs), whereby nanopore measurements of single AAs were identified, even distinguishing leucine and lysine, which have the same molecular weight (*5*). However, the technique relied on the incorporation of target AAs into predesigned charged peptide chains.

In addition to the enhanced chemical complexity of proteins compared with nucleic acids, they also come in folded configurations that usually prevent readout in a linear fashion. Here, nature may provide a helping hand in the form of molecular machines, such as unfoldases that could be combined with nanopores to facilitate peptide linearization. A few strategies implemented such guided protein translocation through a nanopore, including using negatively charged DNA (*1*, *6*), carrier peptides (*5*), unfoldases (*7*, *8*), or denaturants (*9*). These approaches improve protein identification by nanopores and may aid understanding of isoform diversity on a polypeptide level, but longer peptide reads remain desirable.



A nanopore sequencer composed of a nanopore (turquoise) and a helicase (red) can reread DNA-protein conjugates (yellow-purple) with single–amino acid resolution.

The first attempts to use nanopore sequencing for peptide discrimination were recently reported, but the approach is limited to reading very short peptides (*6*). By contrast, Brinkerhoff *et al.* present a method to increase the accuracy by using a helicase that is present in solution and is allowed to bind repeatedly to the DNA molecules. This ingenious design ensures that the same polypeptide can be read, pulled back into the nanopore by an electric field, and then reread with the help of a new helicase. The ability to read out the same molecule in a nanopore with such high resolution is a major breakthrough in the field. Crucially, the rereading of the same peptide allows the authors to reduce nanopore read errors to one in a million with only 30 rereads.

The approach of Brinkerhoff *et al.* is not only elegant but also compatible with commercially available nanopore platforms. However, the path toward proteomics will remain challenging. There are ~40 million proteins in a simple eukaryotic cell (*10*). At an average length of 400 AA, the complete proteome yields 16 billion AAs. For a low error rate, 20 rereads for a 26-AA polypeptide requires 2 min. A single nanopore system would need 20.5 million hours to analyze the protein in an entire cell. Even current commercial systems with 144,000 pores would only reduce the read time of the cellular proteome to ~1 week. As in genomics, the question is likely to become whether the full proteome is needed or whether a more targeted approach will suffice.

A combination approach with other protein analysis methods seems inevitable. One method involves single-molecule fluorescence on immobilized proteins, which can identify hundreds of proteins by labeling specific AAs (*11*). Additionally, protein identification has been achieved through protein fragmentation and subsequent nanopore readout (*12*). However, it is challenging to adopt these approaches for quantitative proteomics.

PTMs are another obstacle for nanopore-based protein sequencing. There are hundreds of PTMs that increase the potential complexity of nanopore signals (*13*). Nevertheless, identification of PTMs should be attainable with nanopore sensing and has the potential to be combined with PTM-specific labeling (*14*). However, the molecular mass of some modifications, such as glycosylation, can substantially surpass that of AAs and stop peptide translocation through nanopores. An enzymatic glycan-removal step could be necessary before nanopore analysis of native proteins, but with the inevitable loss of information on PTMs. Nanopore sensing for PTMs seems possible with biological nanopores where the diameter is adapted to the target. Simultaneous sensing with a range of different nanopores could lead to a comprehensive understanding of proteome and protein isoform diversity.

For the goal of single-molecule proteomics, more developments are required. Spatial proteomics with single-cell resolution will require the positioning of target cells on nanopore arrays. The design of new nanopores with improved chemical recognition to aid AA and PTM discrimination seems to move closer to reality with the advent of protein-folding predictions. The abundant ionic current data will pose new challenges for data analysis. There are still many barriers to overcome; however, the future looks bright for true single-molecule readout-based proteomics. ■

REFERENCES AND NOTES

1. H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev, C. Dekker, *Science* **374**, 1509 (2021).
2. Y. Liu, A. Beyer, R. Aebersold, *Cell* **165**, 535 (2016).
3. J. A. Alfaro *et al.*, *Nat. Methods* **18**, 604 (2021).
4. J. Clarke *et al.*, *Nat. Nanotechnol.* **4**, 265 (2009).
5. H. Ouldali *et al.*, *Nat. Biotechnol.* **38**, 176 (2020).
6. S. Yan *et al.*, *Nano Lett.* **21**, 6703 (2021).
7. S. Zhang *et al.*, *Nat. Chem.* **13**, 1192 (2021).
8. J. Nivala, D. B. Marks, M. Akeson, *Nat. Biotechnol.* **31**, 247 (2013).
9. P. Tripathi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016262118 (2021).
10. B. Ho, A. Baryshnikova, G. W. Brown, *Cell Syst.* **6**, 192 (2018).
11. C. V. de Lannoy, M. Filius, R. van Wee, C. Joo, D. de Ridder, *iScience* **24**, 103239 (2021).
12. F. L. R. Lucas, R. C. A. Versloot, L. Yakovlieva, M. T. C. Walvoort, G. Maglia, *Nat. Commun.* **12**, 5795 (2021).
13. G. A. Khoury, R. C. Baliban, C. A. Floudas, *Sci. Rep.* **1**, 90 (2011).
14. L. Restrepo-Pérez, C. H. Wong, G. Maglia, C. Dekker, C. Joo, *Nano Lett.* **19**, 7957 (2019).

10.1126/science.abn0001

PHOTO: CEES DEKKER LAB TU DELFT/SCIXEL

SEMICONDUCTORS

Nitride perovskite becomes polar

An oxygen-free polar perovskite offers several advantages over perovskite oxides

By **Xia Hong**[1,2]

There is a strong drive behind the quest for thin-film materials that are oxygen-free and polar. Oxygen hinders the integration of ferroelectric oxides with semiconductors, which affects efforts to develop nonvolatile memory—that is, a memory that can sustain its information without power. Ideally, one would use single-crystalline perovskite films to construct these devices so that the polarization can be maximized. However, when depositing crystalline polar perovskite oxides onto silicon or germanium, a nonpolar oxide buffer layer (*1*) or a native oxide layer (*2*) can be present at the interface, compromising device performance. A nitrogen-based perovskite may overcome this limitation (*3*). On page 1488 of this issue, Talley *et al.* (*4*) report the synthesis of lanthanum tungsten nitride ($LaWN_3$) thin films, which marks the first demonstration of polar nitride perovskite. This may lead to oxygen-free integration of functional perovskite on a semiconductor platform.

Perovskites, defined by their ABX_3 composition and structure, are one of the most abundant naturally occurring crystal structures on Earth. Despite the simplicity of their pseudocubic structure, perovskite materials form a versatile playground for fundamental exploration and technological development. For example, in perovskite oxides (ABO_3), the B-site cation is surrounded by the oxygen octahedron, which not only defines the crystal structure but also facilitates electron hopping and magnetic exchange. This enables the fine-tuning of the energy scales of various material phenomena, such as charge correlation, electron-phonon coupling, and spin ordering, all by manipulating the crystal structure. The competition between these energy scales has produced a diverse roster of functional materials that spans the entire electronic spectrum, ranging from superconductors to insulators and from dielectrics to ferroelectrics and even multiferroic materials.

Perovskite oxides whose structures are not inversely symmetric tend to exhibit piezoelectricity (generation of polarization in response to applied mechanical stress). Ferroelectrics form a subset of the polar piezoelectric oxides that possess a spontaneous polarization or ordered electric dipole moments, which can be switched by an electric field. Tailoring the ordering of these dipole moments at the nanoscale can lead to a plethora of technologically relevant quantum phenomena, such as a polar vortex state with applications in piezoelectronics (*5*), a negative capacitance effect for low-power logic applications (*6*), and topologically protected domain walls for nonvolatile memories (*7*). Besides oxides, another class of perovskite with polar characteristics is the hybrid halide perovskite, which has drawn enormous research interest as a promising photovoltaic material. It has been suggested that being polar is one of the critical factors that lead to its high power-conversion efficiency (*8*, *9*).

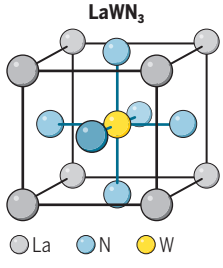
Despite numerous theoretical efforts to study perovskite nitrides (*10*–*13*), there have only been a few successful experimental demonstrations of oxygen-free nitride perovskites—and mostly in non–single-crystalline, powder forms (*14*). Among the many possible cation combinations, $LaWN_3$ has been predicted to be thermodynamically stable (*10*). Talley *et al.* synthesized polycrystalline $LaWN_3$ films on fused silica and p-type silicon substrates and confirmed the crystal structure with x-ray scattering and electron microscopy. Characterizations by means of spectroscopy techniques further showed that these films are free of oxygen content.

Perhaps the most interesting finding of the $LaWN_3$ film is that it is polar. Talley *et al.* narrowed down the possible crystal structure to either noncentrosymmetric rhombohedral or centrosymmetric tetragonal. Further characterization by using piezo-response force microscopy revealed a large piezoelectric coefficient (d_{33}) of about 40 pm/V. This confirmed that the material is indeed polar as a noncentrosymmetric rhombohedral structure, which agrees with the theoretical prediction (*11*). Compared with the piezoelectric oxide and nitride reference samples, the d_{33} value of $LaWN_3$ is considerably higher than that of $LiNbO_3$ and $Al_{0.92}Sc_{0.08}N$ and only smaller than that of $PbZr_{0.52}Ti_{0.48}O_3$, whose composition is close to the structural phase boundary. This makes the polar $LaWN_3$ highly competitive for applications such as mechanical energy harvesters.

Theoretically, $LaWN_3$ has been predicted to be ferroelectric, with a large remnant polarization of about 61 μC/cm^2 and a small energy barrier of about 110 meV for switching the polarization (*11*). If this is true, these properties can be used to develop nonvolatile memory applications with high efficiency and low operation power. Although the global measurements of the switching characteristics of polarization made by Talley *et al.* did not yield conclusive results, they observed hysteresis behavior in the piezo-response force microscopy measurements, which may have originated from ferroelectric switching. At this stage, the existence of ferroelectricity in $LaWN_3$ remains inconclusive.

The demonstration of oxygen-free nitride perovskite with competitive piezoelectric response paves the way for integrating the rich functionalities of polar perovskites with the mainstream semiconductor industry (see the figure). In addition to thin-film bulk acoustic resonators, a range of innovative applications may be possible, including mechanical energy harvesters (*3*), sensors (*3*), thermoelectronics (*12*), nonvolatile memories (*7*), neuromorphic devices (*15*), negative capacitance transistors (*6*), and photovoltaic devices (*3*). Because many of these device con-

Research opportunities for polar nitride perovskites



PROPERTY	APPLICATIONS
Piezoelectricity	• Thin-film bulk acoustic resonator • Mechanical energy harvester
Pyroelectricity	• Thermoelectric energy device • Thermosensor
Ferroelectricity	• Nonvolatile memory • Neuromorphic system-on-chip • Negative capacitance transistor • Photovoltaic device

GRAPHIC: A. MASTIN AND C. SMITH/*SCIENCE*

[1]Department of Physics and Astronomy, University of Nebraska–Lincoln, Lincoln, NE 68588-0299, USA. [2]Nebraska Center for Materials and Nanoscience, University of Nebraska–Lincoln, Lincoln, NE 68588-0299, USA. Email: xia.hong@unl.edu

cepts operate on ferroelectric polarization switching, there are also materials challenges to be tackled. Confirming ferroelectricity in $LaWN_3$ calls for the synthesis of polycrystalline films with large grain size or even single-crystalline films—or identifying strategies for band gap engineering. It is also of strong interest to discover polar nitride perovskites with an intrinsically large band gap and to explore other theoretically predicted properties, such as magnetic order (*10*), distinctive spin textures (*13*), and topological phenomena (*12*). Future progress in this field thus relies on the collective efforts of computationally driven materials search, materials synthesis, property characterization, and device implementation. As an emerging field, it is conceivable that the discovery of functional nitride perovskites provides a promising material platform for fundamental exploration as well as the development of innovative device applications. ■

"The demonstration of oxygen-free nitride perovskite with competitive piezoelectric response paves the way for integrating the rich functionalities of polar perovskites with the mainstream semiconductor industry."

REFERENCES AND NOTES

1. A. Lin *et al.*, *Appl. Phys. Lett.* **78**, 2034 (2001).
2. R. M. Moghadam *et al.*, *Nano Lett.* **17**, 6248 (2017).
3. T. Mikolajick *et al.*, *J. Appl. Phys.* **129**, 100901 (2021).
4. K. R. Talley, C. L. Perkins, D. R. Diercks, G. L. Brennecka, A. Zakutayev, *Science* **374**, 1488 (2021).
5. A. K. Yadav *et al.*, *Nature* **530**, 198 (2016).
6. P. Zubko *et al.*, *Nature* **534**, 524 (2016).
7. J. Ma *et al.*, *Nat. Nanotechnol.* **13**, 947 (2018).
8. J. M. Frost *et al.*, *Nano Lett.* **14**, 2584 (2014).
9. H. Zhu, K. Miyata, Y. Fu, J. Wang, P. P. Joshi, D. Niesner, K. W. Williams, S. Jin, X.-Y. Zhu, *Science* **353**, 1409 (2016).
10. R. Sarmiento-Pérez, T. F. T. Cerqueira, S. Körbel, S. Botti, M. A. L. Marques, *Chem. Mater.* **27**, 5957 (2015).
11. Y.-W. Fang *et al.*, *Phys. Rev. B* **95**, 014111 (2017).
12. M.-C. Jung, K.-W. Lee, W. E. Pickett, *Phys. Rev. B* **97**, 121104 (2018).
13. H. J. Zhao *et al.*, *Phys. Rev. B* **102**, 041203 (2020).
14. S. D. Kloß, M. L. Weidemann, J. P. Attfield, *Angew. Chem. Int. Ed.* **60**, 22260 (2021).
15. S. Oh, H. Hwang, I. K. Yoo, *APL Materials* **7**, 091109 (2019).

ACKNOWLEDGMENTS

X.H. acknowledges support by NSF through grant DMR-1710461 and Established Program to Stimulate Competitive Research (EPSCoR) RII Track-1 Award OIA-2044049 (EQUATE).

10.1126/science.abm7179

CANCER IMMUNOLOGY

An atlas of intratumoral T cells

Intratumoral T cell composition is relevant for disease outcome across tumor types

By **Anne M. van der Leun** *and* **Ton N. Schumacher**

The T cells that are present in human tumors display a diversity of cell states. Because not all T cells have an equal capacity to contribute to antitumor responses, understanding this diversity is critical to define their role in natural tumor control and cancer immunotherapy. On page 1462 of this issue, Zheng *et al.* (*1*) describe a "T cell atlas" that contains transcriptional profiles of T cells across 21 cancer types, addressing aspects such as recurring T cell states, cell differentiation trajectories, and prognostic value. Although there are many ways to slice these data, aspects of particular interest are the pan-cancer identification of T cell subsets that may play an active role in tumor control and the observation that the relative abundance of T cells with distinct states has prognostic value that transcends tumor type, which takes a step toward immune type–based patient stratification.

The capacity of T cells to recognize and eliminate tumor cells forms the mechanistic basis for the activity of immune checkpoint–blocking therapies that have revolutionized cancer care. There is compelling evidence that the contribution of individual T cells to tumor control varies strongly and appears to be associated with their cell state. Specifically, analysis of T cell infiltrates in human tumors has demonstrated that only a small fraction of T cells at such sites is tumor reactive (*2*). In some tumor types, the expression of hallmarks of dysfunction (or exhaustion), including expression of inhibitory receptors such as programmed cell death 1 (PD-1), can be used to distinguish tumor-reactive T cells from neighboring "bystander" cells (*3*, *4*). Additionally, within the tumor-reactive T cell compartment, T cells differ in their capacity to convey antitumor effects. Studies in mouse models have, for example, shown that cells with an early dysfunctional cell state, characterized by the expression of transcription factor 7 (*TCF7*) and longer-term renewal potential, are particularly important for durable responses to immune checkpoint therapy (*5*, *6*).

Division of Molecular Oncology and Immunology, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, Netherlands. Email: t.schumacher@nki.nl

A central question that is addressed by Zheng *et al.* is how the properties of the T cell pool with presumed tumor-reactivity compare across cancer types. $CD8^+$ T cells that are enriched at the tumor and that show clonal expansion, as inferred from T cell receptor (TCR) sequencing, are predominantly observed in the cell pool with a (terminal) exhaustion (T_{ex}) phenotype, which is consistent with chronic antigen exposure driving T cell dysfunction. The same T cell states showed enhanced gene expression signatures associated with both cell division and TCR signaling, which might suggest that tumor-reactive T cells in many human cancers are actively responding to tumor cells even in the absence of therapy. Similarly, use of the latter criteria on the $CD4^+$ T cell compartment identified a regulatory T cell (T_{reg} cell) population that expresses tumor necrosis factor (TNF) receptor superfamily member 9 (*TNFRSF9*), which encodes the activation marker 4-1BB, as most highly enriched for potential tumor reactivity. This demonstrates that both cytotoxic and suppressive tumor-reactive T cell populations may potentially be identified across cancer types according to their transcriptional characteristics, holding promise for both diagnostic and therapeutic applications.

Continuous antigen exposure forms a major driver of T cell dysfunction, but the varied presence of antigen as well as additional cell-bound and soluble factors, such as immune checkpoint ligands and transforming growth factor–β (TGF-β), at the tumor site provide ample opportunity for potential diversification in this process. On the basis of their pan-cancer dataset, Zheng *et al.* propose a model in which two distinct differentiation paths lead to a state of terminal dysfunction, an observation that extends recent work in non–small cell lung cancer (NSCLC) (*7*). Both T_{ex} cell differentiation paths, characterized by the presence of either a granzyme K (*GZMK*$^+$) or zinc finger protein 683 (*ZNF683*$^+$) intermediate $CD8^+$ T cell state, were shown to coexist in a substantial part of tumors. This may be explained by intratumoral heterogeneity in the signals that drive dysfunction but could also reflect the developmental origin of these T cell populations outside of the tumor—for example, because of differential imprinting of naïve T cells during priming, a matter that deserves further attention. The preferential connection

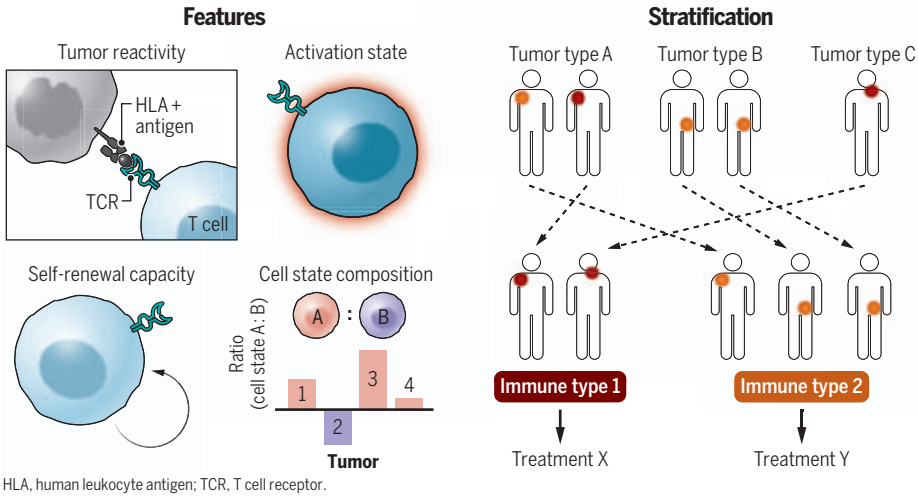
of the *TCF7*$^+$ T_{ex} population with T cell states from the *GZMK*$^+$ differentiation path raises the question of whether there are additional T cell populations with a comparable level of stemness or whether the capacity for self-renewal may differ between these trajectories.

Analysis of the relative abundance of different T cell states across cancers made it possible to distinguish eight "immune types" of cancer, according to their immune composition, characterized by (among others) greater or lesser abundance of T_{ex} cells, *TNFRSF9*$^+$ T_{reg} cells, and various memory T cell states with a lower level of dysfunction. Two of these immune types share a sizable T_{ex} cell population but differ in the abundance of *TNFRSF9*$^+$ T_{reg} cells, a difference that is likely to affect the activity of different immunotherapeutic strategies. Furthermore, Zheng *et al.* observed that patients with either of these "T_{ex}-high" immune-type cancers had reduced survival compared with "T_{ex}-low" immune types that show less T cell dysfunction. Similarly, T_{ex}-low melanomas showed an improved response to immune checkpoint blockade compared with that of T_{ex}-high tumors (*8*). Given the enrichment of tumor-reactive T cells in the dysfunctional T cell pool (*9*), these data argue for a model in which patient outcome is less determined by the presence of a sizable tumor-reactive T cell pool but more so by the capacity to maintain T cell reactivity over time. It will be important to translate the concept of immune types, reflecting aspects such as capacity for T cell renewal and tumor recognition, into assays that can be incorporated into routine diagnostics (see the figure).

As with any atlas, the inclusion of additional layers of information should further increase its value in the coming years. For example, information on the epigenetic state may help to understand to what extent T cell populations can (durably) be reactivated through therapy. Furthermore, a combined analysis of T cells and other immune cell types appears attractive because cross-talk between immune cell subsets likely explains part of the diversity in T cell states observed. Arguably, the most valuable addition touches on a much more central aspect of any map: its spatial resolution. T cells that reside in human tumors may be present either in stromal or parenchymal areas or in tertiary lymphoid structures or other immune cell niches (*10*), determining the signals to which these cells are exposed. A future map that couples T cell state to their presence in defined tumor areas will help to dissect how specific T cell pools shape their local environment and how the functional state of T cells is influenced by the cellular neighborhood in which they grow up. ■

T cell immune types of cancer

Various characteristics of T cells that infiltrate tumors are relevant to therapeutic response, such as the presence of tumor-reactive T cells, their activation state, self-renewal capacity, and the balance between suppressive and effector T cell states. These characteristics may define cancer immune types that can stratify patients to optimize therapy.



HLA, human leukocyte antigen; TCR, T cell receptor.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. L. Zheng *et al.*, *Science* **374**, eabe6474 (2021).
2. W. Scheper *et al.*, *Nat. Med.* **25**, 89 (2019).
3. E. J. Wherry, M. Kurachi, *Nat. Rev. Immunol.* **15**, 486 (2015).
4. A. Gros *et al.*, *J. Clin. Invest.* **124**, 2246 (2014).
5. I. Siddiqui *et al.*, *Immunity* **50**, 195 (2019).
6. B. C. Miller *et al.*, *Nat. Immunol.* **20**, 1556 (2019).
7. P. Gueguen *et al.*, *Sci. Immunol.* **6**, eabd5778 (2021).
8. M. Sade-Feldman *et al.*, *Cell* **175**, 998 (2018).
9. Y. Simoni *et al.*, *Nature* **557**, 575 (2018).
10. D. S. Thommen, T. N. Schumacher, *Cancer Cell* **33**, 547 (2018).

10.1126/science.abm9244

QUANTUM CHEMISTRY

Radical quantum oscillations

Laser spectroscopy reveals spin quantum beats in electron transfer reactions

By **P. J. Hore**

Chemical reactions can be accelerated by weakening chemical bonds in the reactant molecules, which can be done by increasing their thermodynamic free energy. However, nothing much will happen if the increase in free energy per molecule is much less than the thermal energy k_BT, which is the Boltzmann constant multiplied by the temperature of the molecules. Therefore, it may be surprising that magnetic interactions one-hundred thousandth the magnitude of k_BT can influence the course of certain chemical transformations (*1, 2*). The reason that some reactions can be so sensitive to weak magnetic fields lies in the oscillations between different electronic states of the reaction intermediates. These oscillations, known as "quantum beats," play a pivotal role in these exceptions to the rule (*3, 4*). On page 1470 of this issue, Mims *et al.* (*5*) report a technique for revealing these normally hidden oscillations and providing insights into this intriguing class of reactions (see the figure).

Radicals—molecules that contain an odd number of electrons—can be formed as pairs when one molecule transfers an electron to another molecule. In particular, organic radical pairs have distinctive features that give rise to their unusual behavior. The pair can exist as singlet or triplet states—i.e., possessing antiparallel or parallel electron spins, whose reactions obey selection rules in which spin is conserved. Moreover, if the radicals are not too close together, the singlet and triplet states, because of their similar energies, would interconvert coherently in a process known as quantum beating, which can be fine-tuned by weak magnetic interactions. The singlet and triplet fractions beat at frequencies that are characteristic

Department of Chemistry, University of Oxford, Oxford, UK.
Email: peter.hore@chem.ox.ac.uk

of the magnetic fields experienced by the electrons. One source of these fields is the hyperfine interactions with the spins of atomic nuclei, such as hydrogen and nitrogen. Another is the Zeeman interaction with externally applied magnetic fields.

Because the thermodynamic constraint mentioned above does not apply to radical pairs when the spins are far from equilibrium with their surroundings, weak magnetic interactions control the instantaneous probability of the radical pair being singlet or triplet, as well as the likelihood for the pair to react spin-selectively to form distinct singlet or triplet products. The yields of the two competing reaction pathways depend in a complex fashion on the timing of the coherent spin dynamics. A frustrating aspect of the radical pair mechanism is that it is rarely possible to observe these quantum beats directly. To do so, one would need to differentiate between the two states using, for example, ultrafast electronic spectroscopy (*6, 7*). However, this is normally impossible because singlet and triplet pairs have very similar energies, often within 10^{-6} eV, such that their absorption spectra are indistinguishable. It is easier to tell the reaction products apart, but because they are usually formed on a time scale much slower than that of singlet-triplet interconversion, any oscillations in their concentrations are seldom apparent.

The "pump-push" technique devised by Mims *et al*. avoids these obstacles by using laser pulses to provide snapshots of the spin state of the radical pair at different times after its creation. Radical pairs are produced, initially in a singlet state, by a short flash of light, known as the "pump," which causes an electron to jump from a donor molecule to an acceptor. Coherent spin motion ensues, driven by hyperfine interactions. After a variable delay, the radicals are subjected to a second short, strong laser flash with a longer wavelength—i.e., the "push" pulse. The result, for both singlet and triplet states, is an electronically excited and much more reactive radical pair, which undergoes rapid and spin-selective reverse electron transfer. If the "pushing" and the ensuing reactions are both faster than the spin motion, this effectively instantaneous sampling of the spin state of the radical pair would be detectable as an abrupt change in the absorption or fluorescence of the donor and/or acceptor. By repeating the measurement with different pump-push delays, one can map out the coherent spin dynamics.

The molecule selected for these proof-of-principle experiments comprised an aromatic tertiary amine electron donor and a perylene diimide electron acceptor separated by a dihydroanthracene bridge.

"The "pump-push" technique... avoids these obstacles by using laser pulses to provide snapshots of the spin state of the radical pair at different times after its creation."

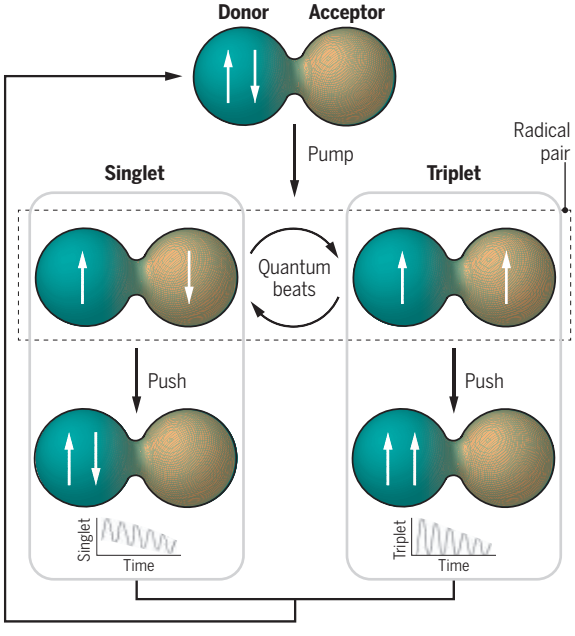
Although structurally complex, the radical pair has a beautifully simple spin system. The hyperfine fields that drive the coherent spin dynamics are dominated by a single nitrogen in the donor radical, with a smaller contribution from four identical hydrogens in the acceptor radical. This simplicity ensures a small number of quantum beat frequencies, which are more easily recorded than in the case of radicals with many coupled nuclei.

The pump-push method has many potential applications. It could be used to investigate photoactive proteins known as cryptochromes that, among other things, regulate plant growth and circadian rhythms. Many cryptochromes bind a flavin adenine dinucleotide cofactor, the excitation of which with blue light triggers three or four sequential electron transfers along a chain of tryptophan amino acid sidechains within the protein. The radical pair so formed shows pronounced magnetic sensitivity down to the millitesla range (*8, 9*) and has been proposed (*10, 11*) as the receptor that allows migratory songbirds to sense the direction of Earth's magnetic field (*12*). Direct observations of the coherent spin motion would reveal otherwise obscured information on the origin of the magnetic sensitivity. The pump-push method is also likely to find rewarding applications in studying magnetic field effects on electron-hole pairs in organic light-emitting diodes, whose magnetoresistance and magneto-electroluminescence can be explained by the radical pair mechanism (*13*).

Another potential application, relevant in the emerging field of quantum biology (*14*), might be to determine the extent to which radical pair magnetoreception is truly quantum. This question was recently approached by asking whether one can only account for the coherent spin dynamics in cryptochromes using quantum mechanics or whether a description using classical oscillations would suffice (*15*). A more satisfactory answer could be provided by measuring the quantum beats instead of inferring them from the yields of reaction products. ■

Observing quantum beats in the reactions of radical pairs

Coherent interconversion of the singlet (↑↓) and triplet (↑↑) states of radical pairs normally hidden from view. Mims *et al*. reveal these oscillations by using "pump" and "push" laser pulses to move an electron from one end of a donor-bridge-acceptor molecule to the other and back again.



GRAPHIC: N. DESAI/*SCIENCE*

REFERENCES AND NOTES

1. U. E. Steiner, T. Ulrich, *Chem. Rev.* **89**, 51 (1989).
2. A. R. Jones, *Mol. Phys.* **114**, 1691 (2016).
3. P. J. Hore, H. Mouritsen, *Annu. Rev. Biophys.* **45**, 299 (2016).
4. P. J. Hore, K. L. Ivanov, M. R. Wasielewski, *J. Chem. Phys.* **152**, 120401 (2020).
5. D. Mims, J. Herpich, N. N. Lukzen, U. E. Steiner, C. Lambert, *Science* **374**, 1470 (2021).
6. P. Gilch, F. Pollinger-Dammer, C. Musewald, M. E. Michel-Beyerle, U. E. Steiner, *Science* **281**, 982 (1998).
7. V. A. Bagryansky, V. I. Borovkov, Y. N. Molin, *Russ. Chem. Rev.* **76**, 493 (2007).
8. K. Maeda *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4774 (2012).
9. J. Xu *et al*., *Nature* **594**, 535 (2021).
10. T. Ritz, S. Adem, K. Schulten, *Biophys. J.* **78**, 707 (2000).
11. S. Y. Wong *et al*., *Neuroforum* **27**, 141 (2021).
12. H. Mouritsen, *Nature* **558**, 50 (2018).
13. T. Grünbaum *et al*., *Faraday Discuss.* **221**, 92 (2019).
14. Y. Kim *et al*., *Quantum Rep* **3**, 1 (2021).
15. T. P. Fay, L. P. Lindoy, D. E. Manolopoulos, P. J. Hore, *Faraday Discuss.* **221**, 77 (2019).

10.1126/science.abm9261

MEDICINE

Biological basis of cannabinoid medicines

Mechanistic insights into cannabinoid signaling could improve therapeutic applications

By **Erik Keimpema[1], Vincenzo Di Marzo[2,3], Tibor Harkany[1,4]**

Since its first mention ~4000 BCE, *Cannabis sativa* has evolved through selective cultivation from being a source of durable fiber (hemp) to a plant enriched in bioactive ingredients. Currently, >100 potentially bioactive phytocannabinoids from *Cannabis* spp. have been cataloged, yet their precise structure-function relationships are mostly unclear (*1*). Δ^9-Tetrahydrocannabinol (THC) and cannabidiol (CBD) are primarily studied, particularly because high-grade *Cannabis* subspecies can produce over 20% yield of either compound. The variety of bioactive constituents in *C. sativa*, together with their defined ratios, suggests that they have potential application in many illnesses (*1*). Possibly due to many phytocannabinoids producing similar pharmacological effects through different mechanisms, selecting which to study for a disease remains a formidable challenge.

THC action in humans is dependent on the CB_1 cannabinoid receptor (CB_1R) (*2*), which is the most abundant G protein–coupled receptor (GPCR) in the brain, as well as its ortholog, CB_2R. When activated, CB_1R in the plasma membrane signals through $G_{(i/o)}$ proteins to inhibit either $G\alpha_i$-mediated SRC–signal transducer and activator of transcription (STAT) or $G\beta\gamma$-mediated adenylyl cyclase, AKT, or extracellular signal–regulated kinase (ERK) cascades. For excitable cells, such as neurons, activated CB_1R inhibits Ca^{2+} influx through voltage-gated Ca^{2+} channels, thus reducing neurotransmitter release.

Although most studies focus on THC because of its psychostimulant effects, CBD is another abundant (up to 40 to 50% of total phytocannabinoid content) and yet non-psychotropic *Cannabis* component. CBD is thought to have anti-inflammatory and tissue-protective effects. This is because CBD action is putatively mediated by more than one receptor, including transient receptor potential cation channel V1 (TRPV1), GPCR *55* (GPR*55*), and peroxisome proliferator-activated receptor–γ (PPARγ). Additionally, CBD enhances antioxidant cellular defenses by scavenging hydroxyl radicals and can counteract THC action intracellularly, through CB_1Rs on mitochondrial membranes (*3*). Indeed, CBD opposes the THC-induced disruption of oxidative phosphorylation, at the level of complex I (*3*), thereby protecting from the deleterious consequences of THC-induced reduction in cellular respiration. Notably, these potential cellular and molecular sites of action are not limited to the brain but apply to, e.g., pancreas, muscle, liver, and gut—suggesting that cannabinoids may have applications in diverse settings.

It is important to recognize that *C. sativa* is more than just THC and CBD. Many phytocannabinoids that exist in lesser amounts (e.g., cannabivarin, cannabigerol, and THC acid) (*4*) in plant preparations could individually be biologically powerful and even supersede or modify THC and/or CBD action. Thus, the continued structure-function study of phytocannabinoids is warranted. Accordingly, combinations of phytocannabinoids might deliver relief to disorders with complex etiology (*1*).

Receptor-mediated actions of phytocannabinoids, particularly THC, center on displacing high-affinity endocannabinoids, innate ligands that bind CB_1R, CB_2R, and alternative receptors including TRPV1, GPR*55*, and PPARγ. Two such molecules, 2-arachidonoylglycerol (2-AG) and arachidonoylethanolamide (anandamide), with their respective biosynthetic and catabolic enzyme machineries, form the molecular backbone of the endocannabinoid system (*5, 6*). 2-AG and anandamide are functionally redundant. Yet, endocannabinoid signaling can adopt cell-type–specific configurations to support cell-autonomous (e.g., self-inhibition in neurons), intercellular (e.g., metabolic interplay), and intracellular (e.g., cellular respiration through CB_1R on mitochondria) signaling in the brain and in peripheral tissues. Such functional flexibility is made possible by the different half-lives (minutes versus hours) and tissue distributions of 2-AG and anandamide, together with the diversity of available receptors. These arrangements allow for systemic actions of endocannabinoids and their modulation by THC and CBD (*1*).

An expanding catalog of endocannabinoid-like molecules are being recognized for their modulation of TRP channels, PPARs, and orphan GPCRs. This "endocannabinoidome" offers a more comprehensive physiological substrate for phytocannabinoid action than the core endocannabinoid system. Moreover, the array of endogenous CB_1R and CB_2R ligands now includes allosteric modulators derived from steroids and hemoglobin fragments, which can antagonize THC intoxication (*7*) and reduce neuropathic pain (*8*), respectively.

Functional interrogation of 2-AG action on CB_1Rs in the brain revealed that endocannabinoids modulate synaptic plasticity by limiting Ca^{2+}-dependent neurotransmitter release (*9*). This mechanism relies on CB_1Rs partitioned at presynaptic termini with endocannabinoid ligand production in postsynaptic neurons. Because endocannabinoids are eicosanoid lipids that travel by nonvesicular diffusion, their activity-dependent production and fast degradation ensure phasic availability and short-lived action. Accordingly, neuropsychiatric effects of THC-containing *Cannabis* preparations (e.g., euphoria, hypomotility, and amnesia) occur through CB_1R hyperactivity at the plasma membrane and intracellularly, which disrupts temporal precision and prolongs endocannabinoid-dependent synaptic inhibition. In the adult brain, the near-complete recovery of endocannabinoid action (and reinstatement of synaptic neurotransmission) may take hours to weeks after THC consumption. Yet, given that only a small fraction of first-time consumption progresses to chronic heavy use, *Cannabis* is classified as moderately addictive in adults.

A prominent clinical niche being explored for cannabinoid medicines is treatment-resistant epilepsy (*10*). Phytocannabinoids that trigger CB_1R-mediated signaling at glutamatergic synapses dampen neuronal hyperexcitability in epileptic foci in rodents and humans. Repeated epileptic seizures provoke neurodegeneration in, e.g., the hippocampus. Therefore, although still speculative, rescuing neurons from oxidative damage and mitochondrial dysfunction by, e.g., CBD, could partly underpin the substantial reduction in seizure frequency seen in clinical trials (*10*).

Phytocannabinoids may also have applications in diseases associated with aging. This is largely because aging tissues contain increased amounts of immune cells, which remove cellular debris or extracellular (proteinaceous) deposits. THC-induced activation of CB_2Rs during aging may offer relief in neu-

[1]Department of Molecular Neurosciences, Center for Brain Research, Medical University of Vienna, Vienna, Austria. [2]Institute of Biomolecular Chemistry, Consiglio Nazionale delle Ricerche, Pozzuoli, Italy. [3]Institut Universitaire de Cardiologie et de Pneumologie de Québec and Institut sur la Nutrition et les Aliments Fonctionnels-Centre NUTRISS, Université Laval, Québec, Canada. [4]Department of Neuroscience, Karolinska Institutet, Solna, Sweden. Email: tibor.harkany@ki.se

rological disorders (Alzheimer's, Parkinson's, and Huntington's diseases) by limiting inflammatory cytokine release from activated microglia, which can be harmful to neurons (see the figure). Alternatively, THC can dampen errant synaptic neurotransmission by CB_1R activation, which might improve cognition. Long-term, low-dose THC treatment in a mouse model of Alzheimer's disease rescued memory deficits, neuronal morphology, and aberrant gene transcription (*11*). This duality of correcting synaptic neurotransmission and protecting neurons from degeneration is a typical example of cannabinoid polypharmacology: beneficial effects at more than one cellular target (*1*). This also exemplifies biphasic responses to cannabinoids, manifesting as beneficial effects at low concentrations and harmful effects at high concentrations. Apart from the multitarget nature of these compounds, these biphasic effects may depend on biased cannabinoid action at different CB_1R populations and, hence, on the baseline (patho) physiological substrate these molecules act on. A further obstacle to the therapeutic use of THC per se is that CB_1Rs, unlike CB_2Rs, often exacerbate inflammation (*1*).

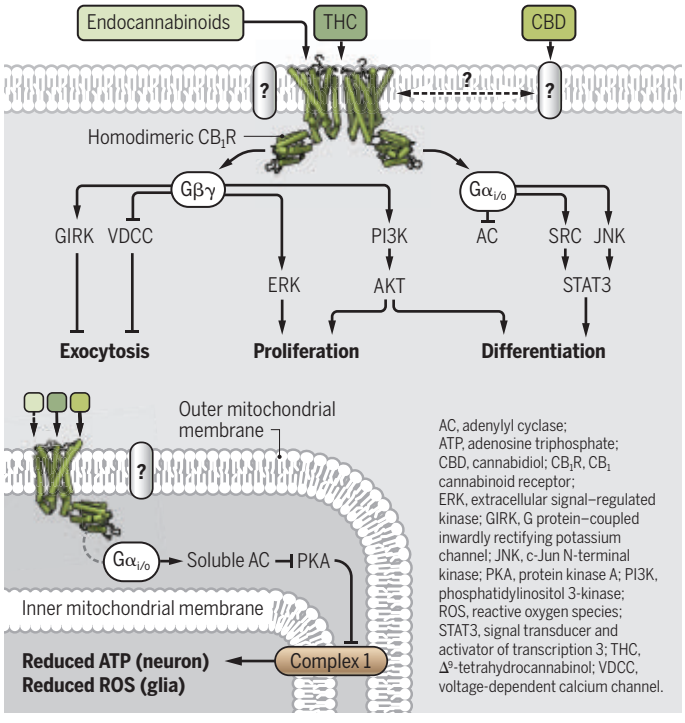
Cannabinoid therapy is also potentially applicable for the management of neuropathic pain, induced by a lesion or disease of the somatosensory nervous system, because dampening excitatory neurotransmission at the level of spinal neurocircuits can reduce hyperalgesia in rodents (*12*). Another druggable target emerges in inflammatory pain because of peripheral sensitization in the skin, where accumulation of anandamide facilitates pain processing and proinflammatory signaling through TRPV1 on primary sensory afferents (Aδ and C fibers). TRPV1 activation increases the transcription of nerve growth factor receptors [tropomyosin receptor kinase A (TrkA) and p75], whose activation augments pathological touch sensitivity. Thus, inactivating TRPV1 by sequential activation and desensitization, as some phytocannabinoids do (*1*), might be medically relevant.

In contrast to the transient and reversible effects of THC on synaptic neurotransmission at mature synapses, the situation is different during pre- and postnatal brain development. The precisely timed activation of CB_1R, CB_2R, and likely GPR*55* is critical for cell fate decisions during organ development. In the brain, cell-autonomous signaling contributes to neurite outgrowth (*13*). Alternatively, intercellular endocannabinoid action determines the size of neural stem cell pools and lineage commitment of daughter cells, their migration, synaptogenesis, and synapse maintenance in vertebrates. Thus, the coincidence of THC exposure with these processes, which occur in the human brain from the second trimester during pregnancy until late adolescence, can imprint adverse and life-long modifications on the structural integrity of the brain. Accordingly, exposure of (pre-) adolescent children to *Cannabis* is associated with an increased number of hospitalizations for neurological complications (*14*).

Like CB_2R on neural stem cells, the activation of CB_1R in adipocytes, TRPV1 in pancreas, or GPR*55* in skin and salivary gland defines tissue size by modulating the rate of cell proliferation during organ development. Thereafter, CB_1R (or TRPV1) activation in differentiating progeny also determines cell survival. These findings support the exploration of phytocannabinoid-based therapy for disorders in which errant cell-cycle regulation is pathogenic, including cancer. A leading concept is that possible antitumoral effects of THC (and likely CBD) may involve increased ceramide production, triggering autophagy-mediated cancer cell death (*15*).

Cannabinoid signaling

Endocannabinoids signal through G protein–coupled receptors to control exocytosis, proliferation, differentiation, and respiration. In disease, phytocannabinoids (only THC and CBD are shown for simplicity) can affect CB_1R signaling, although the mechanisms are not fully elucidated. Many of these signaling principles could also apply to CB_2R which, for example, regulates cytokine release.



Increasing knowledge of endocannabinoid mechanisms and *Cannabis* constituents has led to the development of synthetic cannabinoids, encompassing THC analogs and structurally unrelated compounds such as high-affinity and selective CB_1R antagonists and inhibitors of a variety of lipases and hydrolases that catalyze endocannabinoid synthesis and inactivation. Some of these synthetic ligands have already entered clinical practice (e.g., rimonabant, nabilone, orlistat) or are in clinical trials (e.g., ABX-1431). Thus, the expanding repertoire of drugs targeting the endocannabinoid system and the endocannabinoidome is of broad therapeutic appeal.

Given the abundance and subcellular partitioning of CB_1R and CB_2R and the endocannabinoid enzymatic machinery, organ-specific targeting by synthetic ligands remains a key pharmacological challenge. Dissecting how subcellular pools of CB_1R bring about differential drug action and define functional outcome will be key to improving disease-specific applications. Similarly, delivering single components versus phytocannabinoid mixtures requires decisions on targeting specific signaling pathways rather than harnessing the pleiotropy and synergy of coadministered phytocannabinoids (*1*). Nonetheless, the expanding knowledge of the endocannabinoid system is lifting phytocannabinoids from fringe utilization to potentially safe and effective medicines in adults. ■

REFERENCES AND NOTES

1. J. S. Brodie *et al.*, *Trends Pharmacol. Sci.* **36**, 802 (2015).
2. W. A. Devane *et al.*, *Mol. Pharmacol.* **34**, 605 (1988).
3. D. Jimenez-Blasco *et al.*, *Nature* **583**, 603 (2020).
4. S. E. Turner *et al.*, *Prog. Chem. Org. Nat. Prod.* **103**, 61 (2017).
5. W. A. Devane *et al.*, *Science* **258**, 1946 (1992).
6. R. Mechoulam *et al.*, *Biochem. Pharmacol.* **50**, 83 (1995).
7. M. Vallée *et al.*, *Science* **343**, 94 (2014).
8. E. F. Toniolo *et al.*, *Peptides* **56**, 125 (2014).
9. N. Stella *et al.*, *Nature* **388**, 773 (1997).
10. E. A. Thiele *et al.*, *JAMA Neurol.* **78**, 285 (2021).
11. A. Bilkei-Gorzo *et al.*, *Nat. Med.* **23**, 782 (2017).
12. A. Calignano *et al.*, *Nature* **394**, 277 (1998).
13. E. Keimpema *et al.*, *J. Neurosci.* **30**, 13992 (2010).
14. M. Di Forti *et al.*, *Lancet Psychiatry* **2**, 233 (2015).
15. M. Salazar *et al.*, *J. Clin. Invest.* **119**, 1359 (2009).

ACKNOWLEDGMENTS

The authors are supported by the European Research Council (ERC-2015-AdG-695136; T.H.), the Austrian Research Fund (P 34121-B; E.K.), and the Tri-Agency of the Canadian Federal Government (CERC programme, Canada Foundation for Innovation Leaders Fund, and Sentinelle Nord-Apogée programme; V.D.).

10.1126/science.abf6099

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

BOOKS *et al.*

In the absence of national curriculum standards, some students may never learn about climate change.

EDUCATION

Climate change in the classroom

Inconsistent and agenda-driven K–12 curricula leave US citizens ill-equipped to confront environmental problems

By **Ashley Huderson**

"How you gonna win when you ain't right within," sings Lauryn Hill in "Doo Wop (That Thing)," a track that appears on her 1998 debut solo album, *The Miseducation of Lauryn Hill*. This lyric singularly captures the essence of Katie Worth's new, similarly titled book, *Miseducation: How Climate Change Is Taught in America,* which provides a detailed account of the intersection between education, politics, policies, personal beliefs, and the power of information and argues that US public school curricula leave the country's citizens ill-prepared to win the fight against climate change.

Worth is a US journalist who has spent her 15-year career winning awards for both breaking news and long-form investigative narratives. With *Miseducation*, she uses her journalistic skills to examine the evolution of the climate change discussion over time. She provides evidence from all sides of the debate, describing teachers who unsuccessfully walk the line between curriculum requirements and their beliefs and policy leaders who have personal and professional stakes in the issues at hand.

In the US, there are about 50 million children enrolled in 100,000 public schools who are taught by 3 million teachers. However, there are no national standards for a curriculum on climate change. In her research for this book, Worth assembled a state-by-state database of various curricula and reviewed dozens of textbooks used in school systems across the country.

Despite an overwhelming scientific consensus that climate change is "real, it's us, it's bad, and there's hope," Worth writes, the US has created an educational environment in which "children in some places are required by law to learn about the phenomenon… while in others, students may not hear the words 'climate change' in class at all."

Worth writes of an Advanced Placement science teacher in Oklahoma who refuses to teach anthropogenic climate change because her family is in the oil and gas business. Another teacher, who does not believe in climate change, chooses to discuss the topic but refuses to cite any scientific data and laces his teaching with his own personal beliefs, casting doubt in the young minds he has been charged with molding.

Many districts and states forbid the inclusion of climate change in the curriculum. These states are filled with teachers who cast doubt on the concept of human-induced climate change, textbook publishers eager to avoid upsetting school boards, and editors who rewrite commissioned science pieces to fit political formulas. "These patterns are no accident of history," Worth concludes. "Rather, they are the product of successful disinformation campaigns, animated not by science but by ideology."

Such practices create adults who struggle to understand the current state of affairs surrounding climate change and have resulted in a political divide, argues Worth. "Classrooms have emerged as a battleground in the American political war over climate change because what kids learn about climate change now will directly impact the speed and ambition of action taken for decades to come," she notes.

Worth attributes the origins of the climate debate to long-standing tensions between science and religious fundamentalism, tracing the roots of these tensions back to Darwin's theory of evolution, which fundamentally shifted how we view the environment and our role in it. Demonstrating how people from different backgrounds can be presented with the same facts and interpret them differently, she reveals how variations in our worldviews further expand the political divide, playing out in how we vote and who we vote for.

Miseducation: How Climate Change Is Taught in America
Katie Worth
Columbia Global Reports, 2021. 184 pp.

Miseducation is a cautionary tale of the wide-ranging impacts that political agendas can have when deployed in educational settings. Taking lessons from earlier debates over evolution and tobacco, oil corporations, state legislatures, school boards, think tanks, lobbyists, and textbook publishers are now sowing uncertainty, confusion, and distrust about climate science. Until we confront these wrongs within the US educational system, the battle to mitigate climate change will never be won. ■

The reviewer is at the American Society of Mechanical Engineers, Washington, DC 20036, USA; the Department of Biology, University of the District of Columbia, Washington, DC 20008, USA; and STEM Innovation Consulting, Washington, DC 20018, USA.
Email: ahuderson@steminnovationconsulting.com

10.1126/science.abl9313

PHOTO: ALEKSEI GORODENKOV/ALAMY STOCK PHOTO

SOIL SCIENCE

Searching for solutions to our soil woes

Could a controversial carbon storage plan help restore degraded lands?

By **Daniel D. Richter**

Poets, philosophers, farmers, and scholars have long recognized the mutual dependency of human beings and soils. We are utterly reliant on soil for food and fiber, water purification, carbon sequestration, infrastructure support, and even new drugs. That soil is dependent upon people may first have been asserted in *The Georgics* by Virgil, who framed soils and all of nature as being highly vulnerable to human action and in need of human care. Virgil celebrated the endless work required for sustaining soil and agriculture—work needed to stave off a degraded world where we might find ourselves shaking oak trees for acorns, "frantic for something to eat" (*1*).

A georgic ethos with the land resonates today, as the world's farmers work with soil to produce nearly 95% of the food supply for 8 billion human beings. To meet these times, microbial scientist Jo Handelsman, who served as a White House science adviser during the Obama administration, has written *A World Without Soil*, which presents a manifesto for improved soil conservation and management.

With more than half of the world's soils actively managed, Handelsman raises serious concerns about the many ways that soils are being degraded more rapidly than they can be formed. Her goal is to promote new soil policy that can help reverse the course of soil loss.

A World Without Soil is well written, even eloquently so. "Beneath the bustle of cities, towns, farms, forests, and highways lies the silent, dark ribbon of life, rock, and water that binds the past and future," reads one evocative passage. But it is the need for new soil management that drives the book, specifically the need to increase soil organic carbon and control soil erosion.

Handelsman's book adds to the literature warning of soil crises and urging reform, which includes the voices of G. P. Marsh, E. B. Balfour, H. H. Bennett, S. W. Trimble, and D. R. Montgomery (*2*–*6*). Like these authors, Handelsman emphasizes the precarious vulnerability of human-soil relations.

The reviewer is at the Nicholas School of the Environment, Duke University, Durham, NC 27708, USA, and coauthor (with Daniel Markewitz) of *Understanding Soil Change* (Cambridge Univ. Press, 2001). Email: drichter@duke.edu

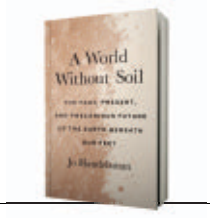
What distinguishes Handelsman from her predecessors is her optimism about our ability to reverse the course of soil loss. While recognizing that soil degradation is a challenging problem long in the making, Handelsman considers it "one that can be remedied quickly" and "with relatively little short-term cost." Handelsman's positive outlook is based on her confidence that new policy can stimulate soil-carbon storage that will mitigate climate change and have co-benefits that improve soil health and fertility.

Nearly 10,000 years of agriculture have reduced soil carbon by more than 100 Pg (*7*), a loss that has compromised the functioning of many soils. Because this loss of soil carbon is equal to about a decade of current CO_2 emissions from fossil fuel combustion, if the world's soils could sequester a fraction of the carbon they have lost, Handelsman and others reason that this could provide more time for societies to reduce CO_2 emissions from combustion of fossil fuel. Handelsman thus sees improved soil management as "a powerful mitigative tool to address climate change" and strongly supports international efforts (*8*) to modify agricultural practices "to advance the world's climate mitigation and improve soil."

Many soil scientists, agronomists, and carbon scientists suggest that such massive initiatives are impractical and overly optimistic and lack a basis in the science of soil carbon dynamics (*9*–*11*). Handelsman acknowledges these criticisms but is not dissuaded.

Ultimately, Handelsman has crafted a book for a broad audience that will widen discussion and interest in soils and soil degradation. Perhaps the book can also spur broader science and policy discussion around whether soil policy should be used to manage the global carbon cycle. Handelsman's bold linkage of climate and soil will need much more robust examination in preparation for what may be some of the most important land-management policy decisions of our time. ■

A World Without Soil: The Past, Present, and Precarious Future of the Earth Beneath Our Feet
Jo Handelsman
Yale University Press, 2021. 272 pp.



REFERENCES AND NOTES

1. D. Ferry, Transl., *The Georgics of Virgil* (Farrar, Straus and Giroux, 2004).
2. G. P. Marsh, *Man and Nature* (Scribner, 1864).
3. E. B. Balfour, *The Living Soil* (Faber and Faber, 1943).
4. H. H. Bennett, *Our American Land* (US Department of Agriculture, 1946).
5. S. W. Trimble, *Man-Induced Soil Erosion on the Southern Piedmont* (Soil and Water Conservation Society, 1974).
6. D. R. Montgomery, *Dirt: The Erosion of Civilizations* (Univ. of California Press, 2007).
7. J. Sanderman, T. Hengl, G. J. Fiske, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9575 (2017).
8. B. Minasny *et al.*, *Geoderma* **292**, 59 (2017).
9. R. Amundson, L. Biardeau, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11652 (2018).
10. W. H. Schlesinger, R. Amundson, *Global Change Biol.* **25**, 386 (2019).
11. T. Searchinger, J. Ranganathan, "INSIDER: Further explanation on the potential contribution of soil carbon sequestration on working agricultural lands to climate change mitigation" (World Resources Institute, 2020).

10.1126/science.abm4765

PODCAST

Best books of 2021

A playful immunology primer packed with vivid illustrations, a myth-busting reappraisal of early human history, a fact-filled romp through human-wildlife conflict, and a timely critique of the state of modern physics are among our favorite science books published in 2021. This week on the podcast, *Science*'s book review editor, Valerie Thompson, joins host Sarah Crespi to discuss the books we reviewed and loved this year. https://scim.ag/3CXXO2w

10.1126/science.abn2905

PHOTO: SERPHOTO/SHUTTERSTOCK

Research Funding Available

Up to $1M per lab
Proposals due March 1, 2022

Ionis is funding investigators with extraordinarily original ideas to identify next-wave technologies for healthcare through its Ion-ARPA program. The current program announcement **"Novel Therapeutic Payloads"** is for proposals to develop non-conventional technologies to modulate the expression of mammalian genes to overcome a diseased state. General areas of interest include, but are not limited to, novel strategies for silencing or activating genes and gene editing beyond existing CRISPR/Cas9- or deaminase-based approaches.

TO ACCESS MORE INFORMATION SEE:
www.ionispharma.com/ionis-innovation/ion-arpa-initiative/

IONIS®

Partner with the global leader involved in over 50% of COVID-19 testing to help solve the most intractable crisis we face today.

We offer:

- **Accuracy and reliability**—advanced assay designs compensate for current and emerging mutations like the Delta (B.1.617.2) and Alpha (B.1.1.7) variants
- **Service partnerships**—help ensure minimal downtime, allowing operations to continue running at their peak
- **Single-source convenience**—one supplier reliably providing all consumables, instruments, and servicing to enable uninterrupted testing

Keep moving forward in the fight against COVID-19 and learn about our offerings by scanning the QR code above.

Get a free infographic on lab-based and rapid PCR testing at
thermofisher.com/covid19scienceinfographic

ThermoFisher SCIENTIFIC

© 2021 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **COL25296 0721**

LETTERS

Edited by **Jennifer Sills**

Editorial Expression of Concern

On 21 July 2017, *Science* published the Report "Chiral Majorana fermion modes in a quantum anomalous Hall insulator–superconductor structure" by Q. L. He *et al.* (*1*). Since that time, raw data files were offered by the authors in response to queries from readers who had failed to reproduce the findings. Those data files did not clarify the underlying issues, and now their provenance has come into question. While the authors' institutions investigate further, we are alerting readers to these concerns.

H. Holden Thorp
Editor-in-Chief

REFERENCES AND NOTES

1. Q. L. He *et al.*, *Science* **357**, 294 (2017).

10.1126/science.abn5849

Track Omicron's spread with molecular data

On 26 November, the newly emerged variant Omicron was designated a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant of concern (VOC) (*1*). Rapid polymerase chain reaction (PCR) test results could improve estimates of the prevalence of Omicron around the world. The widely used Thermo Fisher TaqPath COVID-19 PCR assay was valuable in tracking the spread of the Alpha (B.1.1.7) VOC (*2*) because a deletion of amino acids 69 and 70 in Alpha's spike gene (Δ69–70) yields a distinct absent S-gene (S^-) despite positive test results. The Delta VOC lacks this deletion and is therefore S-gene positive (S^+) on TaqPath PCR tests (*3*). The Omicron VOC shares the spike Δ69–70 deletion with Alpha, which has dropped to negligible levels worldwide. Therefore, the frequency of S^- results can be used as a rapid proxy for the frequency of Omicron cases, provided initial detection of local circulation had been confirmed by sequencing.

To put these data to use, countries should prioritize the release of daily counts of cases, hospitalizations, and deaths disaggregated by S^+, S^-, and unknown [e.g., (*4*)] as much as possible while taking logistical and privacy concerns into account. S-gene data could serve as a proxy for estimates of Omicron VOC prevalence (*5*) and help us to understand the fraction of infections caused by Omicron (versus Delta) and the severity of Omicron cases, as measured by mortality and hospitalization. In low-resource settings where genomic sampling is absent, infrequent, or characterized by long turnaround times (*6*), S-gene data will help reveal the risk Omicron poses to pandemic control. Finally, through synthesis with serological data (*7*), S-gene data—shared in real time—could help to evaluate the degree of immune protection conferred by natural- and vaccine-elicited immunity in Omicron cases.

Although S-gene data will be informative, preferential sequencing of samples with an S^- result will lead to virus genomic datasets that are unrepresentative of the true underlying spatiotemporal prevalence of Omicron. To provide adequate context for genome sequences, depositors to the Global Initiative on Sharing All Influenza Data (GISAID) database should use the newly introduced nonmandatory "sampling strategy" field to note how cases are selected and sampled for virus genome sequencing, including whether samples were specifically targeted for sequencing based on S^- PCR results. [We have used this field to plot the first 115 Omicron submissions to GISAID, stratified by sampling strategy (*8*).] Virus genomic datasets then can be compiled from cases known to have been sampled randomly from a given population and analyzed to generate more-accurate estimates of Omicron's growth relative to other variants. Standard sampling strategies include random community sampling [the preferred sampling strategy for estimating lineage growth (*6*, *9*)], targeted surveillance of defined subpopulations (e.g., vaccine breakthrough cases or international travelers), and enhanced sampling to investigate specific outbreaks or clusters.

Tracking SARS-CoV-2 lineages and variants, including Omicron, through GISAID (*10*), Pango lineages (*11*), and NextStrain (*12*) has provided valuable information about their spread in close to real time. However, genome sequencing intensities and turnaround times vary substantially across the world; in most countries, it takes more than 21 days after sample collection to deposit data in GISAID (*6*). Moreover, sampling strategies used to select samples for sequencing are heterogeneous across geographic regions (*6*) and often not reported in virus genome metadata. To evaluate risk and guide policy, there is an urgent need to incentivize the quick sharing of well-annotated genomic and S-gene–stratified surveillance data globally. By acting with speed, transparency, and consistency, we can establish

Polymerase chain reaction testing could provide rapid insights into the spread of the COVID-19 Omicron variant.

PHOTO: HANNAH BEIER/BLOOMBERG/GETTY IMAGES

norms to support better global responses to newly emerging variants.

Lesley Scott[1], Nei-yuan Hsiao[2,3], Sikhuline Moyo[4], Lavanya Singh[5,6], Houriiyah Tegally[5,6], Graeme Dor[1], Piet Maes[7], Oliver G. Pybus[8,9], Moritz U. G. Kraemer[8], Elizaveta Semenova[10], Samir Bhatt[11,12,13], Seth Flaxman[14], Nuno R. Faria[8,12,13,15], Tulio de Oliveira[5,6]*

[1]Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa. [2]Division of Medical Virology, Wellcome Centre for Infectious Diseases in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa. [3]National Health Laboratory Service, Cape Town, South Africa. [4]Botswana Harvard AIDS Institute Partnership and Botswana Harvard HIV Reference Laboratory, Gaborone, Botswana. [5]Centre for Epidemic Response and Innovation, Stellenbosch University, Stellenbosch, South Africa. [6]KwaZulu-Natal Research Innovation and Sequencing Platform, University of KwaZulu-Natal, Durban, South Africa. [7]Laboratory of Clinical and Epidemiological Virology, Katholieke Universiteit Leuven, Leuven, Belgium. [8]Department of Zoology, University of Oxford, Oxford, UK. [9]Royal Veterinary College, London, UK. [10]Department of Mathematics, Imperial College London, London, UK. [11]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [12]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, UK. [13]The Abdul Latif Jameel Institute for Disease and Emergency Analytics, School of Public Health, Imperial College London, London, UK. [14]Department of Computer Science, University of Oxford, Oxford, UK. [15]Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.
*Corresponding author. Email: tulio@sun.ac.za

REFERENCES AND NOTES

1. World Health Organization (WHO), "Classification of Omicron (B.1.1.529): SARS-CoV-2 variant of concern" (2021); www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern.
2. E. Volz *et al.*, *Nature* **593**, 266 (2021).
3. S. Mishra *et al.*, *EClinicalMedicine* **39**, 101064 (2021).
4. Office for National Statistics, "Coronavirus (COVID-19) Infection Survey, UK Statistical bulletins" (2021); www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases.
5. A. J. Kucharski, E. B. Hodcroft, M. U. G. Kraemer, *Lancet Reg. Health Eur.* **9**, 100215 (2021).
6. A. F. Brito *et al.*, *medRxiv* 10.1101/2021.08.21.21262393 (2021).
7. R. K. Arora *et al.*, *Lancet Infect. Dis.* **21**, e75 (2021).
8. "COVID-19 Omicron Repository," GitHub (2021); https://github.com/CADDE-CENTRE/covid19omicron.
9. WHO, "Genomic sequencing of SARS-CoV-2: A guide to implementation for maximum impact on public health" (2021); www.who.int/publications/i/item/9789240018440.
10. Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017).
11. A. O'Toole, "Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2," Virological.org (2021); https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592.
12. J. Hadfield *et al.*, *Bioinformatics* **34**, 4121 (2018).

Published online 9 December 2021
10.1126/science.abn4543

PHOTO: SUSIE GALLEGOS

News stories must account for gender bias

The *Science*Insider piece "Max Planck director loses post after probe of misconduct" (A. Curry, 5 November, p. 671) is the latest in a series of reports of dismissals, demotions, and conflicts involving prominent women in academic research. This story and previous reports have highlighted leadership issues and bullying by women not only at the Max Planck Institute (*1*) but also in top academic positions at ETH Zurich (*2*) and the University of London (*3*). We urge caution

PAST AS PROLOGUE

A uranium miner's daughter

After serving in Vietnam, my dad moved to Grants, New Mexico, to mine uranium. Every day, he drilled out uranium in deep, poorly ventilated, confined, hot, and dangerous underground tunnels. After work, my mom washed his overalls and lunch bucket, soiled with radioactive dirt. One day, when I was in fourth grade, my dad came home early from the graveyard shift and said he was not going back. At the time, I did not understand the circumstances, but I later learned that the industry had collapsed due to declining uranium prices, leaving the local economy in shambles.

Uranium mining has always been controversial. Uranium fuels non–carbon-emitting nuclear energy, but uranium and its radioactive decay progeny may pose health concerns. Even so, my family is proud of my dad's work in the mines because it afforded my parents a livelihood and the means to send their three kids to college, a luxury not given to them. With that opportunity, I pursued degrees in environmental engineering. For my PhD, I moved to Michigan to study iron sulfide–based media for use in cleanup of arsenic-contaminated groundwater. It was a difficult transition moving from sunny New Mexico to the cold, snowy upper Midwest, devoid of blue skies, New Mexico green chile sustenance, and, most importantly, my family. I began to question why I had started down this road, so far away from home without a clear vision of my destination.

The author's father, shown here, mined uranium in Grants, New Mexico.

Fortunately, a series of events allowed me to see the horizon. While on a summer research fellowship in Korea in 2006, I stumbled upon one of the only books in English in the institute's library. It was about the Grants mineral belt. I was amazed to see a book about my hometown halfway around the world. Soon after, I began seeing articles in the Grants newspaper about contamination from former uranium extraction operations, more than three decades after their closure.

I realized that I could apply my expertise to research uranium! The following year, I accepted a Mendenhall postdoc position at the US Geological Survey to study the environmental impacts of uranium mining. Now, with over 14 years of uranium research stimulating my curiosity, I have returned to New Mexico seeking new insights for managing mine waste. It is fitting that my passion for science brought me home again, where it was nurtured from the beginning by a humble, hard-working uranium miner and his wife.

Tanya J. Gallegos
Mineral Resources Program, US Geological Survey, Albuquerque, NM 87113, USA.
Email: tgallegos@usgs.gov

10.1126/science.abn0706

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.

in reporting such stories given that women face more obstacles to attaining leadership positions and are often held to a different standard than men when it comes to how their behavior is interpreted.

Professional women face many biases that disproportionately delay their advancement along the career track and compromise their effectiveness and even their tenure in positions of power and authority (*4*, *5*). The same biases may result in greater and more detrimental visibility for conflicts involving women. Although gender bias in individual institutions can be difficult to assess because of the small numbers of women in leading scientific positions, the Max Planck Society has a large number of directors. It would be useful to know how many Max Planck Institute directors have left their positions before retirement, whether women are disproportionately represented in this group, and whether the publicity accompanying the departures differed between men and women.

To provide fair coverage, news stories should always pursue the question of possible gender bias, both in the treatment of women in positions of academic leadership and in the reporting on cases of leadership conflicts.

Janet G. Hering[1,2,3]*, Roberta Croce[4], Beate I. Escher[5], Anne E. Magurran[6], Beatriz Noheda[7]
[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland. [2]ETH Zürich, Zürich, Switzerland. [3]EPFL, Lausanne, Switzerland. [4]Department of Physics and Astronomy, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [5]Department of Cell Toxicology, Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany. [6]Centre for Biological Diversity and Scottish Oceans Institute, School of Biology, University of St. Andrews, St. Andrews, UK. [7]Zernike Institute for Advanced Materials, Groningen, Netherlands.
*Corresponding author.
Email: janet.hering@eawag.ch

The authors are members of *Science*'s Board of Reviewing Editors.

REFERENCES AND NOTES

1. K. Kupferschmidt, *Science* 10.1126/science.aav0199 (2018).
2. G. Vogel, *Science* 10.1126/science.aav9167 (2018).
3. *Science* **361**, 734 (2018).
4. A. S. Kramer, A. B. Harris, *It's Not You, It's the Workplace: Women's Conflict at Work and the Bias That Built It* (Mobius, 2019).
5. J. Gillard, N. Okonjo-Iweala, *Women and Leadership: Real Lives, Real Lessons* (MIT Press, 2021).

10.1126/science.abn2559

Editor's Note

We thank J. G. Hering and colleagues for raising important questions about gender bias in cases of alleged bullying at the Max Planck Society and elsewhere. We agree that those issues need to be explored. In our coverage of Nicole Boivin, including a follow-up article published on 6 December (*1*), we note accusations of institutional misogyny at Max Planck and the small proportion of women directors. Our piece on the case at ETH Zurich devoted several paragraphs to the scarcity of women on the physics faculty and possible double standards for judging the behavior of women and men. Because bullying can impact the well-being and careers of young researchers, we believe serious cases should be covered regardless of who is accused—as our stories over the years have shown.

Tim Appenzeller
News Editor

REFERENCES AND NOTES

1. A. Curry, Science 10.1126/science.acx9778 (2021).

10.1126/science.abn5820

International Human Frontier Science Program Organization

CALL FOR LETTERS OF INTENT FOR RESEARCH GRANTS:

AWARD YEAR 2023

Initiation deadline: 24 March 2022
Submission deadline: 31 March 2022

The Human Frontier Science Program (HFSP) supports innovative basic research that applies novel and interdisciplinary approaches to understand fundamental biological problems and includes scientific exchanges across national and disciplinary boundaries. Since 1990, over 7000 researchers from over 70 countries have been supported, 28 HFSP awardees have gone to receive the Nobel Prize.

HFSP research projects extend the frontiers of knowledge. Successful applications will entail risk and aim to develop novel lines of research different from the applicants' ongoing work. The participation of scientists from disciplines outside the traditional life sciences such as chemistry, engineering, mathematics, nanoscience or physics is a key requisite in HFSP grant applications.

To stimulate novel, daring ideas and innovative approaches, preliminary results are not required. Applicants are expected to develop new lines of research through the collaboration.

Awards are for 3 years and made to international (preferably intercontinental) teams of 2 – 4 members who have not collaborated before. **Research Grants – Early Career** are for teams of researchers who are all within 5 years of establishing an independent laboratory and within 10 years of obtaining their PhDs. **Research Grants – Program** are for independent scientists at all stages of their careers, and early career scientists are specifically encouraged to participate in these. The amount is dependent upon team size, up to $500,000 per year for a team of four.

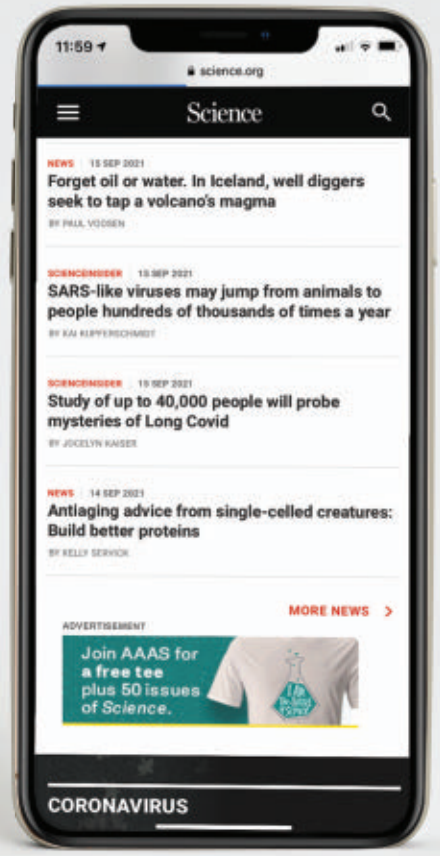
Applicants are advised to use the quiz on the HFSP website to check their eligibility and to read the guidelines carefully (www.hfsp.org).The application site will open at the end of January 2022 for principal applicants. The deadline to initiate an application is March 24, 2022, and the submission deadline for the Letter of Intent is March 31, 2022.

Specific enquiries: grant@hfsp.org

NEWS FROM Science

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

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



Science.org/news

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

RADIATIVE COOLING

A passive turnoff

Passive radiative cooling technology uses the infrared atmospheric window to allow outer space to be a cold sink for heat. However, this effect is one that is only helpful for energy savings in the warmer months. Wang *et al.* and Tang *et al.* used the metal-insulator transition in tungsten-doped vanadium dioxide to create window glass and a rooftop coating that circumvent this problem by turning off the radiative cooling at lower temperatures. Because the transition is simply temperature dependent, this effect also happens passively. Model simulations suggest that these materials would lead to energy savings year-round across most of the climate zones in the United States. —BG

Science, abg0291, abf7136, this issue p. 1501, p. 1504

New materials for passive radiative cooling of homes and buildings can be deactivated at cooler temperatures by a temperature-dependent transition.

NEUROSCIENCE

Locomotion-related signals in the brain

To calculate where we are in space, continuous knowledge of one's speed is necessary. How does the brain know how fast the body is traveling during locomotion? Using in vivo calcium imaging, electrophysiology, optogenetics, cell tracing, and histology, Farrell *et al*. identified neurons in the rodent supramammillary nucleus of the hypothalamus that encode future locomotor speed and potently drive locomotion when stimulated. Because these locomotor neurons have extensive axons in brain areas that support spatial navigation, this cell type distributes this information selectively to areas that require knowledge of speed. This nucleus is functionally positioned between input from a higher-order cognitive center and the downstream midbrain where locomotor nuclei reside. —PRS

Science, abh4272, this issue p. 1492

MAGNETISM

Peeking into magnetic textures

Topological spin textures hold promise as robust carriers of information and have been observed in bulk materials with a specific crystal structure. One of these materials, manganese germanide (MnGe), exhibits unusual textures in bulk form. Repicky *et al*. used spin-polarized scanning tunneling microscopy to study surface magnetism in thin films of MnGe. Achieving high spatial resolution, the researchers observed stripe-like features consistent with a helical state. In regions where the film was slightly curved due to strain, the intersection of domain walls led to characteristic closed patterns that could be manipulated with current/voltage pulses. —JS

Science, abd9225, this issue p. 1484

PHYSIOLOGICAL STRESS

Dangerous terrain

As climate change leads to regions of the world becoming increasingly uninhabitable, unregulated human migration is likely to increase. Migrants often traverse dangerous terrain, and the environmental conditions they encounter can be deadly. Campbell-Staton *et al*. used an approach commonly used to predict spatially explicit regions of physiological challenge in animal species to create a hazard landscape for the border crossing between the United States and Mexico. Their predictions of high risk, particularly due to dehydration, coincided with regions of high migrant mortality. Such detailed predictions may help to prevent these tragedies. —SNV

Science, abh1924, this issue p. 1496

CELL BIOLOGY

Elusive inhibin B co-receptor identified

Inhibin B secreted by the ovarian follicle in response to follicle-stimulating hormone (FSH) plays an essential role in

CREDITS: (THIS PAGE PHOTO) KALULU/ISTOCK.COM; (OPPOSITE PAGE LEFT TO RIGHT) GOOGLE QUANTUM AI; *WALLIS ET AL., DEV. CELL* **56**, P3192 (2021)/CC-BY

the pituitary-gonadal feedback loop, which suppresses pituitary FSH secretion and allows for the development of only a single follicle in humans. More than 20 years ago, it was demonstrated that inhibin A displays specific activity by binding to an activin receptor and an inhibin A–specific co-receptor betaglycan. Brule *et al*. identified the inhibin B co-receptor, which has evaded discovery for so long, as the transmembrane protein TGFBR3L (see the Focus by Woodruff). The authors demonstrate the critical role that this co-receptor plays in female fertility, suggesting that targeting this pathway may provide new methods to improve the regulation of fertility. —TPB

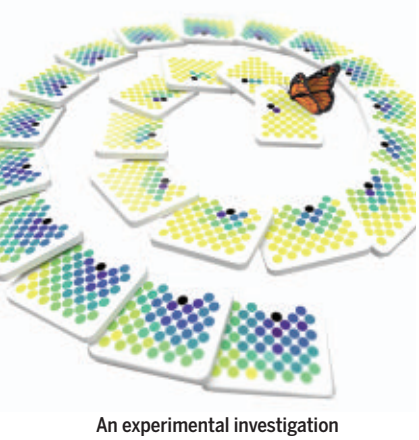
Sci. Adv. 10.1126/sciadv.abl4391, 10.1126/sciadv.abn1373 (2021).

QUANTUM SIMULATION

Quantum scrambling

Information spreading in interacting quantum systems is of relevance to a wide range of settings, from black holes to strange metals. Mi *et al*. used the Sycamore quantum processor to study this process. Through judicial design of quantum circuits, the researchers were able to separate the contributions of operator spreading and operator entanglement. Measuring the mean value and fluctuations of a specific correlator enabled quantifying these distinct contributions. —JS

Science, abg5029, this issue p. 1479



An experimental investigation of quantum scrambling on a superconducting chip yields observations of the so-called butterfly velocity.

QUANTUM SIMULATION

Establishing order, time after time

The formation of discrete time crystals, a novel phase of matter, has been proposed for some many-body quantum systems under periodic driving conditions. Randall *et al*. used an array of nuclear spins surrounding a nitrogen vacancy center in diamond as their many-body quantum system. Subjecting the system to a series of periodic driving pulses, they observed ordering of the spins occurring at twice the driving frequency, a signature that they claim establishes the formation of a discrete time crystal. Such dynamic control is expected to be useful for manipulating quantum systems and implementing quantum information protocols. —ISO

Science, abk0603, this issue p. 1474

MUCOSAL IMMUNOLOGY

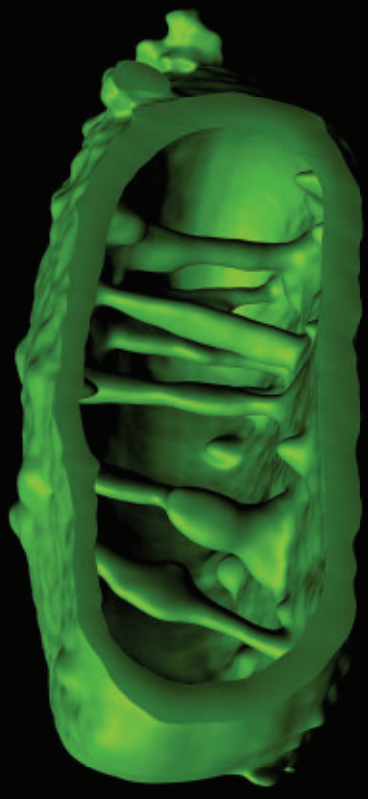
A fluent defense

Secretory immunoglobulin A (IgA) is known to be a key effector molecule in establishing effective antiviral immunity in the lungs, but the specific cell types producing this mucosal IgA and their physical locations are unclear. Oh *et al.* analyzed pulmonary IgA-secreting cells in mice after prior influenza infection or after intranasal immunization with an adjuvanted recombinant neuraminidase influenza vaccine. Both treatments enhanced antiviral immunity and IgA production when administered intranasally rather parenterally. The IgA responses were mediated through a combination of lung-resident memory B cells, plasmablasts, and plasma cells. These findings add to growing evidence that mucosal vaccination strategies show enhanced efficacy in establishing frontline mucosal immunity against respiratory pathogens. —IRW

Sci. Immunol. **6**, eabj5129 (2021).

IN OTHER JOURNALS

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



CELL BIOLOGY

Force regulation during repair

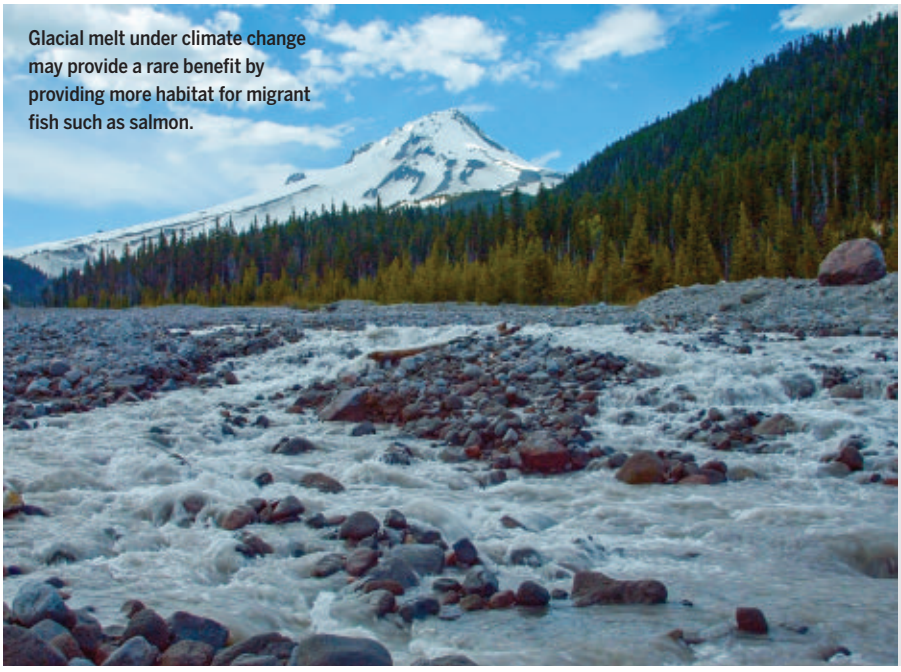
The cell nucleus is surrounded by a double membrane structure known as the nuclear envelope (NE). Ruptures in the NE compromise nuclear-cytoplasmic compartmentalization and contribute to genome instability and pro-inflammatory responses. Resealing of the wounded NE is mediated by the Endosomal Sorting Complex Required for Transport (ESCRT) machinery, a highly conserved membrane-remodeling pathway. During repair, the mechanical strain imposed by the cytoskeleton needs to be relieved to counteract intranuclear pressure. Wallis *et al.* found that the ESCRT-associated protein BROX regulates the mechanical properties of the NE during repair. BROX bound to the nucleoskeleton-cytoskeleton linker protein Nesprin-2G and promoted its removal from compression sites, which facilitated efficient membrane resealing and protected genetic material from damage. —SMH *Dev. Cell* **56**, P3192 (2021).

Resealing of a wounded nuclear envelope, shown in this 3D reconstruction, is essential for maintaining genome integrity.

MICROBIOLOGY

Prince among tiny vampires

Life clings on in even the most unpromising of Earth's habitats. A series of hypersaline alkaline lakes on the northeastern Mongolian plateau is home to several unique communities, including extremophile bacteria. Yakimov *et al.* discovered that the anaerobic, purple sulfur photosynthetic bacterium



Glacial melt under climate change may provide a rare benefit by providing more habitat for migrant fish such as salmon.

CLIMATE CHANGE

The bright side

Human-induced climate change is already wreaking havoc in ecosystems. Even if we ceased all fossil fuel activities today, much damage has already been done. However, not all may be lost. Pittman *et al.* modeled stream systems emerging from the melting of glaciers in the mountainous Pacific Northwest region of the United States and identified those with the potential for salmon habitat and colonization under a range of climate-warming scenarios. The authors found that glacier retreat could create more than 6000 kilometers of new salmon habitat, a third of which would be suitable for spawning. These emerging habitats present a rare opportunity for proactive conservation. —SNV *Nat. Commun.* **12**, 6816 (2021).

Halorhodospira, which is found in the most saline of these lakes, not only has to survive extraordinarily stressful conditions of the prevailing chemistry, temperature, and light, but is also hunted down by another microbe. Unusually, this ultrasmall organism has been successfully cultured and observed to attach to cell walls of the halobacterium. By sipping its host's cytoplasm (it is related to *Vampiricoccus* spp.), it gains most of the metabolic precursors it needs. Despite its genetic minimalism, this type of vampire bacterium may constitute up to half of the Earth's bacterial diversity. —CA

Environ. Microbiol. 10.1111/1462-2920.15823 (2021).

PLANT ECOLOGY

Cumulative impacts of tree pathogens

Invasive pests and pathogens are an increasing problem for tree species, especially in temperate forests heavily disturbed by human activities. Many tree species are hosts to unique communities of obligate associated species (particularly invertebrates), which are also at risk on the demise of their host. The risk to biodiversity extends further if two host species are lost at once: Mitchell *et al.* found that the loss of associated species of oak (*Quercus* spp.) and ash (*Fraxinus excelsior*) in the United Kingdom is likely to be greater than the sum of the obligate associates, potentially affecting hundreds of species. These findings call for particular attention to be paid to the management of forest landscapes where multiple dominant tree species are threatened. —AMS

J. Ecol. 10.1111/1365-2745.13798 (2021).

COSMIC DUST

Disappearing interstellar silicon carbide

Grains of stardust condense in the outflows from old asymptotic giant branch (AGB) stars. Infrared observations of carbon-rich AGB stars show strong emission from silicon carbide (SiC) grains, but such emissions have never been observed in the interstellar medium (ISM). Chen *et al.* used density functional theory and astrochemical rate calculations to determine whether SiC grains could be destroyed by oxygenation reactions in the diffuse ISM. They found that the reaction proceeds efficiently and without an energy barrier, but the rate is too low to overcome SiC production by AGB stars. Some unknown process must be consuming SiC grains in the ISM or hiding them from observation. —KTS

Mon. Not. R. Astron. Soc. 10.1093/mnras/stab3175 (2021).

HYDROGELS

Clearer and cleaner

Many materials have been adapted to use sunlight to degrade pollutants or remove accumulated dirt, for example, but these uses have largely been limited to photocatalytic reactions at a surface. For water purification, it would be far more effective and efficient to be able to use materials in bulk form. Kuckhoff *et al.* designed a hydrogel based on a high-transmittance acrylamide copolymerized with a photocatalytic unit containing benzothiadioazole at the 2% level. The material can degrade organic pollutants, including rhodamine B, methyl phenyl sulfide, and the herbicide glyphosate, as well as inorganic chromium VI compounds. Test volumes up to a half liter showcased the ability to degrade these compounds throughout the hydrogel material, which can be recovered and dehydrated for transport and repeated use. —MSL

Chem. Mater. 10.1021/acs.chemmater.1c02180 (2021).

STEM WORKFORCE

15,000 years of peer review

Scientific publishing relies heavily on the peer review system. Peer review often is done voluntarily, as part of scholarly service. Researchers provide comments to improve manuscripts and judge their quality, which requires highly specialized knowledge. Using publicly available data, Aczel *et al.* determined that, globally, researchers spent more than 100 million hours in 2020 on peer review, equivalent to more than 15,000 years. In the United States, the estimated monetary value of voluntary peer review was more than 1.5 billion dollars. Although these data emphasize the massive amount of voluntary time that researchers provide to scientific publishing, the authors stress that these numbers are likely underestimates and discuss the importance of considering alternative ways of structuring and paying for peer review. —MMc

Res. Integr. Peer Rev. **6**, 14 (2021).

PHOTO: TYLER HULETT/GETTY IMAGES

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

SPIN CHEMISTRY

Quantum oscillations in radical pairs

The spin dynamics of photoinduced radical pairs, involving an interconversion between singlet and triplet spin states, plays an important role in nature, for example, in avian magnetoreception. The spin interconversion is a truly quantum process with characteristic coherent oscillations (quantum beats) that should be reflected in the reaction kinetics. However, their experimental observation has remained challenging. Mims *et al*. developed an optical readout technique that can directly monitor the singlet-triplet interconversion quantum beats, as demonstrated for a photoinduced, charge-separated state of an electron donor–acceptor dyad (see the Perspective by Hore). The present work opens a new way to monitor the spin evolution in radical pairs, which will be important not only in biological physics but also in organic solar cells and other practical applications. —YS

Science, abl4254, this issue p. 1470; see also abm9261, p. 1447

PEROVSKITES

Nitrides join the perovskite club

Perovskite structured materials have a variety of uses as photovoltaics, capacitors, and micromechanical actuators, along with other applications. Oxides, halides, and chalcogenides all have large numbers of perovskite structured materials. Examples of perovskite nitrides are conspicuously absent, but Talley *et al*. managed to synthesize one (see the Perspective by Hong). Lanthanum tungsten nitride in the perovskite structure turns out to be piezoelectric, which is ideal for a variety of applications. Perovskite structured nitrides are very attractive because they could easily integrate with the large number of nitride-based semiconducting devices already in use. —BG

Science, abm3466, this issue p. 1488; see also abm7179, p. 1445

FOREST ECOLOGY

An experimental forest ecosystem drought

Drought is affecting many of the world's forested ecosystems, but it has proved challenging to develop an ecosystem-level mechanistic understanding of the ways that drought affects carbon and water fluxes through forest ecosystems. Werner *et al*. used an experimental approach by imposing an artificial drought on an entire enclosed ecosystem: the Biosphere 2 Tropical Rainforest in Arizona (see the Perspective by Eisenhauer and Weigelt). The authors show that ecosystem-scale plant responses to drought depend on distinct plant functional groups, differing in their water-use strategies and their position in the forest canopy. The balance of these plant functional groups drives changes in carbon and water fluxes, as well as the release of volatile organic compounds into the atmosphere. —AMS

Science, abj6789, this issue p. 1514; see also abn1406, p. 1442

BIOTECHNOLOGY

Reading amino acids by nanopore

Nanopore technology enables sensing of minute chemical changes at the single-molecule level by detecting differences in an ion current as molecules are drawn through a membrane-embedded pore. The sensitivity is sufficient to discriminate between nucleotide bases in nanopore sequencing, and other applications of this technology are promising. Brinkerhoff *et al*. developed a nanopore-based, single-molecule approach in which a protein was sequentially scanned in single-amino-acid steps through the narrow construction of a nanopore, and ion currents were monitored to resolve differences in the amino acid sequence along the peptide backbone (see the Perspective by Bošković and Keyser). The peptide reader was capable of reliably detecting single-amino-acid substitutions within individual peptides. An individual protein could be re-read many times, yielding very high read accuracy in variant identification. These proof-of-concept nanopore experiments constitute a promising basis for the development of a single-molecule protein sequencer. —DJ

Science, abl4381, this issue p. 1509; see also abn0001, p. 1443

CARBON CAPTURE

A hydrophobic CO_2 physisorbent

Most materials for carbon dioxide (CO_2) capture of fossil fuel combustion, such as amines, rely on strong chemisorption interactions that are highly selective but can incur a large energy penalty to release CO_2. Lin *et al*. show that a zinc-based metal organic framework material can physisorb CO_2 and incurs a lower regeneration penalty. Its binding site at the center of the pores precludes the formation of hydrogen-bonding networks between water molecules. This durable material can preferentially adsorb CO_2 at 40% relative humidity and maintains its performance under flue gas conditions of 150°C. —PDS

Science, abi7281, this issue p. 1464

LIVER DISEASE

Halting a hepatocyte lipotoxicity driver

Despite its prevalence and seriousness, nonalcoholic steatohepatitis (NASH) still lacks a treatment. Zhang *et al*. show that the lipoxygenase ALOX12 increases NASH severity in mice, pigs, and macaques independently of its enzymatic function by stabilizing acetyl-CoA carboxylase 1 (ACC1), altering lysosomal degradation, increasing hepatocyte inflammation, and impeding ketogenesis. In a separate study, Zhang *et al*. demonstrate that a small molecule effectively disrupts the ALOX12-ACC1 interaction in vivo, halting the development of liver steatosis, inflammation, and fibrosis in mice and macaque models of NASH without eliciting the hyperlipidemia that typically results from inhibiting the more canonical enzymatic function of ACC1. —CAC

Sci. Transl. Med. **13**, eabg8116, eabg8117 (2021).

METABOLISM

Insulin resistance from IP_3Rs

Dysregulation of calcium homeostasis in adipose tissue is associated with lipid accumulation and obesity. Guney *et al.* investigated the role of IP_3Rs, ligand-gated channels that release Ca^{2+} from intracellular stores, in adipose tissue inflammation during obesity. IP_3R levels and activity increased in the adipose tissue of mice fed a high-fat diet. Mice lacking $IP_3R1/2$ in adipocytes still gained weight when fed a high-fat diet but had reduced inflammatory cell infiltration of adipose tissue and did not develop some of the adverse metabolic effects of obesity, such as insulin resistance. —WW

Sci. Signal. **14**, eabf2059 (2021).

MEDICINE

Targeting the endocannabinoid system

The endocannabinoid system has diverse functions throughout the body, affecting neural development, neuron

excitation, cell division, metabolism, and inflammation. The endocannabinoid system is the target of phytocannabinoids, including but not limited to Δ9-tetrahydrocannabinol (THC) and cannabidiol (CBD), which are found in *Cannabis sativa*. In a Perspective, Keimpema *et al.* discuss the underlying mechanisms through which phytocannabinoids and synthesized agents might function and their potential therapeutic applications. Although there are encouraging results for endocannabinoid modulation in some types of epilepsy, there is much more that must be understood about the functional effects of phytocannabinoids and their targets before they can be used safely and effectively as therapeutics. —GKA

Science, abf6099, this issue p. 1449

CANCER IMMUNOLOGY

An atlas of cancer-associated T cells

The tumor microenvironment contains many different kinds of immune cells, the composition, function, and roles of which are unclear. Using single-cell RNA sequencing of T cells in 21 cancer types from more than 300 patients, Zheng *et al*. identified differences in transcript composition that could be used to catalog different T cell types (see the Perspective by van der Leun and Schumacher). These annotations identified the different roles of specific types of $CD4^+$ and $CD8^+$ T cells among the different tumor types. Some of these clusters revealed evidence for two developmental paths for T cells, one of which shows a trajectory toward the "exhausted" T cell state, and an understanding of this may be useful in developing future cancer immunotherapies. —LMZ

Science, abe6474, this issue p. 1462; see also abm9244, p. 1446

CORONAVIRUS

Vaccination and disease

The United Kingdom has high rates of vaccination for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), exceeding 80% of adults. As immunity wanes and social distancing is relaxed, how are rates of illness and severe disease affected by more infectious variants? Elliott *et al*. used reverse transcription PCR data from the REACT-1 study, which showed exponential transmission as the Alpha variant (B.1.1.7) was replaced by the Delta variant (B.1.617.2). After adjusting for age and other variables, vaccine effectiveness for the new variant averaged 55% in June and July of 2020. Despite the slower growth of the pandemic in the summer, it looks as if increased indoor mixing in the autumn will sustain transmission of the Delta variant despite high levels of adult vaccination. —CA

Science, abl9551, this issue p. 1463

GENOMICS

Giraffe pangenomes

Genomes within a species often have a core, conserved component, as well as a variable set of genetic material among individuals or populations that is referred to as a "pangenome." Inference of the relationships between pangenomes sequenced with short-read technology is often done computationally by mapping the sequences to a reference genome. The computational method affects genome assembly and comparisons, especially in cases of structural variants that are longer than an average sequenced region, for highly polymorphic loci, and for cross-species analyses. Siren *et al*. present a bioinformatic method called Giraffe, which improves mapping pangenomes in polymorphic regions of the genome containing single nucleotide polymorphisms and structural variants with standard computational resources, making large-scale genomic analyses more accessible. —LMZ

Science, abg8871, this issue p. 1461

RESEARCH ARTICLE SUMMARY

GENOMICS

Pangenomics enables genotyping of known structural variants in 5202 diverse genomes

Jouni Sirén†, Jean Monlong†, Xian Chang†, Adam M. Novak†, Jordan M. Eizenga†, Charles Markello, Jonas A. Sibbesen, Glenn Hickey, Pi-Chuan Chang, Andrew Carroll, Namrata Gupta, Stacey Gabriel, Thomas W. Blackwell, Aakrosh Ratan, Kent D. Taylor, Stephen S. Rich, Jerome I. Rotter, David Haussler, Erik Garrison, Benedict Paten*

INTRODUCTION: Modern genomics depends on inexpensive short-read sequencing. Sequenced reads up to a few hundred base pairs in length are computationally mapped to estimated source locations in a reference genome. These read mappings are used in myriad sequencing-based assays. For example, through a process called genotyping, mapped reads from a DNA sample can be used to infer the combination of alleles present at each site in the reference genome.

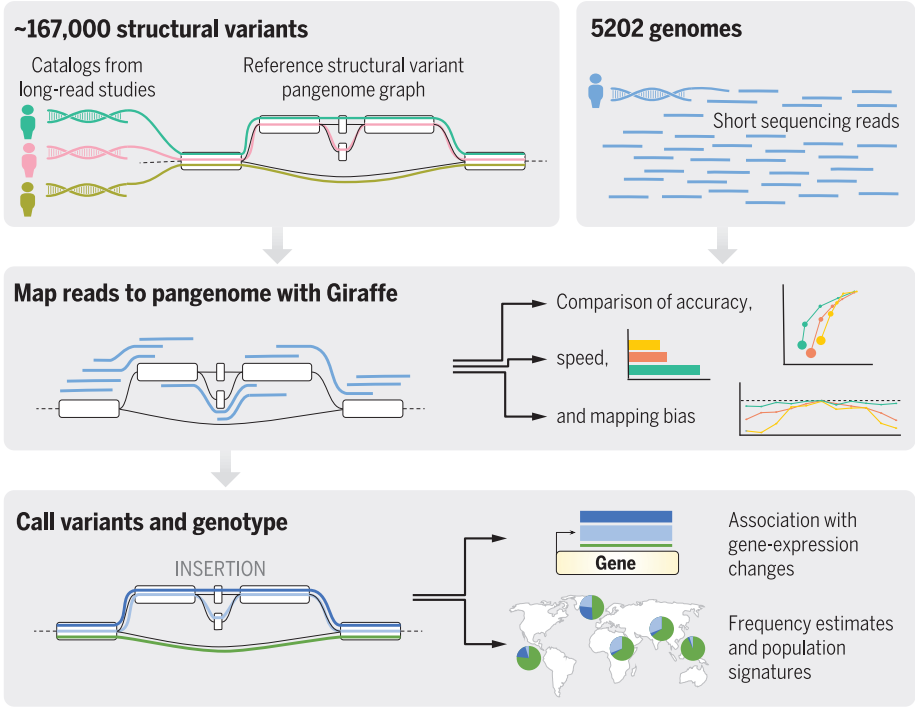
RATIONALE: A single reference genome cannot capture the diversity within even a single person (who gets a genome copy from each parent), let alone in the whole human population. Genomes differ not only by point variations, where one or a few bases are different, but also by structural variations, where differences can be much larger than an individual read. When a person's genome differs from the reference by a structural variation, the reference may contain no location to correctly map the corresponding reads. Although newer long-read sequencing allows structural variation to be more directly observed in sequencing reads, short-read sequencing is still less expensive and more widely available.

RESULTS: We present a short read–mapping tool, Giraffe. Giraffe maps to a pangenome reference that describes many genomes and the differences between them. Giraffe can accurately map reads to thousands of genomes embedded in a pangenome reference as quickly as existing tools map to a single reference genome. Simulations in which the true mapping for each read is known show that Giraffe is as accurate as the most accurate previously published tool. Giraffe achieves this speed and accuracy by using a variety of algorithmic techniques. In particular, and in contrast to previous tools, it focuses on mapping to the paths in the pangenome that are observed in individuals' genomes: the reference haplotypes. This has two key benefits. First, it prioritizes alignments that are consistent with known sequences, avoiding combinations of alleles that are biologically unlikely. Second, it reduces the size of the problem by limiting the sequence space to which the reads could be aligned. This deals effectively with complex graph regions where most paths represent rare or nonexistent sequences.

Using Giraffe in place of a single reference genome reduces mapping bias, which is the tendency to incorrectly map reads that differ from the reference genome. Combining Giraffe with state-of-the-art genotyping algorithms demonstrates that Giraffe mappings produce accurate genotyping results.

Using mappings from Giraffe, we genotyped 167,000 recently discovered structural variations in short-read samples for 5202 people at an average computational cost of $1.50 per sample. We present estimates for the frequency of different versions of these structural variations in the human population as a whole and within individual subpopulations. We identify thousands of these structural variations as expression quantitative trait loci (eQTLs), which are associated with gene-expression levels.

CONCLUSION: Giraffe demonstrates the practicality of a pangenomic approach to short-read mapping. This approach allows short-read data to genotype single-nucleotide variations, short insertions and deletions, and structural variations more accurately. For structural variations, this allowed the estimation of population frequencies across a diverse cohort of 5000 individuals. A single reference genome must choose one version of any variation to represent, leaving the other versions unrepresented. By making more broadly representative pangenome references practical, Giraffe attempts to make genomics more inclusive. ■



Overview of the experiments. Variant calls from long read–based and large-scale sequencing studies were used to construct pangenome reference graphs (top). Giraffe (and competing mappers) mapped reads to the graph or to linear references, and mapping accuracy, allele coverage balance, and speed were evaluated (middle). Then, mapped reads were used for variant calling, and variant call accuracy was evaluated (bottom). Structural variant calls were analyzed alongside expression data to identify eQTLs and population frequency estimates.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: bpaten@ucsc.edu
†These authors contributed equally to this work.
Cite this article as J. Sirén *et al.*, *Science* 374, eabg8871 (2021). DOI: 10.1126/science.abg8871

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg8871

RESEARCH ARTICLE

GENOMICS

Pangenomics enables genotyping of known structural variants in 5202 diverse genomes

Jouni Sirén[1]†, Jean Monlong[1]†, Xian Chang[1]†, Adam M. Novak[1]†, Jordan M. Eizenga[1]†, Charles Markello[1], Jonas A. Sibbesen[1], Glenn Hickey[1], Pi-Chuan Chang[2], Andrew Carroll[2], Namrata Gupta[3], Stacey Gabriel[4], Thomas W. Blackwell[5], Aakrosh Ratan[6], Kent D. Taylor[7], Stephen S. Rich[6], Jerome I. Rotter[7], David Haussler[1,8], Erik Garrison[9], Benedict Paten[1]*

We introduce Giraffe, a pangenome short-read mapper that can efficiently map to a collection of haplotypes threaded through a sequence graph. Giraffe maps sequencing reads to thousands of human genomes at a speed comparable to that of standard methods mapping to a single reference genome. The increased mapping accuracy enables downstream improvements in genome-wide genotyping pipelines for both small variants and larger structural variants. We used Giraffe to genotype 167,000 structural variants, discovered in long-read studies, in 5202 diverse human genomes that were sequenced using short reads. We conclude that pangenomics facilitates a more comprehensive characterization of variation and, as a result, has the potential to improve many genomic analyses.

The field of genomics almost exclusively uses a single reference genome assembly as an archetype of a human genome. Reliance on comparing with the sequences within the reference assembly has created a pervasive bias toward the alleles it contains. This reference allele bias occurs because nonreference alleles are naturally harder to identify when mapping DNA sequencing data to the reference sequences. Reference allele bias is particularly acute for structural variations (SVs), which are complex alleles involving 50 or more nucleotides of divergent sequence. SVs affect millions of bases within each human genome. Because of reference allele bias, SVs are much more poorly characterized than single-nucleotide variants (SNVs) and short insertions and deletions (collectively termed indels) (*1*, *2*). Similarly, characterizing genetic variation in highly polymorphic and repetitive sequences has proven challenging (*3*).

Recent releases of the reference human genome assembly attempted to address these issues by adding additional sequences. These alternate sequences represent diversity in localized regions of the genome (*4*). However, to date, these limited additions have not found widespread use. By contrast, pangenomes encode information about many complete genome assemblies and their homologies (the sequences that are shared between genomes by virtue of descending from a common ancestral sequence). Pangenomes are emerging as a replacement for linear reference assemblies to help mitigate these problems (*5*–*7*). They can particularly improve genotyping of structural variants (*8*).

Pangenomes are frequently formulated as sequence graphs (*9*)—mathematical graphs that represent the homology relationships between multiple sequences. Several algorithms have been developed for mapping sequences to sequence graphs. None has yet made mapping the short sequencing reads from widely used DNA sequencers, such as those made by Illumina, to a structurally complex pangenome a practical option for large-scale applications. The original VG-MAP algorithm (*10*) maps to complex sequence graphs that contain cycles produced by duplications and complex genomic rearrangements (*10*). However, VG-MAP is at least an order of magnitude slower than popular linear genome mappers that have comparable accuracy. Given that mapping is frequently a bottleneck in genome analysis, the cost of VG-MAP has proven prohibitive. Other pangenome mappers have different capabilities and limitations. Some are faster but are limited to acyclic graphs that contain variation at relatively low density (*11*), and some can map to arbitrary sequence graphs but are designed for long reads (*12*). Other tools are not open source and are thus unavailable for general testing and customization (*13*, *14*), and some additionally cannot run on commodity computing environments (*14*).

[1]UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA. [2]Google Inc., Mountain View, CA, USA. [3]Genomics Platform, Broad Institute, Cambridge, MA, USA. [4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [5]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [6]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [7]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor–UCLA Medical Center, Torrance, CA, USA. [8]Howard Hughes Medical Institute, University of California, Santa Cruz, CA, USA. [9]Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA.
*Corresponding author. Email: bpaten@ucsc.edu
†These authors contributed equally to this work.

Results

Giraffe: Fast, haplotype-aware pangenome mapping

When a sequence graph reference (*5*) (fig. S1) is substituted for the traditional linear reference (Fig. 1A), it can reduce reference allele bias by including more alleles (*10*). However, it also expands the size of the alignment search space from a few linear chromosome strings to a combinatorially large number of paths in the graph. This has made our previous graph mappers slower than linear mappers (*10*). Giraffe solves this problem by considering the paths that are observed in individuals' genomes: the reference haplotypes. We use the two haplotypes (one from each parent) that each individual has in their genome and trace them as paths through the sequence graph. The graph describes which positions in the haplotypes are equivalent, whereas the haplotypes describe the subset of the possible paths in the graph to consider. Giraffe uses a graph Burrows-Wheeler transform (GBWT) index (*15*) to store and query a graph's haplotypes efficiently.

Giraffe's strategy of aligning to haplotype paths has two key benefits. First, it prioritizes alignments that are consistent with known sequences, thereby avoiding combinations of alleles that are biologically unlikely. Second, it reduces the size of the problem by limiting the sequence space to which the reads could be aligned. This deals effectively with complex graph regions where most paths represent rare or nonexistent sequences.

We designed Giraffe to minimize the amount of gapped alignment that is performed. Computing gapped alignments, in which sequences are allowed to gain or lose bases relative to each other, is much more expensive than gapless alignment because it requires pairwise dynamic programming algorithms. Most Illumina sequencing errors are substitutions (*16*), and common true indels relative to the traditional linear reference should already be present in the haplotypes; therefore, almost all reads will have a gapless alignment to some stored haplotype. Hence, we try to align each read without gaps before resorting to dynamic programming.

Giraffe follows the common seed-and-extend approach used by most existing mappers [see algorithm in (*17*)]. In this framework, short seed matches between a sequencing read and a genomic reference are found with minimal work, and then only good seeds are extended into mappings of the entire read (*18*–*20*). A visual overview of Giraffe's operation is given in (Fig. 1, B to F). The Giraffe algorithm uses several heuristics for prioritizing alignments.

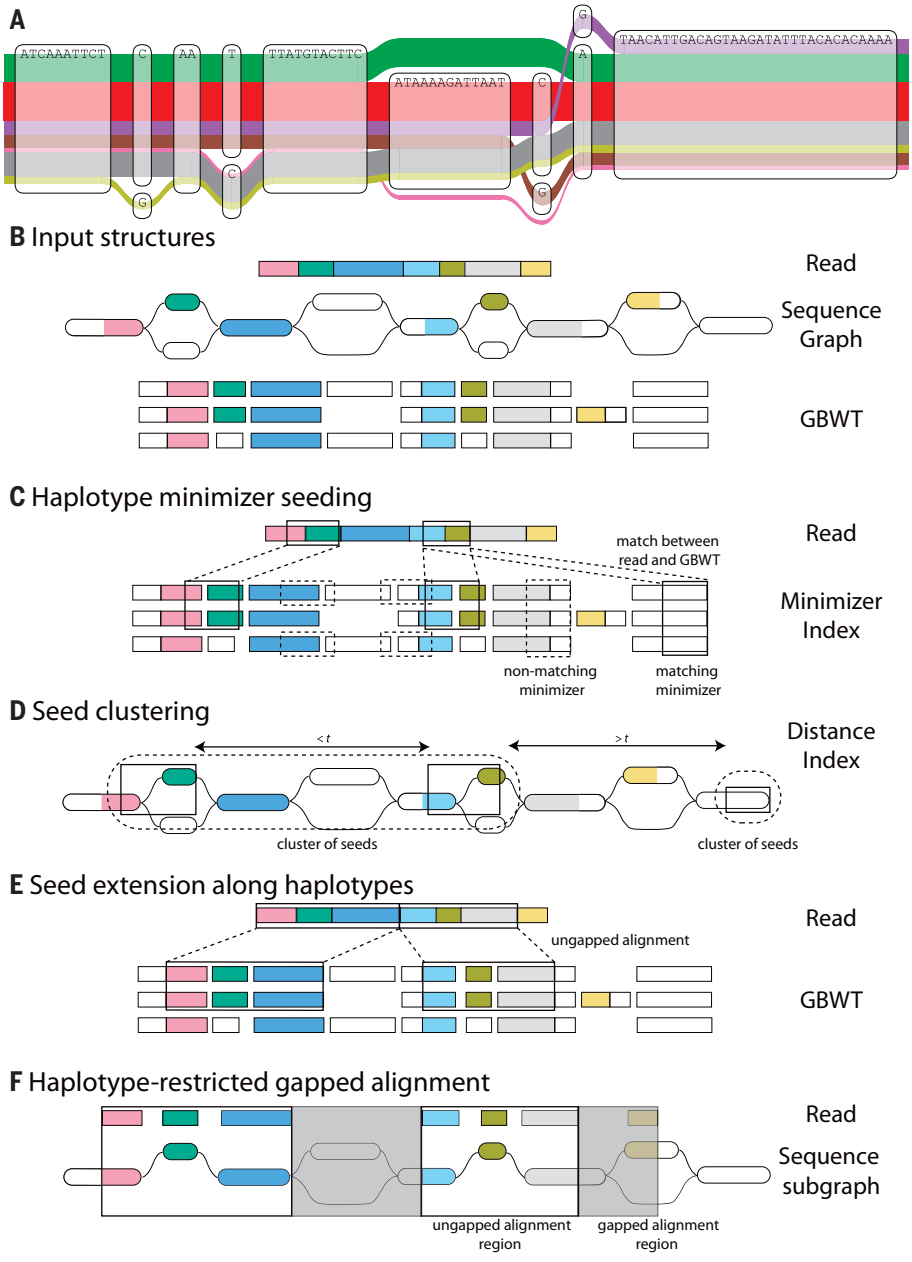


Fig. 1. Haplotype mapping. (**A**) A region of the *CASP12* gene in the 1000GP graph (*17*), illustrating complex local variation. The observed haplotypes (the colored ribbons of width log-proportional to population frequency) represent only a subset of the possible paths through the graph. (**B** to **F**) An overview of Giraffe. Input structures are shown in (B): Giraffe takes as input each read to map, the sequence graph reference to map against, and the GBWT of known haplotypes to restrict to. The input read is represented as a series of colored rectangles. The haplotype sequences in the GBWT are similarly represented as series of rectangles, split according to the nodes they correspond to in the sequence graph. Nodes in the sequence graph and haplotypes in the GBWT are colored according to homology with the read. Haplotype minimizer seeding is shown in (C): Seeds are identified using an index of minimizers (subsets of sequences of specified length k) (*50*) over the sequences of all the GBWT haplotypes. A matching minimizer between the read and the GBWT haplotypes constitutes a seed. The minimizers (black boxes) in the read are enumerated and the matching minimizers in the haplotypes are identified using the minimizer index. Seed clustering is shown in (D): Minimizer instances in the graph are clustered by the minimum graph distance (t, measured in nucleotides) between them (*51*). Seed extension along haplotypes is shown in (E): Minimizers in high-scoring clusters are extended linearly to form maximal gapless local alignments. Haplotype-restricted gapped alignment is shown in (F): Giraffe is designed on the assumption that for most reads, it will be possible to gaplessly extend seed alignments all the way to the ends of the read, allowing the algorithm to stop at the previous step. However, any remaining gaps in the alignment between read and graph are resolved by gapped alignment in this final step.

These heuristics are configurable, and we present two presets: default Giraffe (written as just "Giraffe") balances speed and accuracy, and fast Giraffe optimizes for speed at the expense of some accuracy.

Pangenome references for evaluation

To evaluate Giraffe, we built two human genome reference graphs based on the GRCh38 reference assembly. One (the 1000GP graph) contained mostly small [<50 base pairs (bp)] variants from the 1000 Genomes Project (*21*). The other (the HGSVC graph) contained entirely SVs (≥50 bp) from the Human Genome Structural Variant Consortium (*17*, *22*). The 1000GP graph contained data from 2503 individuals, with one (NA19239) held out for benchmarking. It was built from 76,749,431 SNVs; 3,177,111 small indels (<50 bp); and 181 larger SVs (≥50 bp). The HGSVC graph contained data from three individuals sequenced with long reads: HG00514, HG00733, and NA19240. The HGSVC graph contained 78,106 larger SVs (≥50 bp). Both graphs are available for reuse (see Data and materials availability in the Acknowledgments).

Giraffe and VG-MAP map accurately to human pangenomes

We evaluated Giraffe for mapping human data by simulating paired-end reads for two individuals (*17*): NA19240, who has available genotypes for the HGSVC variants (*22*), and NA19239, who has available genotypes for the 1000GP variants (*21*). Simulated read sets were mapped using Giraffe and competing tools (*17*). We examined the accuracy of single- and paired-end mapping (Fig. 2). We looked at a variety of input read sets and evaluated the calibration of reported mapping quality, which is a standard measure of mapping uncertainty (figs. S2 to S7 and tables S1 to S6). Relative to other tools, at the highest reported mapping quality, VG-MAP and default Giraffe consistently have either higher precision or higher recall across all simulated read technologies and graphs. Their performance is generally similar. Relative to the linear mappers, the Giraffe and VG-MAP lead is larger for the HGSVC graph (Fig. 2, C and D) than for the 1000GP graph (Fig. 2, A and B). This suggests that the gains from using a genome graph are higher when the graph facilitates alignment of genomic sequences from the sample that differ greatly from the linear reference.

Haplotype sampling improves read mapping

Having rare variants or errors in the graph and haplotypes may reduce mapping accuracy by creating opportunities for false-positive mappings (*23*). Mapping reads to regions with many distinct local haplotypes can also be slow. Additionally, Giraffe needs a mechanism to synthesize haplotypes for graph components

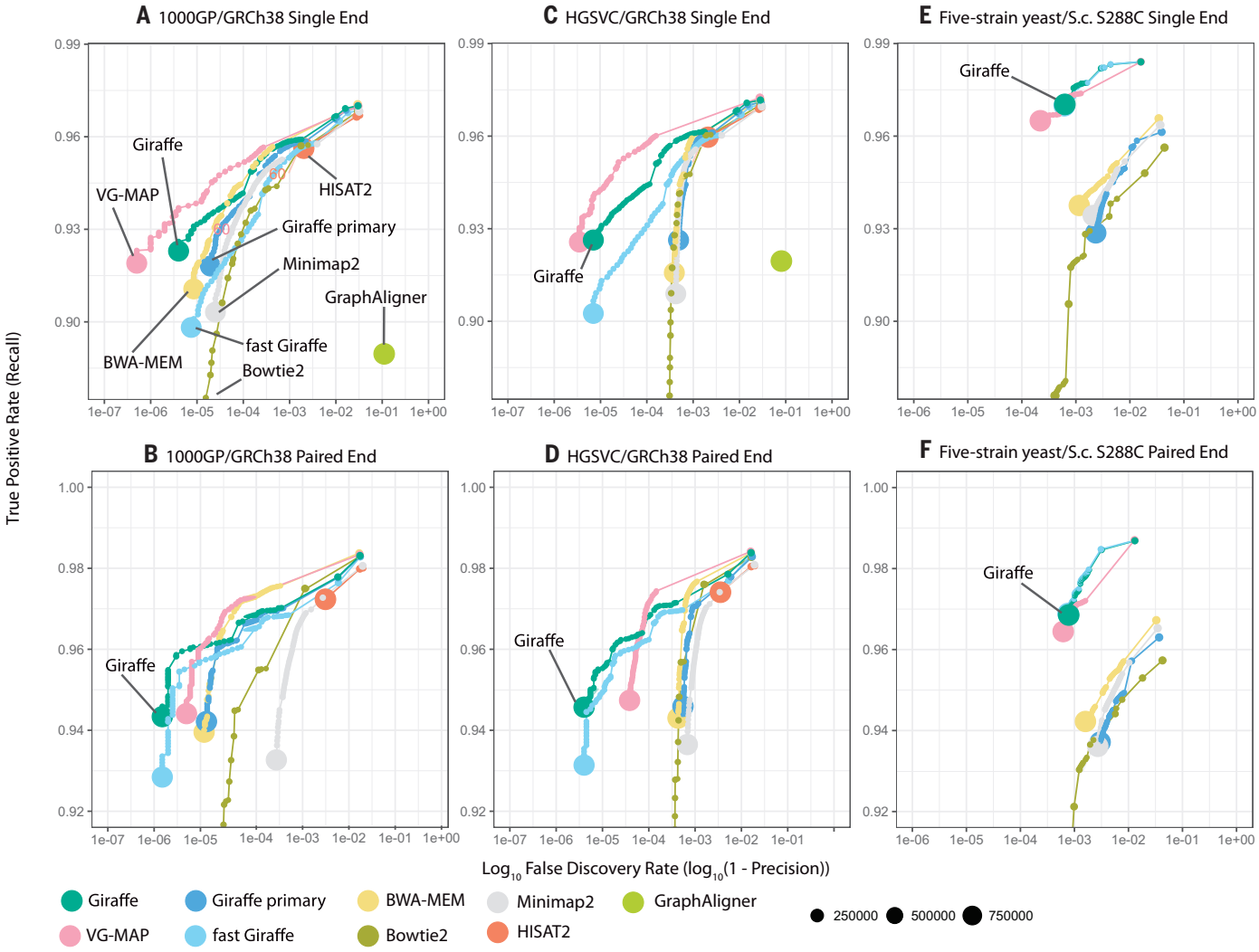


Fig. 2. Simulated read mapping. (**A** to **F**) Each panel shows recall versus false discovery rate (or 1 minus precision) for a simulated read-mapping experiment, comparing Giraffe with linear genome mappers (BWA-MEM, Bowtie2, and Minimap2) and other genome graph mappers (VG-MAP, GraphAligner, and HISAT2). Reads were simulated to match ~150-bp Illumina NovaSeq (for human) or HiSeq 2500 (for yeast) reads, either as single-end reads [(A) to (C)] or as paired-end reads [(D) to (F)] (*17*). Results for each mapper are shown stratified by reported read-mapping quality; the size of each point represents the log-scaled number of reads with the corresponding mapping quality. Three different mapping scenarios are assessed: [(A) and (D)] Comparing mapping to a graph derived from the 1000GP data to mapping to the linear reference genome assembly upon which it is based (GRCh38); [(B) and (E)] comparing mapping to a graph containing larger structural variants from the HGSVC project to mapping to the GRCh38 assembly upon which it is based; and [(C) and (F)] comparing mapping to a multiple sequence alignment–based yeast graph to mapping to the single *S.c. S288C* linear reference, for reads from the DBVPG6044 strain. For mapping with Giraffe, we used the full GBWT that contains six haplotypes to map to the HGSVC graph and the 64-haplotype sampled GBWT to map to the 1000GP graph. "Giraffe primary" represents mapping with Giraffe to the linear reference.

where no haplotype variation is known. To overcome these issues, Giraffe includes mechanisms for creating synthetic haplotype paths. When real haplotypes are available, these synthetic haplotype paths represent local haplotype variation sampled according to haplotype frequency, and we call the result a sampled GBWT (*17*). When no haplotypes are available, we call the result a path cover GBWT. In this case, the synthetic haplotypes represent random walks through the graph. We evaluated the effects of running our mapping evaluations with sampled and path cover GBWTs [fig. S8 and tables S7 and S8; (*17*)]. The mapping benefit of sampling more haplotypes plateaued at 64 haplotypes for the 1000GP graph (which contains around 5000 haplotypes), with higher accuracy than that achieved by mapping to the full haplotype set. We used the HGSVC graph (which contains just six haplotypes) for an experiment on generating path covers without known haplotypes. Path covers alone did not outperform the full underlying haplotype set for the HGSVC graph but came close to matching its performance. We selected the 64-haplotype sampled GBWT for the 1000GP graph and the full GBWT for the HGSVC graph as the best-performing GBWTs, which we use in the rest of the analysis.

Giraffe improves pangenome mapping speed

We measured the runtime (Fig. 3, A and B) and memory usage (Fig. 3, C and D) of Giraffe and competing tools when mapping real reads (*17*). Giraffe was more than an order of magnitude faster than VG-MAP in all conditions.

Fig. 3. Runtime and memory usage. (**A** to **D**) Total runtime [(A) and (B)] and peak memory use [(C) and (D)] for mapping ~600 million NovaSeq 6000 reads using 16 threads. Reads were mapped [(A) and (C)] to the 1000GP derived graph or (for linear mappers) the GRCH38 assembly and [(B) and (D)] to the HGSVC graph or GRCh38 reference, respectively. For HISAT2*, results are shown for the subset 1000GP graph. "Giraffe full" refers to mapping using the full GBWT of all haplotypes. "Giraffe sampled" refers to mapping using the 64-haplotype sampled GBWT.

It was also faster at aligning to human graphs than Bowtie2 or BWA-MEM were at aligning to the corresponding linear reference. For the 1000GP graph, using the 64-haplotype sampled GBWT for mapping instead of the full ~5000-haplotype GBWT was much faster in every case. HISAT2 and fast Giraffe were both about equally fast and were both faster than all other mappers.

Because of the in-memory indexes it uses, Giraffe's memory consumption is higher than the other mappers, except for GraphAligner. However, it can map to the 1000GP graph with the full GBWT in ~80 gigabytes (GB) of memory—an amount readily available on compute cluster nodes (Fig. 3, C and D).

Giraffe reduces allele mapping bias

We assessed Giraffe's reference bias (*17*). We expected Giraffe to be able to use the extra variation information contained in the graph reference to achieve a lower level of bias than a linear mapper. For variants that were heterozygous in NA19239, we found the fraction of reads supporting alternate versus reference alleles at each indel length (Fig. 4A). Giraffe and VG-MAP both show less bias toward the reference allele than a linear mapper, and this difference becomes more pronounced as indel length increases, particularly for larger insertions.

Giraffe genotyping outperforms best practices

We used Illumina's Dragen platform (*14*) to genotype SNVs and short indels using Giraffe mappings to the 1000GP graph, projected onto the linear reference assembly. We compared these results with results using competing graph and linear reference mappers (*17*). No training or optimization was performed for any of the mappings other than those performed by default by Dragen itself. We evaluated the calls using

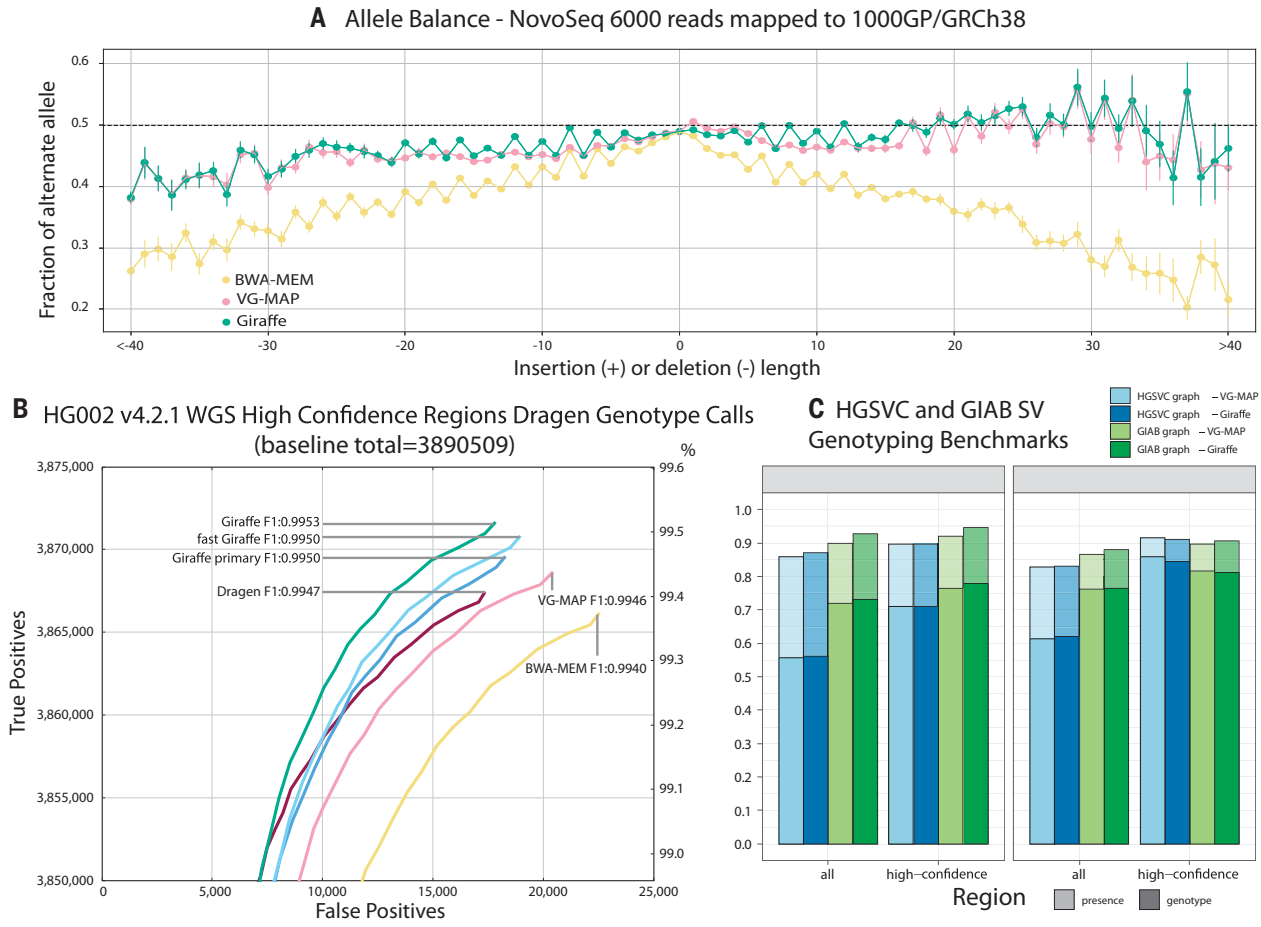


Fig. 4. Evaluating Giraffe for genotyping. (**A**) The fraction of alternate alleles in reads detected for heterozygous variants in NA19239. Reads were mapped to the 1000GP graph with Giraffe and VG-MAP and to GRCh38 with BWA-MEM, and the fraction of reads supporting reference or alternate alleles was found for each indel length. (**B**) Assessing true-positive and false-positive genotypes made using the Dragen genotyper with mappings from Giraffe and other mappers. The line labeled Dragen represents the mapper included with the Dragen system itself. (**C**) Comparing Giraffe with VG-MAP for typing large insertions and deletions. "Presence" (lighter bars) evaluates the detection of SVs without regard to genotype; "genotype" (darker bars) requires the SV to be detected and its genotype to agree with the truth genotype. The *y* axis shows the F1 score. For the HGSVC benchmark, we define high-confidence regions as regions not overlapping simple repeats and segmental duplications. For the GIAB benchmark, we use the set high-confidence regions provided by GIAB.

the Genome in a Bottle (GIAB) v4.2.1 HG002 high-confidence variant-calling benchmark (*24*).

Out of the examined pipelines, Giraffe mappings to the 1000GP graph produce the highest overall F1 score (harmonic mean of precision and recall) at 0.9953 (Fig. 4B and tables S9 and S10). Similar but uniformly higher results were found with higher-coverage, 250-bp reads (fig. S9 and tables S11 and S12). Although one would expect longer reads and higher coverage to produce better variant calls, with all else being equal, Giraffe has a slightly higher F1 score with the 150-bp read set (0.9953) than BWA-MEM with the higher coverage 250-bp read set (0.9952). Restricting comparison only to confident regions that overlap variant calls from the 1000GP variants used in graph construction, Giraffe has the highest F1 score at 0.9995 relative to the other methods (fig. S10 and table S13). Perhaps surprisingly, Giraffe maintains the highest F1 score (0.9528) when performing the converse analysis, restricting the comparison to confident regions that do not overlap 1000GP variant calls (fig. S11 and table S14).

DeepVariant is a highly accurate genotyping tool that requires training (*25*). We trained DeepVariant to use Giraffe mappings and evaluated it on the held-out sample HG003 (*17*). We compared it with the Dragen pipelines tested and DeepVariant using BWA-MEM with the BWA-MEM trained model that the developers provide. The Giraffe-DeepVariant pipeline (F1: 0.9965) outperforms all other tested pipelines (fig. S12 and tables S15 and S16).

Previously, when we used VG-MAP to map reads to SV pangenomes, we found it to perform better than other methods for SV genotyping (*8*). We replicated that evaluation on the HGSVC and GIAB datasets (*1*, *22*) to confirm that the quality of the SV genotypes from Giraffe was competitive (*17*). We observed similar SV genotyping accuracy across SV types, genomic regions, and datasets (Fig. 4C). Of note, GraphTyper (*26*), which was published after our earlier benchmarking analysis (*8*), was also compared with vg as a variant caller but showed lower genotyping performance across SV types, genomic regions, and datasets (fig. S13).

Giraffe generalizes beyond human

We assessed Giraffe's performance mapping to a yeast pangenome for five strains of the *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* yeasts (*17*). This graph was substantially different from the human graphs. It proved challenging because it contains the cycles and duplications typical of graphs generated from genome-wide alignments of more divergent sequences. Using a graph decomposition technique (*27*), we find it contains 1,459,769 variant sites, four times the density of variation in the 1000GP graph. Ninety of these sites are

complex, meaning that they are not directed, not acyclic, or not free of internal source and sink nodes.

Mapping accuracy results for reads from the held-out *S. cerevisiae* DBVPG6044 strain are displayed in Fig. 2, E and F, for the single-end and paired-end reads, respectively. Speed results for mapping real reads are presented in fig. S14. Neither HISAT2 nor GraphAligner could map reads to the yeast graph; in the case of HISAT2, this was because it cannot map to graphs containing cycles. Giraffe is 28 times faster for paired-end mapping than the only other tool that could map to this graph, VG-MAP, while achieving similar accuracy. Both graph mappers are much more accurate than the linear reference methods. Moreover, the gap between the graph methods and the linear reference is even larger on this graph than in the HGSVC graph (Fig. 2, C and D). This lends further support to the hypothesis that graph-mapping methods have the most benefit when facilitating alignment of genomic sequences that differ greatly from the reference, such as the sibling-subspecies-scale differences represented in the yeast graph.

Genotyping the SVs of the 5202 samples

Building on our previous work to genotype SVs (*8*), we demonstrate the value of Giraffe by performing population-scale genotyping of an expanded compendium of SVs in large cohorts of samples sequenced with short reads. We built a comprehensive pangenome containing SVs that combines variants from three catalogs of SVs that were discovered using long-read sequencing (*1*, *22*, *28*). The combined catalog represents 16 samples from diverse human populations and is estimated to cover most of the common insertions and deletions in the human population (*28*). Near-duplicate versions of variants (i.e., SVs with slightly different breakpoints) are often present within and across SV catalogs. A naïve integration of all these variants can lead to redundancy in the graph that can affect read mapping and variant genotyping. We remapped sequencing data and integrated variants iteratively into the graph to progressively build a nonredundant, compact SV graph (*17*). The final SV graph was constructed from 123,785 SVs from the original catalogs: 53,663 deletions and 70,122 insertions. Overall, the graph contained 26.2 Mbp of nonreference sequences in the form of insertions. Using a graph decomposition (*27*), we identified 228,405 subgraphs that represent variant sites. Some of these correspond to smaller variants nested inside larger ones. After combining these cases, there were 96,644 non-nested, nonoverlapping SV subgraphs.

Compared with Hickey *et al.* (*8*), we used a graph containing more SVs and a more recent version of the vg toolkit (see Data and materials availability in the Acknowledgments), including the Giraffe mapper presented here. Our SV genotypes were as accurate as those in (*8*), if not more so (fig. S15, A to C). Thanks to Giraffe and improvements in the variant calling approach, the genotyping workflow used about 12 times less compute on a sample sequenced at about 20× coverage.

SV genotyping was run using the NHLBI BioData Catalyst ecosystem (*29*) (fig. S15D). We genotyped samples from the Multi-Ethnic Study of Atherosclerosis (MESA) cohort within the Trans-Omics for Precision Medicine (TOPMed) program. The MESA cohort is a longitudinal cohort study consisting of 6814 participants at baseline (between the years 2000 and 2002). Participants were ascertained from six sites in the United States and identified themselves as "Spanish/Hispanic/Latino" (22% at baseline) and/or "African American or Black" (28%), "Chinese" (12%), and "Caucasian or White" (39%) (*30*). Two-thousand samples from the MESA cohort (*30*) were selected, using a criterion to maximize the sample diversity (*17*). Using the graph described above and fast Giraffe, it took around 4 days to genotype 2000 samples from the MESA cohort. We used the same workflow to genotype the 3202 diverse samples from the high-coverage 1000GP dataset (*31*) in around 6 days. On average, genotyping a sample took 194.4 central processing unit (CPU)–hours of compute and cost between $1.11 and $1.56 (fig. S15D and tables S17 and S18). The sequencing data were down-sampled in advance to ~20× coverage to reduce compute costs. Benchmarking indicates that this down-sampling has a minimal impact on the genotyping accuracy (fig. S16).

Diverse, clustered SVs

Our SV graph construction approach preserves multiple alleles cataloged at a given SV site and decomposes them into a parsimonious joint representation. In general, these SV representations require more alleles per site than is common for small variants. For example, there may be SNVs and indels within the sequence of an insertion or around a deletion's breakpoints, or copy-number changes in variable number tandem repeat (VNTR) regions. The existence of these potentially recombining subvariants implies the possibility of previously unidentified alleles.

We genotyped a total of about 1.7 million alleles clustered in 167,858 SV sites across the 2000 MESA samples. In most SV sites, we only observed one or a few alleles (~90% of SV sites with five or fewer alleles; Fig. 5A). Additionally, most of the SV sites (~151,000, 90%) contained SV alleles that differed by only small variants (*17*), whereas the rest of the sites showed size variation from polymorphic VNTR regions (fig. S17A). Examples of SV sites that illustrate these different profiles are given in Fig. 5, B and C, and fig. S18. Figure 5, D and E, shows the size and frequency distributions of the most common allele at each site. SVs spanned the size spectrum (50 bp up to 125 kbp), with 89.8% shorter than 500 bp. For 84% of the SVs, simple repeats or low-complexity regions overlapped at least 50% of the SV region. Hence, the SVs genotyped in this study match the original SVs discovered in long-read sequencing studies (*1*, *22*, *28*) in terms of number, size distribution, and sequence context.

We observed similar patterns in the 1000 Genomes Project dataset. We identified 1.9 million alleles clustered in 167,188 SV sites, with a size and frequency distribution similar to that of the MESA cohort (fig. S19). The 1000 Genomes Project dataset also provided 602 trios that we used to estimate the quality of our genotypes. First, we computed the rate of Mendelian error, which was 5.2% for deletions and 4.7% for insertions when considering all variants. This error decreased as the confidence in the genotype increased. For example, the Mendelian error dropped to 2.1 and 2.5% for the ~70% of deletions and insertions, respectively, with the highest genotype qualities (fig. S20A). The most common error by far occurred when a heterozygous variant was predicted in the offspring but both parents were predicted homozygous for the reference allele (table S19). The transmission rate of heterozygous alleles was close to the expected 50%: 40 to 47% for deletions and 43 to 49% for insertions (fig. S20B).

Comprehensive SV frequency estimates

The genotyped SVs were originally discovered with long-read sequencing technology, and many are absent from the population scale SV catalogs that could provide frequency information. Of the SV sites genotyped using our pangenome approach, 93% are missing from the 1000 Genomes Project SV catalog (*32*), and 67% were missing from the Genome Aggregation Database (gnomAD)–SV catalog (*33*) (table S20). This is consistent with the amount of previously unidentified structural variation described in the three studies from which our SV graph is derived (*1*, *22*, *28*). Our results provide frequency estimates across a large and diverse cohort for these SVs.

The frequency distribution resembled the allele frequency distributions in the 1000 Genomes Project SV and gnomAD-SV catalogs (fig. S21A). The frequencies of the subset of variants present in both our catalog and the mentioned public catalogs were largely concordant (fig. S21, B and C). Our frequency distribution looks rather different than that of SVPOP (*28*). However, we note that SVPOP's frequency distribution is markedly different than the 1000 Genomes Project and gnomAD-SV (fig. S21A) and has very different frequency estimates on matched variants (fig. S21, D to F).

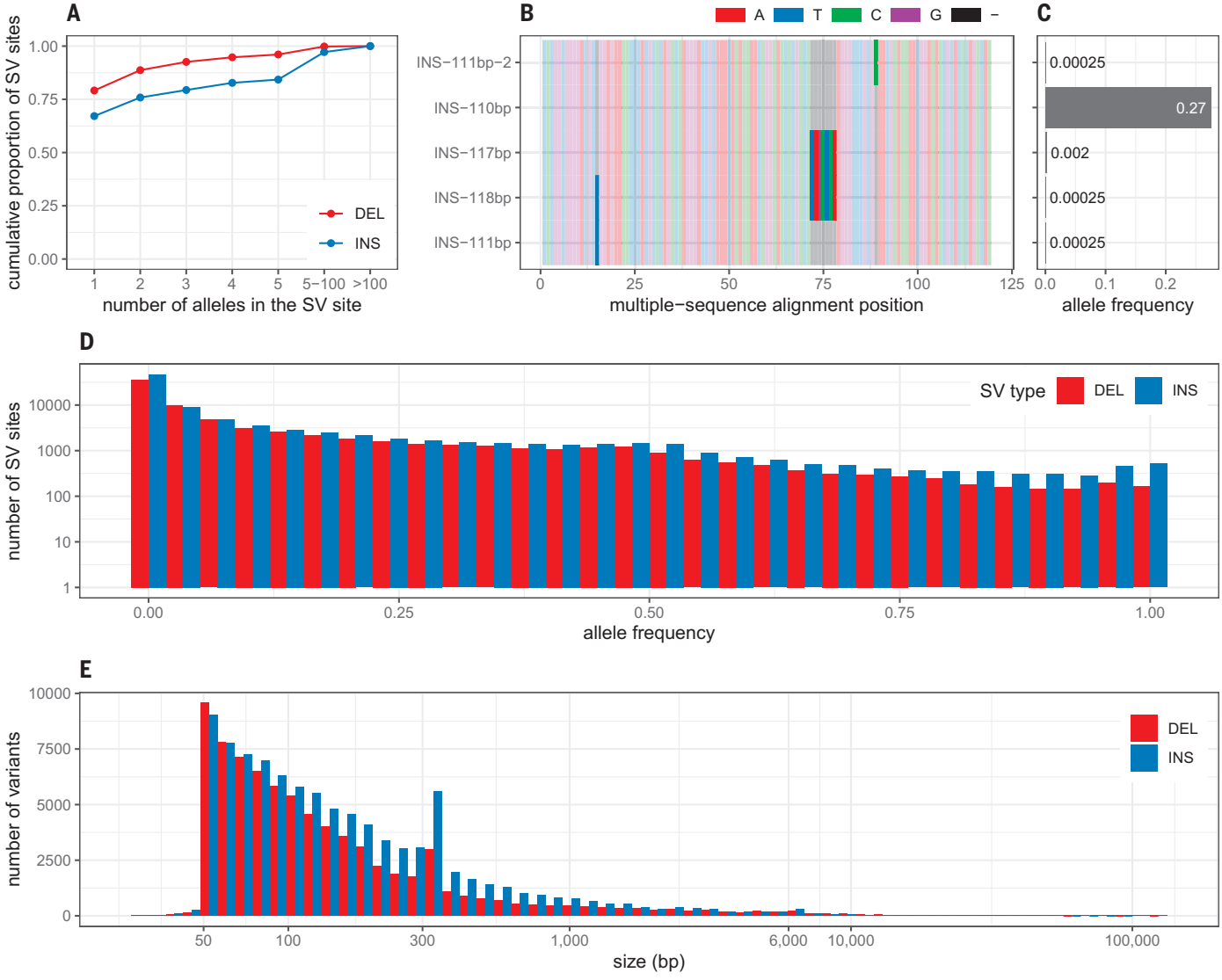


Fig. 5. SVs in the MESA cohort. (**A**) Cumulative proportion of SV sites depending on the maximum number of alleles (*x* axis) in the site. DEL, deletion; INS, insertion. (**B** and **C**) Illustration of an insertion site with five alleles. The alleles differ by three nested indels as shown by the multiple sequence alignment of the inserted sequences represented in (B). Only one allele is frequent in the population (allele frequency of 0.27), as highlighted in (C). (**D**) Allele frequency distribution of the major allele for each SV site. The *y* axis, showing the number of SVs, is log-scaled. (**E**) Size distribution of the major allele for each SV site.

Fine-tuning SVs with frequencies

SVs in the input catalogs may contain errors. When multiple alleles co-occur at an SV site, we often observed that one allele was frequently present in the cohort, whereas other similar alleles were not (Fig. 5, B and C). The other alleles at these sites are either rare or erroneous. In either case, it is useful to identify the major alleles. In 7520 SV sites, only one allele was called in more than 1% of the population, whereas other alleles from the original catalogs were not. Further, the major allele was at least three times more frequent than the second most frequent allele in 6175 of these sites (fig. S17B). As a quality control, we verified that these alleles were more likely to match exactly with the alleles in the GIAB truth set (*1*), which is the SV catalog with the highest base-level confidence [permutation $p < 0.0001$; (*17*)]. Our results thus help fine-tune the sequence resolution of these SVs. More generally, our results identify one major allele for 39,699 multiallelic SV sites.

SV frequency population signatures

Principal components analysis (PCA) of the allele counts at the 166,959 SV sites in the MESA cohort produces a low-dimensional embedding of the samples. This embedding appears similar to the TOPMed consortium's PCA of SNV genotype data from all samples (Pearson correlation of 0.96 to 0.99 for the top three components; fig. S22). This result is expected and provides confirmatory support for the accuracy of our SV genotypes.

We clustered samples with PCA, taking each cluster to be a population (*17*). Allele frequencies vary across these populations for thousands of SV sites (fig. S23, A to C). For example, we found 21,069 SV sites with strong intercluster frequency patterns, defined by a frequency in any population differing by more than 10% from the median frequency across all populations (fig. S23D). The existence of SVs with different frequencies across populations supports the need to develop and test genomic tools and references across multiple populations.

Because there is a risk of circularity when using the same genotype data to define populations and look for patterns across them, we

replicated these observations in the high-coverage 1000 Genomes Project dataset (*31*). Here, again, the PCA of the allele counts organized the samples in a way consistent with the known history of the 1000 Genomes Project "superpopulation" groups (fig. S24). In this analysis, we found 25,960 SV sites with strong inter-superpopulation frequency patterns, defined as for the MESA analysis, but with the 1000 Genomes superpopulations as the sample categories (fig. S25). As a comparison, when the samples were randomly grouped into superpopulations, we observed only 14 SV sites with strong intergroup frequency patterns (*17*). More than 17,000 SV sites with strong inter-superpopulation frequency patterns were enriched or depleted in the African Ancestry (AFR) superpopulation, followed by about 10,000 sites enriched or depleted in the East Asian Ancestry (EAS) superpopulation.

As an example of a newly annotated variant, a deletion of the *RAMACL* gene was genotyped with frequency 46.6% in the AFR super population, 4% in American Ancestry (AMR), and less than 1% in other superpopulations. This deletion is not present in the 1000 Genomes Project SV catalog and was unresolved in version two of the gnomAD-SV catalog. It has been curated in gnomAD-SV v2.1 and shows similar population patterns there to what we found in our reanalysis of the 1000 Genomes Project dataset. Such variants could be falsely identified as putatively pathogenic if analyzed only in European-ancestry populations where the frequency is low.

In addition, our approach is often capable of genotyping repeat-rich variants, such as short tandem repeats that vary in length. For example, a 1-kbp expansion of an exonic VNTR in *MUC6* with a frequency of 14% in the AFR superpopulation was observed only rarely outside of it: 2.3% in AMR and <1% in other superpopulations (Fig. 6A). This repeat expansion is absent from gnomAD-SV and the SV catalog from the 1000 Genomes Project, despite its observed frequency.

SVs, genes, and expression

In the MESA and 1000 Genomes Project datasets, 1563 and 1603 SVs overlapped coding regions of 408 and 380 protein-coding genes, respectively. When including promoters, introns, and untranslated regions, each dataset had overlaps between at least 78,290 SVs and 7641 protein-coding genes. Of these SVs, 10,640 show strong inter-superpopulation frequency patterns in the 1000 Genomes Project dataset (see Fig. 6A).

We searched for associations between SVs and gene expression across 445 samples from the 1000 Genomes Project that have been RNA sequenced by the Genetic European Variation in Disease (GEUVADIS) consortium (*34*). These samples span four European-ancestry populations [Utah residents (CEPH) with Northern and Western European ancestry (CEU), Finnish in Finland (FIN), British in England and Scotland (GBR), and Toscani in Italy (TSI)], and the Yoruba in Ibadan, Nigeria (YRI) population (*34*). A pooled analysis identified 2761 expression quantitative trait loci (eQTLs) across 1270 genes [false discovery rate of 1%; (*17*)]. Of those genes, 878 are protein-coding genes. We note that 58% of the SV-eQTLs are located within simple repeats or low-complexity regions. The distribution of the p values across all tests showed the expected patterns for genome-wide association studies (fig. S26).

Genes with eQTLs, or eGenes, were enriched in gene families involved in immunity, as previously observed (*35*), but we also found significant enrichments in other families (table S21). For example, 3 of the 10 genes in the anoctamins family have SV-eQTLs (adjusted $p = 0.0006$). This gene family is involved in the regulation of multiple processes, including neuronal cell excitability, and mutations in some of its members have been linked to neurologic disorders (*36*). Other families enriched included the survival motor neuron (SMN) complex family (3 out of 10 genes with an SV-eQTL, adjusted $p = 0.0012$) and aldehyde dehydrogenases genes (3 out of 19 genes with

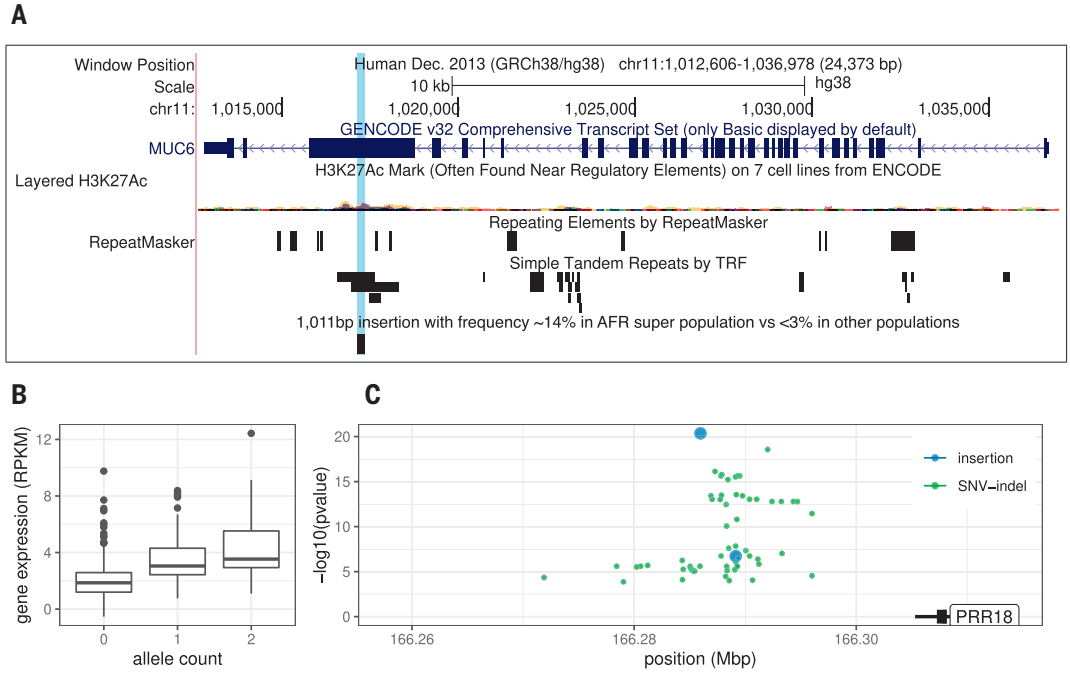


Fig. 6. Population-specific SVs and SV-eQTLs in the 1000 Genomes Project dataset. (**A**) Example of an insertion at appreciable frequency (~14%) in the AFR superpopulation that is rare (<3%) in the other superpopulations. The variant is a 1011-bp expansion of a VNTR in the coding sequence of the *MUC6* gene. chr11, chromosome 11; TRF, Tandem Repeats Finder. (**B** and **C**) Association between a 10,083-bp insertion overlapping a predicted enhancer and the gene expression of the *PRR18* gene. Each allele is associated with an increase in gene expression, as shown in (B). The position of significant eQTLs (SNV-indels in green, insertions in blue) is shown in (C). All the eQTLs are in the intergenic region downstream of the *PRR18* gene. The *y* axis represents the significance of the association, with the top eQTL being the highest point. Of note, the lead eQTL (the 10,083-bp insertion) overlaps a region predicted to be an enhancer by ENCODE. In (B), boxes represent the median and quartiles; whiskers extend from the box up to 1.5 times the interquartile range.

an SV-eQTL, adjusted $p = 0.008$). As expected, SV-eQTLs were strongly enriched in coding, intronic, promoter, untranslated, and regulatory regions (fig. S27). Interestingly, SVs associated with decreased gene expression drove most of the enrichment in coding regions. Separate analysis of the four European-ancestry populations together and the YRI population alone identified, respectively, 44 and 139 SVs where an association with the expression of protein-coding genes was detected only in the smaller analysis (*17*). As expected, a number of these population-specific SV-eQTLs had shown strong inter-superpopulation frequency patterns (see above).

Finally, we performed a joint analysis with available SNV and indel calls. Like previous studies (*37*, *38*), we found that the lead eQTLs (the strongest association for a gene) are enriched in SVs (permutation $p = 0.022$). For example, only 0.5% of the variants tested were SVs, but SVs were the lead eQTL in 5.9% of the genes that had both SV and SNV-indel eQTLs. We did not observe a difference in relative effect size of SV-eQTLs compared with SNV-indel eQTLs, but we noticed that SV-eQTLs were fourfold enriched in the genes with the highest expression (permutation $p = 0.004$; fig. S28). Figure 6, B and C, and fig. S29 show two examples where the SV-eQTL is the strongest association: a 10,083-bp insertion associated with an increased expression of the *PRR18* gene and a 5405-bp deletion associated with a reduced expression of the *SLC44A5* gene. In addition, 39 genes had SV-eQTLs but no SNV-indel eQTLs (table S21). These results show that the SV genotypes produced here can be used to test for phenotypic association.

Discussion

Pangenome references hold great potential as a replacement for standard linear reference genomes. They can represent diverse collections of human genomes, and they have been shown to reduce the bias that arises from using a linear reference (*10*). However, because of the appreciable complexity of the task, previous methods for mapping to pangenomes have been slow or not clearly better than comparable methods for linear genomes. By contrast, Giraffe can map to pangenome graphs consisting of thousands of aligned haplotypes, potentially with complex topologies, with accuracy comparable to that of the best previously published tools and speed surpassing linear reference mappers. Further, we have demonstrated that its mappings can improve genotyping.

Pangenome exchange formats have been coevolving alongside pangenome methods. Giraffe is designed to meet and solidify these emerging standards while also interfacing with the broader genomics ecosystem. The graphical fragment assembly (GFA) format for representing pangenome graphs has increasing tool support, including by vg (*12*, *39*–*43*). In addition, Giraffe can output the graphical mapping format (GAF) read-to–pangenome graph alignment format proposed by Li *et al.* (*39*) and supported by other pangenome mappers (*12*). Giraffe also supports backward compatibility to linear references by allowing mappings to be projected onto an embedded linear reference genome and output in standard formats. The state-of-the-art SNV and short-indel genotyping results described in this study demonstrate the value of this support. These necessary technical advances are starting to nucleate an interoperable tool ecosystem for pangenomics.

For SVs, and for particularly large insertions, we and others have shown that the benefits of pangenomes for genotyping are not merely incremental but transformative (*8*, *39*, *44*). Our approach allowed us to identify duplicate SVs, to refine the canonical definitions of SVs, and to establish the frequencies of these SVs in diverse human populations. Complementing previous surveys of SVs in diverse human populations (*32*, *38*, *45*), we demonstrate that many of the previously unidentified SVs studied here are also differentially distributed across human populations. This frequency information could be used, among other applications, for prioritizing variants to investigate for genomic medicine because variants common anywhere are unlikely to be pathogenic.

We expect accurate and unbiased SV genotyping to be one of the most impactful contributions of pangenomics. Among other applications, this contribution will enable more links from SVs to disease traits and other phenotypes to be identified. For example, we were able to detect thousands of associations between SVs and gene expression. Ebert *et al.* (*38*) recently performed a similar analysis using the same RNA sequencing (RNA-seq) dataset from the GEUVADIS consortium (complemented with 34 new deep RNA-seq experiments) and genotypes for SVs discovered in 32 haplotype-resolved genomes. Although we did not use this new sequencing data and SV catalog, we found a similar number of SV-eQTLs with our pangenomic approach [2761 SV-eQTLs and 1270 eGenes in this study; 2109 SV-eQTLs and 1526 eGenes in Ebert *et al.* (*38*)].

Soon, pangenomes will be built from larger collections of high-quality de novo assembled genomes using accurate long reads. We hope such human pangenomes will enable more comprehensive genotyping of common complex variants (including SVs) from existing catalogs of short-read sequencing data, allowing for the typing of such variants at the scale of existing catalogs of point variation. We expect that unlocking this latent information will ultimately aid with disease association studies and help us further understand how the architecture of the genome contributes to an individual's phenotype.

Methods summary

Evaluation

Read simulation

To evaluate Giraffe for mapping human data, we obtained paired-end sequencing reads from a parent-child pedigree. Reads were obtained from an Illumina NovaSeq 6000 machine for parent NA19239 (accession ERR3239454) and from Illumina HiSeq 2500 and HiSeq X Ten machines for child NA19240 (accession nos. ERR309934 and SRR6691663, respectively). These samples were selected because NA19240 has genotypes for the HGSVC variants (*22*), whereas NA19239 has genotypes for the 1000GP variants (*21*). NA19239 was excluded from the 1000GP graph (*17*). We simulated 1 million read pairs (2 million reads) from each individual's haplotypes (*17*).

Read-mapping accuracy

Simulated read sets were mapped to the graphs using Giraffe, VG-MAP (*10*), HISAT2 (*11*), and GraphAligner (*12*). We were unable to build a HISAT2 index for the full 1000GP graph, and so instead we mapped it to a graph created from a subset of the 1000GP data where all variants with a frequency below 0.001 were filtered out. In addition, we mapped the read sets to the primary graphs using Giraffe and to the linear reference assemblies using the linear sequence mappers BWA-MEM (*19*), Bowtie2 (*18*), and Minimap2 (*20*). Mapping accuracy was evaluated by comparing the positions along embedded, shared linear paths at which reads fell after mapping with similarly determined positions for their original simulated alignments.

Read-mapping speed

We compared mapping runtime, speed, and memory usage on an AWS EC2 i3.8xlarge node with 32 vCPUs and 244 GB of memory. To estimate real-world runtime and memory usage, we aligned a shuffled read set of 600 million NovaSeq 6000 reads from NA19239. We mapped reads to the 1000GP graph, the HGSVC graph, and the GRCh38 linear reference for comparison and measured runtime and memory usage. For each tool, we also separately measured reads mapped per thread per second, ignoring the start-up time of the mapper (fig. S30). This measure gives an estimate of speed that is invariant to read-set size or thread count, except for the effects of long-running work batches and thread synchronization overhead (*17*).

Read-mapping bias

To assess reference allele mapping bias, we mapped 600 million real paired-end NovaSeq 6000 reads for NA19239 to the 1000GP graph

using default Giraffe and VG-MAP. For comparison, we mapped the same reads to GRCh38 with BWA-MEM.

Genotyping accuracy

We compared the performance of Giraffe, VG-MAP, Illumina's Dragen platform, and BWA-MEM for genotyping SNVs and short indels. The design of each calling pipeline is described in section S4 of the supplementary materials (*17*) and the parameters and indexes for each experiment are described in table S22. The variants produced by each pipeline were compared against the GIAB v4.2.1 HG002 high-confidence variant-calling benchmark (*24*) using the RealTimeGenomics vcfeval tool (*46*) and Illumina's hap.py tool (*47*). This benchmark set covers 92.2% of the GRCh38 sequence.

We also evaluated a DeepVariant (*25*) pipeline that uses Giraffe mappings (*17*). Using the default DeepVariant 1.1.0–trained model, we tested genotyping of the HG003 sample across the entire genome. This sample was not used in training the model.

Generalization to yeast

To evaluate Giraffe's performance on more diverged, nonhuman data, we used a yeast graph built from a Cactus multiple sequence alignment for five strains of the *S. cerevisiae* and *S. paradoxus* yeasts (*8*). For the corresponding negative-control primary graph, we used the *S.c. S288C* assembly. We collected basic statistics about the yeast graph and decomposed the graph for analysis using the method of (*27*). We simulated 500,000 read pairs from a held-out *S. cerevisiae* yeast strain, DBVPG6044, not included in the yeast graph, using an error and length model for Illumina HiSeq 2500 reads (*17*).

SV genotyping

We built an SV pangenome from the HGSVC (*22*), GIAB (*1*), and SVPOP (*28*) sequence-resolved catalogs. After filtering out erroneous duplicates using a remapping approach, the SVs were iteratively inserted in the genome graph to minimize the effect of errors and redundancy in the catalog. The SVs were then genotyped across 5202 genomes by aligning short-read sequencing data using Giraffe with a workflow description language (WDL) workflow that we deposited in Dockstore (*48*). Two-thousand samples were selected from the MESA cohort to maximize sample diversity. The remaining 3202 samples are from the 1000 Genomes Project and include 2504 unrelated individuals. The trios available in this latter dataset were used to compute the rate of Mendelian concordance in the genotypes.

The different SV alleles observed in the population were clustered into SV sites based on their reciprocal overlap (for deletions) and sequence similarity (for insertions). We used the frequency profile across alleles within an SV site to identify the major allele and to fine-tune variants with near duplicates in the combined catalog that may have been due to errors. Each variant was then annotated with its presence in existing SV databases (*28*, *32*, *33*), its repeat content, and its location relative to gene annotations. We also compared the frequency distributions across the SV databases and how well the frequency estimates matched for variants shared across databases.

PCA was performed on the SV genotypes, and principal components were compared with those produced from SNV-indel genotypes. We defined strong intercluster or inter-superpopulation frequency patterns by a frequency in any cluster or superpopulation differing by more than 10% from the median frequency across all of them. For the 2000 MESA samples, the clusters were defined using hierarchical clustering on the first three principal components. For the 1000 Genomes Project, we used their "superpopulation" assignments. Permutations were used to contrast the number of SVs with such patterns with an expected baseline.

Finally, we examined the SV genotypes in a subset of the samples that had gene-expression data available from the GEUVADIS consortium (*34*). MatrixEQTL (*49*) identified SV-eQTLs while controlling for sex and population structures, as summarized by the first four principal components. Separate analyses of the four European-ancestry populations together and the YRI population alone were performed similarly. In addition, we performed a joint eQTL analysis with publicly available SNVs and indels (*31*). We used permutation to compute enrichment of SV-eQTLs in gene regions, gene families, or among lead-eQTLs (those with the strongest association for a gene).

REFERENCES AND NOTES

1. J. M. Zook *et al*., A robust benchmark for detection of germline large deletions and insertions. *Nat. Biotechnol.* **38**, 1347–1355 (2020). doi: 10.1038/s41587-020-0538-8; pmid: 32541955
2. M. Mahmoud *et al*., Structural variant calling: The long and the short of it. *Genome Biol.* **20**, 246 (2019). doi: 10.1186/s13059-019-1828-7; pmid: 31747936
3. J. Ebler, A. Schönhuth, T. Marschall, Genotyping inversions and tandem duplications. *Bioinformatics* **33**, 4015–4023 (2017). doi: 10.1093/bioinformatics/btx020; pmid: 28169394
4. D. M. Church *et al*., Modernizing reference genome assemblies. *PLOS Biol.* **9**, e1001091 (2011). doi: 10.1371/journal.pbio.1001091; pmid: 21750661
5. The Computational Pan-Genomics Consortium, Computational pan-genomics: status, promises and challenges. *Brief. Bioinform.* **19**, 118–135 (2016). doi: 10.1093/bib/bbw089
6. R. M. Sherman, S. L. Salzberg, Pan-genomics in the human genome era. *Nat. Rev. Genet.* **21**, 243–254 (2020). doi: 10.1038/s41576-020-0210-7; pmid: 32034321
7. S. Ballouz, A. Dobin, J. A. Gillis, Is it time to change the reference genome? *Genome Biol.* **20**, 159 (2019). doi: 10.1186/s13059-019-1774-4; pmid: 31399121
8. G. Hickey *et al*., Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biol.* **21**, 35 (2020). doi: 10.1186/s13059-020-1941-7; pmid: 32051000
9. J. M. Eizenga *et al*., Pangenome graphs. *Annu. Rev. Genomics Hum. Genet.* **21**, 139–162 (2020). doi: 10.1146/annurev-genom-120219-080406; pmid: 32453966
10. E. Garrison *et al*., Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879 (2018). doi: 10.1038/nbt.4227; pmid: 30125266
11. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019). doi: 10.1038/s41587-019-0201-4; pmid: 31375807
12. M. Rautiainen, T. Marschall, GraphAligner: Rapid and versatile sequence-to-graph alignment. *Genome Biol.* **21**, 253 (2020). doi: 10.1186/s13059-020-02157-2; pmid: 32972461
13. G. Rakocevic *et al*., Fast and accurate genomic analyses using genome graphs. *Nat. Genet.* **51**, 354–362 (2019). doi: 10.1038/s41588-018-0316-4; pmid: 30643257
14. Illumina, Accuracy improvements in germline small variant calling with the DRAGEN platform; https://science-docs.illumina.com/documents/Informatics/dragen-v3-accuracy-appnote-html-970-2019-006/Content/ Source/Informatics/Dragen/dragen-v3-accuracy-appnote-970-2019-006/ dragen-v3-accuracy-appnote-970-2019-006.html.
15. J. Sirén, E. Garrison, A. M. Novak, B. Paten, R. Durbin, Haplotype-aware graph indexes. *Bioinformatics* **36**, 400–407 (2020). pmid: 31406990
16. M. Schirmer, R. D'Amore, U. Z. Ijaz, N. Hall, C. Quince, Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* **17**, 125 (2016). doi: 10.1186/s12859-016-0976-y; pmid: 26968756
17. Materials and methods are available as supplementary materials.
18. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
19. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN] (2013).
20. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi: 10.1093/bioinformatics/bty191; pmid: 29750242
21. A. Auton *et al*., A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
22. M. J. P. Chaisson *et al*., Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* **10**, 1784 (2019). doi: 10.1038/s41467-018-08148-z; pmid: 30992455
23. J. Pritt, N.-C. Chen, B. Langmead, FORGe: Prioritizing variants for graph genomes. *Genome Biol.* **19**, 220 (2018). doi: 10.1186/s13059-018-1595-x; pmid: 30558649
24. J. Wagner *et al*., Benchmarking challenging small variants with linked and long reads. *bioRxiv* 2020.07.24.212712 [Preprint] (2020); doi: 10.1101/2020.07.24.212712
25. R. Poplin *et al*., A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018). doi: 10.1038/nbt.4235; pmid: 30247488
26. H. P. Eggertsson *et al*., GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nat. Commun.* **10**, 5402 (2019). doi: 10.1038/s41467-019-13341-9; pmid: 31776332
27. B. Paten *et al*., Superbubbles, Ultrabubbles, and Cacti. *J. Comput. Biol.* **25**, 649–663 (2018). doi: 10.1089/cmb.2017.0251; pmid: 29461862
28. P. A. Audano *et al*., Characterizing the major structural variant alleles of the human genome. *Cell* **176**, 663–675.e19 (2019). doi: 10.1016/j.cell.2018.12.019; pmid: 30661756
29. National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services, The NHLBI BioData catalyst. *Zenodo* (2020); https://doi.org/10.5281/zenodo.3822858.
30. D. E. Bild *et al*., Multi-ethnic study of atherosclerosis: Objectives and design. *Am. J. Epidemiol.* **156**, 871–881 (2002). doi: 10.1093/aje/kwf113; pmid: 12397006
31. M. Byrska-Bishop *et al*., High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *bioRxiv* 2021.02.06.430068 [Preprint] (2021); https://doi.org/10.1101/2021.02.06.430068.
32. P. H. Sudmant *et al*., An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015). doi: 10.1038/nature15394; pmid: 26432246
33. R. L. Collins *et al*., A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020). doi: 10.1038/s41586-020-2287-8; pmid: 32461652
34. T. Lappalainen *et al*., Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013). doi: 10.1038/nature12531; pmid: 24037378

35. M. Fagny *et al*., Exploring regulation in tissues with eQTL networks. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7841–E7850 (2017). doi: 10.1073/pnas.1707375114; pmid: 28851834
36. E. E. Benarroch, Anoctamins (TMEM16 proteins): Functions and involvement in neurologic disease. *Neurology* **89**, 722–729 (2017). doi: 10.1212/WNL.0000000000004246; pmid: 28724583
37. C. Chiang *et al*., The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692–699 (2017). doi: 10.1038/ng.3834; pmid: 28369037
38. P. Ebert *et al*., Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021). doi: 10.1126/science.abf7117; pmid: 33632895
39. H. Li, X. Feng, C. Chu, The design and construction of reference pangenome graphs with minigraph. *Genome Biol.* **21**, 265 (2020). doi: 10.1186/s13059-020-02168-z; pmid: 33066802
40. S. Koren *et al*., Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017). doi: 10.1101/gr.215087.116; pmid: 28298431
41. H. Li, Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016). doi: 10.1093/bioinformatics/btw152; pmid: 27153593
42. R. R. Wick, M. B. Schultz, J. Zobel, K. E. Holt, Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015). doi: 10.1093/bioinformatics/btv383; pmid: 26099265
43. A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics* **70**, e102 (2020). doi: 10.1002/cpbi.102
44. S. Chen *et al*., Paragraph: A graph-based structural variant genotyper for short-read sequence data. *Genome Biol.* **20**, 291 (2019). doi: 10.1186/s13059-019-1909-7; pmid: 31856913
45. P. H. Sudmant *et al*., Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, aab3761 (2015). doi: 10.1126/science.aab3761; pmid: 26249230
46. J. G. Cleary *et al*., Comparing variant call files for performance benchmarking of next-generation sequencing variant calling pipelines. *bioRxiv* 023754 [Preprint] (2015); doi: 10.1101/023754
47. P. Krusche *et al*., Illumina/hap.py. *GitHub* (2020); https://github.com/Illumina/hap.py.
48. J. Monlong, github.com/vgteam/vg_wdl/vg_mapgaffe_call_sv_cram. *Zenodo* (2020). .doi: 10.5281/zenodo.4290651
49. A. A. Shabalin, Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012). doi: 10.1093/bioinformatics/bts163; pmid: 22492648
50. M. Roberts, W. Hayes, B. R. Hunt, S. M. Mount, J. A. Yorke, Reducing storage requirements for biological sequence comparison. *Bioinformatics* **20**, 3363–3369 (2004). doi: 10.1093/bioinformatics/bth408; pmid: 15256412
51. X. Chang, J. Eizenga, A. M. Novak, J. Sirén, B. Paten, Distance indexing and seed clustering in sequence graphs. *Bioinformatics* **36**, i146–i153 (2020). doi: 10.1093/bioinformatics/btaa446; pmid: 32657356
52. J. Sirén *et al*., Software and products for "Pangenomics enables genotyping known structural variants in 5,202 diverse genomes". *Zenodo* (2021); doi: 10.5281/zenodo.4774364
53. C. A. Sloan *et al*., ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016). doi: 10.1093/nar/gkv1160; pmid: 26527727

ACKNOWLEDGMENTS

We acknowledge the studies and participants who provided biological samples and data for the TOPMed project. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute (NHLBI); the National Institutes of Health (NIH); or the US Department of Health and Human Services. **Funding:** Research reported in this publication was supported by the NIH under award numbers U41HG010972, R01HG010485, U01HG010961, OT3HL142481, OT2OD026682, U01HL137183, and 2U41HG007234. Research reported in this publication was supported by the NHLBI BioData Catalyst Fellows Program of the NIH through the University of North Carolina at Chapel Hill, under award number OT3HL147154. J.A.S. was supported by the Carlsberg Foundation. Computational resources for the project were made available by the NIH and by Amazon Web Services, without full compensation at market value. The high-coverage sequencing data for the 1000 Genomes Project were generated at the New York Genome Center with funds provided by National Human Genome Research Institute (NHGRI) grant 3UM1HG008901-03S1 and can be found on Terra. MESA and the MESA SHARe projects are conducted and supported by the NHLBI in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Funding for SHARe genotyping was provided by NHLBI contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, CA, USA) and the Broad Institute of Harvard and MIT (Boston, MA, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. This work was also supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Whole-genome sequencing (WGS) for the TOPMed program was supported by the NHLBI. WGS for "NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)" (phs001416) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1 and HHSN268201500014C). Core support, including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support, including phenotype harmonization, data management, sample-identity quality control, and general program coordination, was provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). **Author contributions:** Project design: D.H., E.G., B.P. Giraffe implementation: J.S., X.C., A.M.N., J.M.E., B.P. SV analysis: J.M., G.H. Short-variant analysis: C.M., P.-C.C., A.C. The vg implementation: J.S., J.M., X.C., A.M.N., J.M.E., C.M., J.A.S., G.H., D.H., E.G., B.P. Manuscript writing: J.S., J.M., X.C., A.M.N., J.M.E., C.M., J.A.S., G.H., B.P. Data production: N.G., S.G., T.W.B., A.R., K.D.T., S.S.R., J.I.R. **Competing interests:** P.-C.C. and A.C. are employees of Google and own Alphabet stock as part of the standard compensation package. The remaining authors declare no competing interests. **Data and materials availability:** An overview of the data generated for this paper, and key input data to reproduce the analyses, is available at https://cglgenomics.ucsc.edu/giraffe-data/. The dataset is available through InterPlanetary File System (IPFS) at https://ipfs.io/ipfs/QmVo4Q5hCKqUGJJZyYLGJTaiHZdK9JWhJtGJbKa9ojrSjh. Archived copies of the code and final reusable work products have been deposited at Zenodo (*52*). This archive also includes vg, toil-vg, and toil source code and Docker containers used in this work, as well as the giraffe-sv-paper orchestration scripts. "Final" versions of vg and toil-vg, including all features needed to reproduce this work, are 9907ab2 for vg and 99101f2 for toil-vg. The latest version of the vg toolkit, including the Giraffe mapper, is customarily distributed at https://github.com/vgteam/vg. The scripts used for the analysis presented in this study were developed at https://github.com/vgteam/giraffe-sv-paper, a git bundle of which is archived at Zenodo (*52*). Data used in the Giraffe read-mapping experiments—including the 1000GP, HGSVC, and yeast target graphs, the linear control graphs, the graphs used to simulate reads, and the simulated reads themselves—can be found at https://cgl.gi.ucsc.edu/data/giraffe/mapping/. The SV pangenomes and SV catalogs annotated with allele frequencies are hosted at https://cgl.gi.ucsc.edu/data/giraffe/calling/ and archived at Zenodo (*52*). This repository also includes SVs with strong inter-superpopulation frequency patterns, SV-eQTLs, and SVs that overlap protein-coding genes. To build the 1000GP and HGSVC graphs, we used the GRCh38 no-alt analysis set (accession no. GCA_000001405.15) and the hs38d1 decoy sequences (accession no. GCA_000786075.2), both available from the National Center for Biotechnology Information (NCBI), in addition to the variant call files distributed by the respective projects. To train read simulation and evaluate speed, we used human read sets ERR3239454, ERR309934, and SRR6691663 and yeast read sets SRR4074256, SRR4074257, SRR4074394, SRR4074384, SRR4074413, SRR4074358, and SRR4074383, all available from Sequence Read Archive (SRA). The public high-coverage sequencing dataset from the 1000 Genomes Project (*31*) is available at www.internationalgenome.org/data-portal/data-collection/30x-grch38, including European Nucleotide Archive (ENA) projects PRJEB31736 and PRJEB36890. The gene-expression data were download from ArrayExpress E-GEUV-1 (GD462.GeneQuantRPKM.50FN.samplename.resk10.txt.gz). We downloaded the call sets from the ENCODE portal (*53*) (www.encodeproject.org/) with the identifier ENCFF590IMH. Individual WGS data for TOPMed whole genomes are available through dbGaP. The dbGaP accession no. for MESA is phs001416. Data in dbGaP can be downloaded by controlled access with an approved application submitted through their website: www.ncbi.nlm.nih.gov/gap.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg8871
Materials and Methods
Figs. S1 to S31
Tables S1 to S22
References (*54*–*77*)
MDAR Reproducibility Checklist

2 February 2021; accepted 2 November 2021
10.1126/science.abg8871

Science
JOURNALS AAAS

Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLE SUMMARY

CANCER IMMUNOLOGY

Pan-cancer single cell landscape of tumor-infiltrating T cells

Liangtao Zheng†, Shishang Qin†, Wen Si†, Anqiang Wang, Baocai Xing, Ranran Gao, Xianwen Ren, Li Wang, Xiaojiang Wu, Ji Zhang, Nan Wu, Ning Zhang, Hong Zheng, Hanqiang Ouyang, Keyuan Chen, Zhaode Bu*, Xueda Hu*, Jiafu Ji*, Zemin Zhang*

INTRODUCTION: Cancer immunotherapies that target tumor-specific T cells have benefited many cancer patients, but the clinical efficacy varies greatly among different cancer types. Tumor-infiltrating T cells often enter a dysfunctional state, widely known as T cell exhaustion, and the antitumor functions of effector T cells are regulated by multiple factors, including the presence of regulatory T cells (T_{reg} cells). The states and abundances of T cells vary across tumor microenvironments (TMEs) of different cancer types, which may fundamentally influence different clinical parameters such as drug response to immunotherapies.

RATIONALE: To build a high-resolution pan-cancer T cell atlas, we performed single-cell RNA sequencing (scRNA-seq) on tumors, paracancerous tissues, and blood samples from patients of various cancer types and collected additional published scRNA-seq datasets. The diverse data were integrated after correcting confounding factors and batch effects. This atlas was composed of scRNA-seq data from 316 patients across 21 cancer types. T cell receptor (TCR) sequences of individual T cells with gene expression profiles were assembled to characterize the expansion and dynamics of T cells. Various computational methods were applied to investigate the features and abundance of T cells across cancer types.

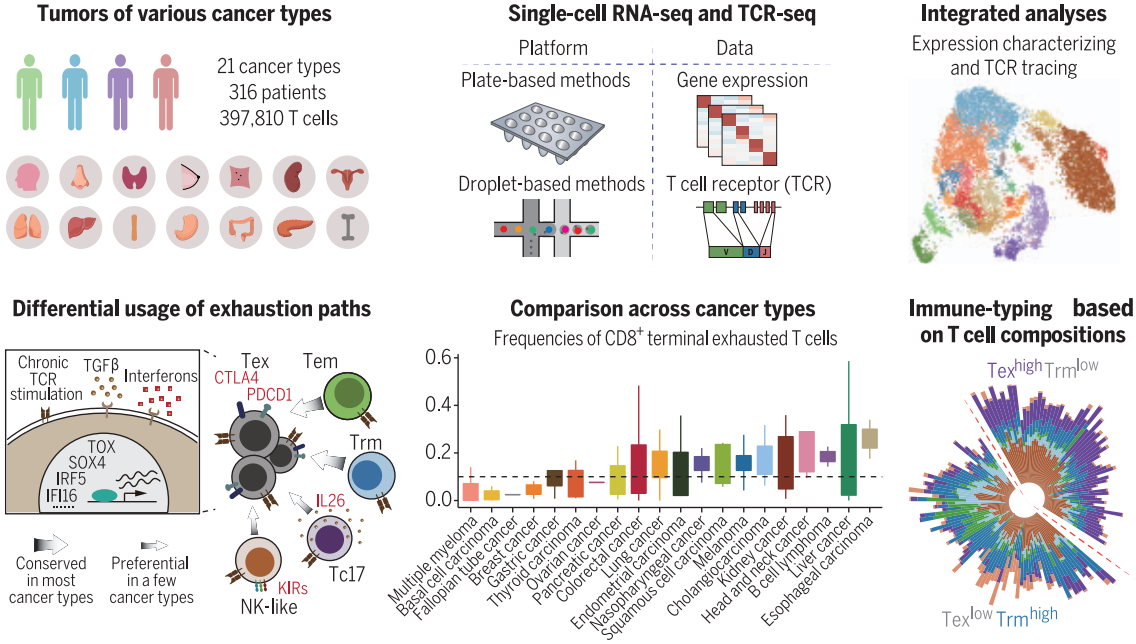
RESULTS: We identified multiple potentially tumor-reactive T cell (pTRT) populations in cancer patients. The states of the pTRTs varied dramatically in the tumor microenvironment of different cancer types. For $CD8^+$ T cells, the major pTRTs were exhausted T cells and exhibited high heterogeneity. We computationally inferred two major developmental paths to T cell exhaustion, through effector memory T cells and tissue-resident memory T cells, respectively, and both were prevalent among cancer types. We also noted the state transitions between terminal exhausted T cells and cells such as natural killer (NK)–like T cells, Type 17 $CD8^+$ T cells (Tc17 cells) cells, and $CD8^+$ T_{reg} cells, but such transitions tend to occur in specific cancer types. For $CD4^+$ T cells, follicular helper T cell (T_{FH})/T helper 1 (T_H1) dual-functional T cells, which appeared to originate from T_{FH} cells, were also notable pTRTs and correlated with the tumor mutation burden. We also found that the transcriptional programs of pTRTs could be affected by transforming growth factor–β (TGF-β) and interferons in the TMEs. The abundances of T cell states vary dramatically depending on cancer types. On the basis of tumor-infiltrating T cell compositions, cancer patients could be immune-typed as a group with high frequencies of terminal exhausted $CD8^+$ T cells and another group with high frequencies of tissue-resident memory $CD8^+$ T cells, and the immune types were associated with clinical traits such as patient survival and responses to immune checkpoint blockade.

CONCLUSION: We depicted the pan-cancer landscape of T cell heterogeneity and dynamics in the TME and established a baseline reference for future temporal or spatial studies associated with cancer treatments. The systematic comparison across cancer types revealed the commonalities and differences of T cell states in different TMEs. Our detailed signature, dynamics, and regulations of tumor-infiltrating T cells will facilitate the development of immunotherapies, and our proposed immune-typing can aid the therapeutic and diagnostic strategies that target T cells. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zemin@pku.edu.cn (Z.Z.); jijiafu@hsc.pku.edu.cn (J.J.); huxueda@abiosciences.com (X.H.); buzhaode@cjcrcn.org (Z.B.)
†These authors contributed equally to this work.
Cite this article as L. Zheng *et al.*, *Science* **374**, eabe6474 (2021). DOI: 10.1126/science.abe6474

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abe6474

Systematic analysis of a human pan-cancer T cell atlas. We analyzed approximately 390,000 T cells from 316 patients of 21 cancer types by means of scRNA-seq. Combining gene expression profiles and T cell receptor sequences, we investigated the heterogeneity and dynamics of tumor-infiltrating T cells and performed a systematic comparison of T cells among cancer types. Additionally, we provided a T cell composition–based immune-typing scheme. KIR, killer cell immunoglobulin-like receptor; IL26, interleukin-26.



RESEARCH ARTICLE

CANCER IMMUNOLOGY

Pan-cancer single-cell landscape of tumor-infiltrating T cells

Liangtao Zheng[1]†, Shishang Qin[2]†, Wen Si[1]†, Anqiang Wang[3], Baocai Xing[4], Ranran Gao[2], Xianwen Ren[2], Li Wang[2], Xiaojiang Wu[3], Ji Zhang[3], Nan Wu[5], Ning Zhang[6], Hong Zheng[7], Hanqiang Ouyang[8,9], Keyuan Chen[8,9], Zhaode Bu[3]*, Xueda Hu[2,10]*, Jiafu Ji[3,11]*, Zemin Zhang[1,2]*

T cells play a central role in cancer immunotherapy, but we lack systematic comparison of the heterogeneity and dynamics of tumor-infiltrating T cells across cancer types. We built a single-cell RNA-sequencing pan-cancer atlas of T cells for 316 donors across 21 cancer types and revealed distinct T cell composition patterns. We found multiple state-transition paths in the exhaustion of CD8$^+$ T cells and the preference of those paths among different tumor types. Certain T cell populations showed specific correlation with patient properties such as mutation burden, shedding light on the possible determinants of the tumor microenvironment. T cell compositions within tumors alone could classify cancer patients into groups with clinical trait specificity, providing new insights into T cell immunity and precision immunotherapy targeting T cells.

Tumor-infiltrating lymphocytes (TILs) are central players in the tumor microenvironment (TME), shaping fundamental clinical properties such as responses to immunotherapies. Immune checkpoint blockade (ICB) has shown tremendous clinical success, but its efficacy varies dramatically across cancer types, suggesting underlying differences of tumor immunity. Within the TME, effector T cells tend to exhibit high expression levels of multiple inhibitory receptors such as PD-1 (programmed cell death 1), TIM3 (T cell immunoglobulin and mucin domain–containing protein 3), TIGIT (T cell immunoreceptor with Ig and ITIM domains), and LAG3 (lymphocyte activating 3) (*1*), which are considered to be hallmarks of a dysfunctional state, widely known as T cell exhaustion. The varied ICB efficacies could be logically linked to the tumor-infiltrating T cell state differences among cancer types, especially the exhaustion differences. In melanoma patients, CD8$^+$ tumor-infiltrating T cells exhibit a linear and continuous progression from predysfunctional cell state to dysfunction (*2*), but in lung cancer patients, there are two pre-exhaustion states that could develop to exhaustion (*3*). Thus, the exhaustion dynamics may differ among TMEs of various cancers. Intrinsically, T cell exhaustion appears to be tightly regulated by several transcription factors (TFs), including TOX (thymocyte selection-associated high mobility group box) (*4*, *5*) and TCF7 (transcription factor 7) (*6*), as well as epigenetic regulators that shape the specific state observed in dysfunctional CD8$^+$ T cells (*7*). In addition, multiple TME factors contribute to the exhaustion phenotype (*8*), and distinct regulatory processes dictating the phenotypes and abundance of T cells may exist within the TMEs of various cancer types. Distinguishable T cell features have been observed in different cancer types. For example, liver and colon cancers have higher fractions of exhausted T cells than that of lung cancer (*9*), and cancer types such as multiple myeloma do not show notable exhausted T cell populations (*10*). However, direct comparative studies have been restricted to only three or four isolated cancer types (*11*, *12*).

We constructed a comprehensive tumor-infiltrating T cell compendium across 21 distinct cancer types through single-cell RNA-sequencing (scRNA-seq). By finding the commonalities and differences of tumor-infiltrating T cells, we aim to reveal the "pan-cancer" features of the T cell states, dynamics, and regulation.

Results

Construction of a pan-cancer single-cell transcriptome atlas of T cells

We compiled a single-cell transcriptome atlas of T cells across 21 cancer types (Fig. 1A). After stringent quality-control filtering, this atlas contained data for 397,810 T cells from 316 donors—derived from their tumors, adjacent normal tissues, and peripheral blood—of which 46.4% cells were newly sequenced in this study, whereas others were from previously published datasets (table S1). We integrated the diverse data generated from multiple technologies (fig. S1) on the basis of "minicluster" and the batch effect correction algorithm Harmony (figs. S2, A to D, and S3, A to D) (*13*, *14*). Both visual and quantitative evaluations showed that cells were well mixed in the integrated data (figs. S2E and S3E). T cell receptor (TCR) sequences from individual cells were assembled for data generated through 10x VDJ (*15*) and Smart-Seq2 (*16*) protocols. A total of 168,901 cells from 92,533 clonotypes spanning 87 donors from 15 cancer types harbored at least one pair of productive TCR α chain and β chain, of which 53.9% were clonal cells (with identical TCR pairs found in at least two cells), corresponding to 14,631 expanded clonotypes (fig. S4).

A total of 17 CD8$^+$ and 24 CD4$^+$ metaclusters were identified, all of which were shared by at least 80% of cancer types (Fig. 1, B and C, and figs. S2F and S3F). Analysis of expression signatures of these metaclusters revealed the presence of both previously described T cell subtypes and new groups, including *granzyme K–positive* (*GZMK*$^+$) effector memory cells (T_{em} cells), terminally differentiated effector memory or effector cells (T_{emra} cells), and interferon-stimulated genes (ISG)–positive T cells in both CD4$^+$ and CD8$^+$ compartments; killer cell immunoglobulin-like receptor (KIR)–positive natural killer (NK)–like T cells, *ZNF683*$^+$*CXCR6*$^+$ tissue-resident memory T cells (T_{rm} cells), and four exhausted CD8$^+$ T cell (T_{ex} cell) populations in the CD8$^+$ compartment; and three follicular helper T cell (T_{FH} cell)–related populations [C-X-C motif chemokine receptor 5-positive (CXCR5+) pre-T_{FH}, classical *IL21*$^+$ T_{FH}, and *IFNG*$^+$ T_{FH}/T helper 1 (T_H1) dual-functional T cells] and four regulatory T cell (T_{reg} cell) populations in the CD4$^+$ compartment (fig. S5 and table S2). For the CD8$^+$ metacluster c16, nearly half of the cells harbored the semi-invariant TCR α chains of mucosal-associated invariant T cells (MAIT) (fig. S6A), and cells with or without such TCR α chains both highly expressed genes related to Type 17 CD8$^+$ T cells (Tc17 cells) (fig. S6B) (*17*, *18*),

[1]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. [2]BIOPIC, Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Peking University, Beijing 100871, China. [3]Gastrointestinal Cancer Center, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. [4]Department of Hepatopancreatobiliary Surgery I, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. [5]Department of Thoracic Surgery II, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. [6]Department of Urology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. [7]Department of Gynecologic Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. [8]Department of Orthopaedics, Peking University Third Hospital, Beijing 100191, China. [9]Beijing Key Laboratory of Spinal Disease Research, Peking University Third Hospital, Beijing 100191, China. [10]Analytical Biosciences Limited, Beijing 100084, China. [11]Department of Biobank, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China.
***Corresponding author. Email: zemin@pku.edu.cn (Z.Z.); jijiafu@hsc.pku.edu.cn (J.J.); huxueda@abiosciences.com (X.H.); buzhaode@cjcrcn.org (Z.B.)**
†These authors contributed equally to this work.

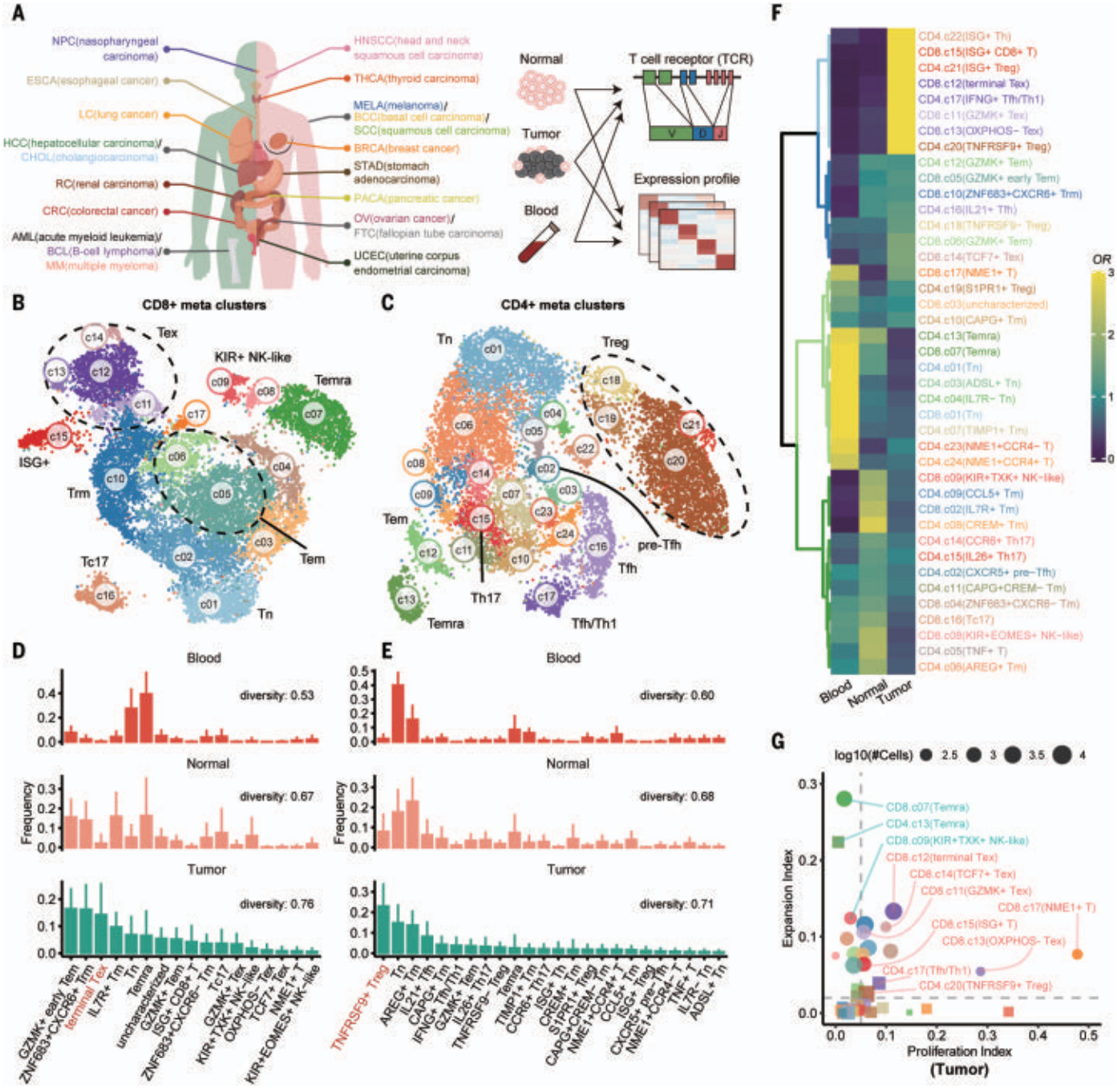


Fig. 1. Pan-cancer T cell profile at the single-cell resolution.
(**A**) Schematics of pan-cancer single-cell transcriptome and TCR profiling of T cells. (**B**) UMAP visualization of CD8$^+$ T cell metaclusters. Selective metaclusters are highlighted by using their functional annotation: T_{ex}, exhausted T cells; ISG, interferon-stimulated genes; T_{emra}, terminally differentiated effector memory or effector; T_{em}, effector memory T cells; T_{rm}, tissue-resident memory T cells; T_n, naïve T cells; and KIR, killer cell immunoglobulin-like receptors. (**C**) The same plot as in (B) applied to CD4$^+$ T cells. (**D**) Bar plots showing the CD8$^+$ T cell compositions in the blood (n = 46 patients), normal (n = 82 patients), and tumor (n = 197 patients) from treatment-naïve patients (mean ± SD). Average diversity measured with the Shannon equitability index for each tissue is shown. A high index indicated similar frequencies across different metaclusters. (**E**) The same plots as in (D) applied to CD4$^+$ T cells from blood (n = 15 patients), normal (n = 51 patients), and tumor (n = 163 patients). (**F**) Heatmap showing the ORs of metaclusters occurring in each tissue. OR > 1.5 indicates that the metacluster is preferred to distribute in the corresponding tissue. Hierarchical clustering based on cosine distance is applied for rows. The naming, numbering, and colors of the metaclusters are in accordance with (B) and (C). (**G**) Scatter plot showing the expansion index and the proliferation index of metaclusters in the tumor. Metaclusters of high expansion (expansion index > 0.1, P < 0.01) are highlighted in blue, and tumor-enriched metaclusters with significant expansion (P < 0.01) and proliferation (index > 0.05) are highlighted in red. Size indicates the number of cells (in log10 scale).

indicating that this metacluster contained both MAIT and non-MAIT Tc17 cells (designated Tc17 hereafter).

The compositions of T cells from different tissue-of-origin of treatment-naïve patients displayed prominent differences because the diversities measured with Shannon equitability index (*19*) in normal tissues and tumors were significantly higher than that in the blood ($P < 0.01$, two-sided Wilcoxon tests), and for $CD8^+$ T cells, the diversity was even higher in tumors than that in normal tissues (Fig. 1, D and E, and fig. S7 A and B). $CD8^+$ cells in the tumor were featured by the emergence of exhausted T cells (Fig. 1D and fig. S7C), whereas among $CD4^+$ T cell populations, the most abundant population was the *TNFRSF9*$^+$ T_{reg} cell, which showed significantly lower frequencies in both blood and normal tissues ($P < 0.01$, two-sided Wilcoxon tests) (Fig. 1E and fig. S7D).

To distinguish the T cells reacting to tumors from bystander T cells, we jointly analyzed the features of tissue distribution, transcriptional phenotypes, proliferation, and clonal expansion. The characteristics of proliferation and clonal expansion of tumor-enriched T cells have been viewed as evidence of their tumor reactivity (*20*). From the odds ratio (OR) analysis, the naïve T cells and T_{emra} cells (both $CD4^+$ and $CD8^+$) showed a strong distribution preference in blood, whereas *TNFRSF9*$^+$ T_{reg} cells, T_{FH}/T_H1 cells, $CD8^+ISG^+$ T cells, and four $CD8^+$ T_{ex} cell groups appeared to be tumor-enriched (Fig. 1F and table S2). On the basis of the STARTRAC (single T cell analysis by RNA-seq and TCR tracking) analysis (*9*), such tumor-enriched metaclusters exhibited expansion and ongoing proliferation (Fig. 1G), implying their clonal expansion in response to tumor antigens. In addition, most of the tumor-enriched metaclusters with high expansion and proliferation indices (except for *TCF7*$^+$ T_{ex} cells and OXPHOS- T_{ex} cells) tended to exhibit high activities in TCR signaling, confirming their high antigen reactivities (fig. S8). At the individual tumor level, we identified groups of cells that share clonotypes with specific expanded TILs that had high TCR signaling or proliferation, and these cells were collectively considered as potentially tumor-reactive T cells (pTRTs). For pTRTs in tumors, the most frequently observed cell states were terminal T_{ex} cells and *TNFRSF9*$^+$ T_{reg} cells for $CD8^+$ and $CD4^+$ T cells, respectively (fig. S9A), although their occurrence varied among different cancer types (fig. S9C). Meanwhile, the T_{emra} cells of both $CD4^+$ and $CD8^+$ compartments exhibited significant expansion ($P <$ 0.01, permutation test) but low proliferation (<2%) in all tissues (Fig. 1G and fig. S10). The STARTRAC migration indices, which quantify the extent of tissue migration, revealed that both $CD8^+$ and $CD4^+$ T_{emra} cells had the highest mobility between blood and normal or tumor tissues in most tested cancer types (fig. S11, A and B). Those observations suggested that those T cells were activated and expanded outside the tumor and circulated in the blood. Compared with healthy donors, cancer patients harbored more $CD8^+$ T_{emra} cells in the blood (fig. S11C). In addition, in certain tumors, most of the pTRTs were T_{emra} cells (fig. S9A). Further, the most frequent cell state of pTRTs in the blood was also T_{emra} cell (fig. S9B), and this pattern applied to multiple cancer indications (fig. S9D).

Taken together, potentially tumor-reactive T cells emerged—T_{FH}/T_H1 cells, *TNFRSF9*$^+$ T_{reg} cells, $CD8^+ISG^+$ T cells, and four T_{ex} cell populations, representing a local antitumor immune response—whereas the expanded $CD8^+$ T_{emra} cells might also harbor tumor-specific TCRs, which is consistent with the notion of a systemic immune response (*21*).

Common themes of CD8$^+$ T$_{ex}$ cell heterogeneity and dynamics

In the $CD8^+$ compartment, the major potentially tumor-reactive metaclusters were the four $CD8^+$ T_{ex} cell populations, all highly expressing multiple exhaustion markers, including *TOX*, *TIGIT*, *CTLA4* (cytotoxic T lymphocyte–associated protein 4), and *TNFRSF9* (TNF-receptor superfamily 9), but they differed in gene expression and pathway activities (Fig. 2A and fig. S12). A major population was terminal T_{ex} cells, which exhibited higher expression of the gene *ENTPD1* (ectonucleoside triphosphate diphosphohydrolase 1), which is related to terminal differentiation (*22*, *23*). The terminal T_{ex} cells also highly expressed *IFNG* (interferon-γ) and *GZMB* (granzyme B), implying its intrinsic antitumor effector potential, and certain genes with unknown roles in T cell exhaustion, including *MYO1E* (myosine IE) and *MYO7A* (Myosin VIIA). A relatively rare T_{ex} cell population was *TCF7*$^+$ T_{ex} cells, which had a lower level of *HAVCR2* (hepatitis A virus cellular receptor 2) and *LAG3* (lymphocyte activating 3) but specifically expressed a high level of *TCF7* (Fig. 2A). *TCF7* has been considered to be the key regulator of stem-like T cells in tumors (*24*). Additionally, *TCF7*$^+$ T_{ex} cells highly expressed *CD200*, *GNG4* (G protein subunit γ 4), *IGFBP4* (insulin-like growth factor binding protein 4), *IGFL2* (insulin growth factor-like family member 2), and genes related to lymph node migration [such as *CCR7* (C-C motif chemokine receptor 7) and *SELL* (selectin L)] (Fig. 2A).

Next, we combined gene expression and TCR data to dissect the trajectories of T cell exhaustion. First, on a global scale, the diffusion map (*25*) and RNA velocity (*26*) showed that $CD8^+$ T cells could develop from naïve T cells to either T_{emra} or T_{ex} cells (fig. S13, A and B), which is consistent with previous reports (*3*, *9*). Second, for the T_{ex} branch, combining Monocle3 (*27*), RNA velocity, and graph inference based on uniform manifold approximation and projection (UMAP) (*28*) and STARTRAC pairwise transition indexes (pTrans) (*14*), we inferred two paths from naïve to T_{ex} cell: the first path (P1) going through *GZMK*$^+$ T_{em} cells [naïve cells to *IL7R*$^+$ memory T cells (T_m cells) to *GZMK*$^+$ T cells to terminal T_{ex} cells], and the second (P2) going through *ZNF683*$^+$ T_{rm} cells (naïve cells to *IL7R*$^+$ T_m cells to *ZNF683*$^+$*CXCR6*$^+$ T_{rm} cells to terminal T_{ex} cells) (Fig. 2, B and C, and figs. S13, C to E, and S14). The state transition to T_{ex} cells from both paths could also be observed in individual tumors (fig. S15). Although certain tumors exhibited preferential usage of P1 or P2, the state transition of both paths was high in other tumors, implying that both T_{em} and T_{rm} cells were involved in the antitumor immunity in those tumors (fig. S15C). The terminal T_{ex} cells had moderate pTrans with ISG^+ T cells, which in turn were highly connected with multiple metaclusters (fig. S16A). The ISG^+ state was not an independent state, but a mixture of T_{em}, T_{rm}, and other cells (fig. S17). In addition, TCR clonotypes that contained multiple cell states—including *ZNF683*$^+$*CXCR6*$^+$ T_{rm} cells or *GZMK*$^+$ T_{ex} cells, ISG^+ T cells, and terminal T_{ex} cells—could be clearly identified in tumors (fig. S16, B and C). Because the ISG^+ state represents an activation state possibly driven by TCR-triggered interferon-γ (IFN- γ) or induced by interferons directly (*29*), these observations suggested that nonexhausted T cells in P1 or P2 could become interferon-responsive, before entering "exhaustion," which is reminiscent of observations in chronic virus infections in which the $CD8^+$ T cell–intrinsic type I interferon signaling skewed the differentiation to a more terminal effector state (*30*) or exhaustion state (*31*).

It has been hypothesized that the progenitor cells that express *CXCR5* or *TCF7* give rise to terminal T_{ex} cells (*23*, *32*, *33*). We found that *TCF7*$^+$ T_{ex} cells had a strong state transition connection with *GZMK*$^+$ T_{ex} cells and even terminal T_{ex} cells (fig. S18). A fraction of *GZMK*$^+$ early T_{em} cells also expressed high levels of *CXCR5* or *TCF7* (fig. S19A). Those *CXCR5*$^+$ or *TCF7*$^+$ cells from *GZMK*$^+$ early T_{em} cells were located near the branch point to T_{emra} or T_{ex} cells (fig. S13, A and D), representing cells with the developmental potential to different fates. By contrast, the *TCF7*$^+$ T_{ex} cells more likely represented committed T_{ex} cells with certain stemness. Compared with *GZMK*$^+$ early T_{em} cells, those *CXCR5*$^+$ or *TCF7*$^+$ cells from the *TCF7*$^+$ T_{ex} cell population had a higher frequency of cells expressing *TOX* and inhibitory receptors such as *PDCD1*, *TIGIT*, and *CTLA4* (fig. S19B). *TCF7*$^+$ T_{ex} cells did not have a strong state transition with *GZMK*$^+$ early T_{em} cells but were highly connected with *NME1*$^+$ (NME/NM23 nucleoside diphosphate kinase 1)

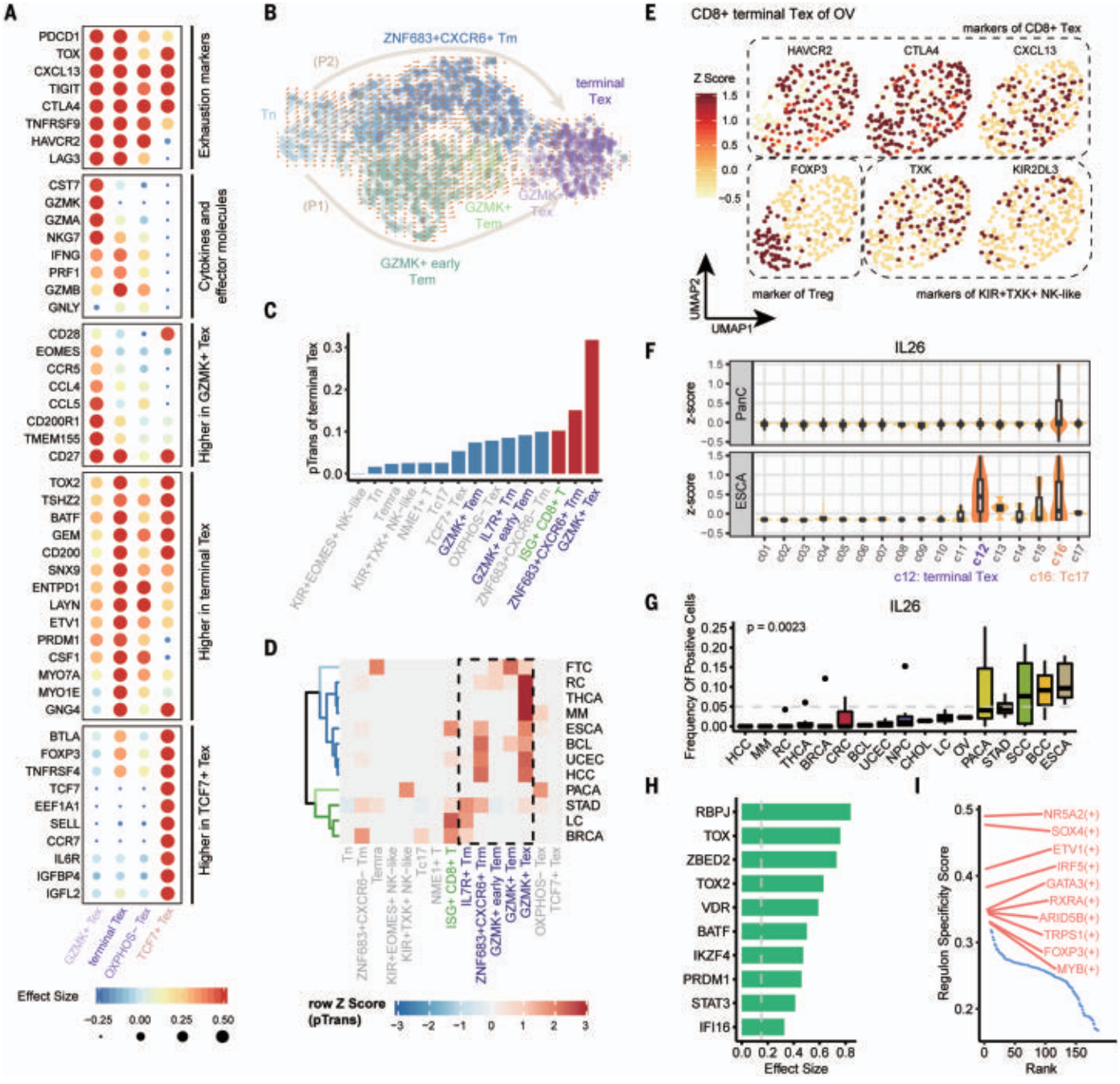


Fig. 2. Heterogeneity and dynamics of CD8$^+$ exhausted T cells. (**A**) Dot plot showing the expression of signature genes of the four CD8$^+$ exhausted T cells. Both color and size indicate the effect size. (**B**) RNA velocities overlaid on UMAP showing two major state transition paths from naïve to exhaustion. Arrows on a grid show the RNA velocity field, and dots are colored by metaclusters. (**C**) The pair-wise transition index (pTrans) of terminal T_{ex} cells. Three metaclusters, with which terminal T_{ex} cells are highly connected (pTrans > 0.1), are highlighted. (**D**) Heatmap showing the pTrans between terminal T_{ex} cells and other metaclusters, stratified by cancer types. Color represents the *z*-score scaled pTrans value by row. Metaclusters belonging to the two major paths to exhaustion are highlighted with a dashed line box. pTrans are whited out if the value < 0.1, representing low transition potential. (**E**) UMAP visualization showing the expression of selected marker genes in terminal T_{ex} cells of ovarian cancer (OV). Color represents *z*-score–scaled gene expression values. (**F**) Violin and boxplot showing the expression of *IL26* across metaclusters in pan-cancer (PanC) and esophageal cancer (ESCA). The expression pattern in pan-cancer serves as a "reference," comparing with the "ectopic" expression of genes *IL26* in the terminal T_{ex} cells of ESCA. (**G**) Boxplot showing the frequency of *IL26*$^+$ cells in terminal T_{ex} cells among different cancer types. The *P* value is calculated with the Kruskal-Wallis test. (**H**) Bar plot showing the effect size of the top 10 universal TFs of terminal T_{ex} cells. (**I**) Scatter plot showing the specificity scores of regulons of terminal T_{ex} cells. The top 10 regulons are highlighted.

T cells (figs. S18). Because of the scarcity of these two cell populations, it remains to be seen whether *TCF7*$^+$ T_{ex} cells were derived from *NME1*$^+$ cells. Thus, we identified a T cell exhaustion path that cannot be defined with either *CXCR5* or *TCF7* alone. Taken together, we evidenced a more complex cellular pathway to T cell exhaustion.

Distinct paths to T cell exhaustion across different cancer types

To examine the prevalence of the major exhaustion paths among cancer types, we further stratified the TCR sharings by cancer types. In 11 out of the 12 cancer types with strong TCR sharings between terminal T_{ex} cells and other metaclusters, the terminal T_{ex} cells exhibited high pTrans values with *GZMK*$^+$ T_{ex} cells, *ZNF683*$^+$ T_{rm} cells, or other metaclusters in P1 or P2 (Fig. 2D), implicating P1 and P2 as the universal paths for T cell exhaustion. We also observed strong TCR sharings between terminal T_{ex} cells and metaclusters not in P1 or P2 in certain cancer types, including those with KIR$^+$*TXK*$^+$ NK-like T cells in pancreatic cancer and those with Tc17 in breast cancer (Fig. 2D). These TCR sharing patterns suggested the presence of heterogeneous paths to T cell exhaustion in addition to the common paths through T_{em} and T_{rm} cells.

In addition to intercluster TCR sharing, the developmental connections between terminal T_{ex} cells and other cell populations could be confirmed by the expression of partial signature of nonexhaustion states in terminal T_{ex} cells. A subset of terminal T_{ex} cell of ovarian cancer expressed T_{reg} cell–dominant TF *FOXP3* (forkhead box P 3) (Fig. 2E and fig. S20A), and these CD8$^+$*FOXP3*$^+$ T cells tended to share TCRs with *CXCL13*-expressing cells (fig. S20B), implying intracluster transition between CD8$^+$*FOXP3*$^+$ T cells and *CXCL13*-expressing cells. Similarly, a subset of terminal T_{ex} cell of ovarian cancer expressed *KIR2DL3* (killer cell immunoglobulin like receptor, two Ig domains, and long cytoplasmic tail 3) and *TXK* (TXK tyrosine kinase), which are part of the signature genes of KIR$^+$*TXK*$^+$ T cells (Fig. 2E and fig. S20A), and these cells shared TCRs with *CXCL13*-expressing cells (fig. S20B). In addition, the RNA velocity of *KIR2DL3*$^+$ cells pointed to *CXCL13*$^+$ cells (fig. S20C), suggesting the transition from KIR$^+$*TXK*$^+$ to exhaustion. Such state transitions might also occur in other cancer types because *FOXP3*-expressing cells, the KIR$^+$ T cell signature (fig. S20D), and similar TCR sharing patterns (fig. S20B) could be found in T_{ex} cells of multiple cancer types, although their frequencies varied among cancer types.

We further identified genes with cancer-type preference in terminal T_{ex} cells, including *IL26* (interleukin 26), *IL17A*, and *RORC* (RAR-related orphan receptor C). These genes are primarily expressed in CD4$^+$ T_H17 cells or CD8$^+$ Tc17 (*18*). We confirmed their expression in Tc17 (table S3), but in certain tumors of esophageal cancer, squamous cell carcinoma, and stomach adenocarcinoma, they were also expressed in a fraction of terminal T_{ex} cells (fig. S21A), especially for *IL26* (Fig. 2, F and G). Additionally, expanded clonotypes containing *IL26*-, *IL17A*-, or *RORC*-expressing T_{ex} cells tended to contain *CXCL13*-expressing T_{ex} cells (fig. S21B), implying the state transition connection between *IL26*-, *IL17A*-, and *RORC*-expressing cells and *CXCL13*-expressing cells. Furthermore, a fraction of the terminal T_{ex} cells expressed the semi-invariant α chain of MAIT, the signatures of exhaustion, and also a partial signature of Tc17 (fig. S22). Thus, there appeared to be a subset of T_{ex} cells with the capacity of secreting cytokines of type 17 response, which were likely derived from Tc17. The preferential expression of those genes—exemplified by *IL26*, which functions as an inflammatory mediator that induces the production of inflammatory cytokines in the mucosal tissues (*34*)—implied the multifunctional characteristic of T_{ex} cells in certain cancer types.

Universal and cancer type–specific transcriptional regulation of CD8$^+$ T cell exhaustion

The identification of *TOX* as a critical TF of T cell exhaustion has sparked interest in finding additional regulators (*4*, *5*). We systematically identified TFs associated with exhaustion. The signature genes of terminal T_{ex} cells encoding TFs and with significantly high expression [effect size > 0.15, false discovery rate (FDR) < 0.01 by meta-analysis] in >80% of cancer types were designated universal T_{ex} cell regulators. *TOX*, *TOX2*, *RBPJ* (recombination signal binding protein for immunoglobulin κ J region), *ZBED2* (zinc finger BED-type containing 1), *PRDM1* (PR domain zinc finger protein 1), *VDR* (vitamin D receptor), *IKZF4* (IKAROS family zinc finger 4), *BATF* (basic leucine zipper ATF-like transcription factor), *STAT3* (signal transducer and activator of transcription 3), and *IFI16* (interferon γ inducible protein 16) were ranked by effect size as the top 10 universal TFs (Fig. 2H, fig. S23A, and table S3). *TOX*, for example, showed statistical significance in all cancer types. Also, SCENIC analysis (*35*), which reconstructs regulons (TFs and their target genes), identified TFs that target *TOX*, including *NR5A2* (nuclear receptor subfamily 5 group A member 2), *ETV1* (ETS variant transcription factor 1), and *ARID5B* (AT-rich interaction domain 5B), which exhibited high regulon specificity in terminal T_{ex} cells (Fig. 2I and table S4) and showed statistical significance in >50% of cancer types (fig. S23A). Supporting such *TOX* regulation, a reanalysis of single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq) data of basal cell carcinoma (*36*) revealed that the high-accessibility peaks in the promoter or distinct intragenic enhancer of *TOX* in the terminal T_{ex} cell matched the motifs of these three TFs (fig. S23B). Although the exhaustion of T_{em} and T_{rm} cells shared multiple up-regulated TFs—such as *TOX*, *RBPJ*, and *ETV1*—pointing to these as common driving forces of T cell exhaustion, different exhaustion paths might also differ in the usage of TFs. For example, *BHLHE40* (basic helix-loop-helix family member E40) and *ZBTB32* (zinc finger and BTB domain containing 32) were featured in the late stage of P1, whereas *STAT1* and *IKZF3* were higher in the late stage of P2 (fig. S24). These observations revealed a finer regulation process of exhaustion.

We also identified key transcription regulators not previously known to be associated with T cell exhaustion, including *SOX4* and *FOXP3* (Fig. 2I and table S3). *SOX4*, in particular, had a high regulon specificity score and showed statistical significance in two-thirds of all cancer types (fig. S23A). As a downstream target of transforming growth factor–β (TGF-β), *SOX4* has been reported to play an important role in *CXCL13*-producing T_H cells (*37*) and to up-regulate the expression of exhaustion marker *ENTPD1* (CD39) in T_{reg} cells (*38*). Thus, we inferred that *SOX4* exerted similar functions in T_{ex} cells, although this should be further verified. *FOXP3*, also with a high regulon specificity score, showed statistical significance in only 47% of all cancer types (fig. S23A). Because *FOXP3* is important for T_{reg} cell functions, the connection between CD8$^+$*FOXP3*$^+$ T cells (CD8$^+$ T_{reg} cells) and exhaustion deserves further investigation. The frequencies of cells expressing *SOX4* or *FOXP3* varied significantly among cancer types ($P < 0.01$, Kruskal-Wallis tests) (figs. S20D and S23C), reflecting the differential impact of distinct TMEs on the phenotypes of T_{ex} cells.

Properties of potentially tumor-reactive T cells in the CD4$^+$ compartment

In the CD4$^+$ compartment, the major potentially tumor-reactive metaclusters were *IFNG*$^+$ T_{FH}/T_H1 and *TNFRSF9*$^+$ T_{reg} cells. The global diffusion map and RNA velocity analyses revealed that CD4$^+$ T cells could develop from naïve T cells to T_{emra} cells, T_{FH}/T_H1 cells, or *TNFRSF9*$^+$ T_{reg} cells separately (Fig. 3A). To gain a finer resolution of their developmental trajectories, we performed trajectory inference on each direction separately. The two T_{FH} cell–related metaclusters showed a gradual transition process from the classical *IL21*$^+$ T_{FH} cell to *IFNG*$^+$ T_{FH}/T_H1 cells (Fig. 3, A and B, and fig. S25, A and B). Along this transition process, the type I response-related cytokines and cytotoxic effector molecules—including *IFNG*, *GZMB*, and *PRF1*—significantly increased (FDR < 0.01,

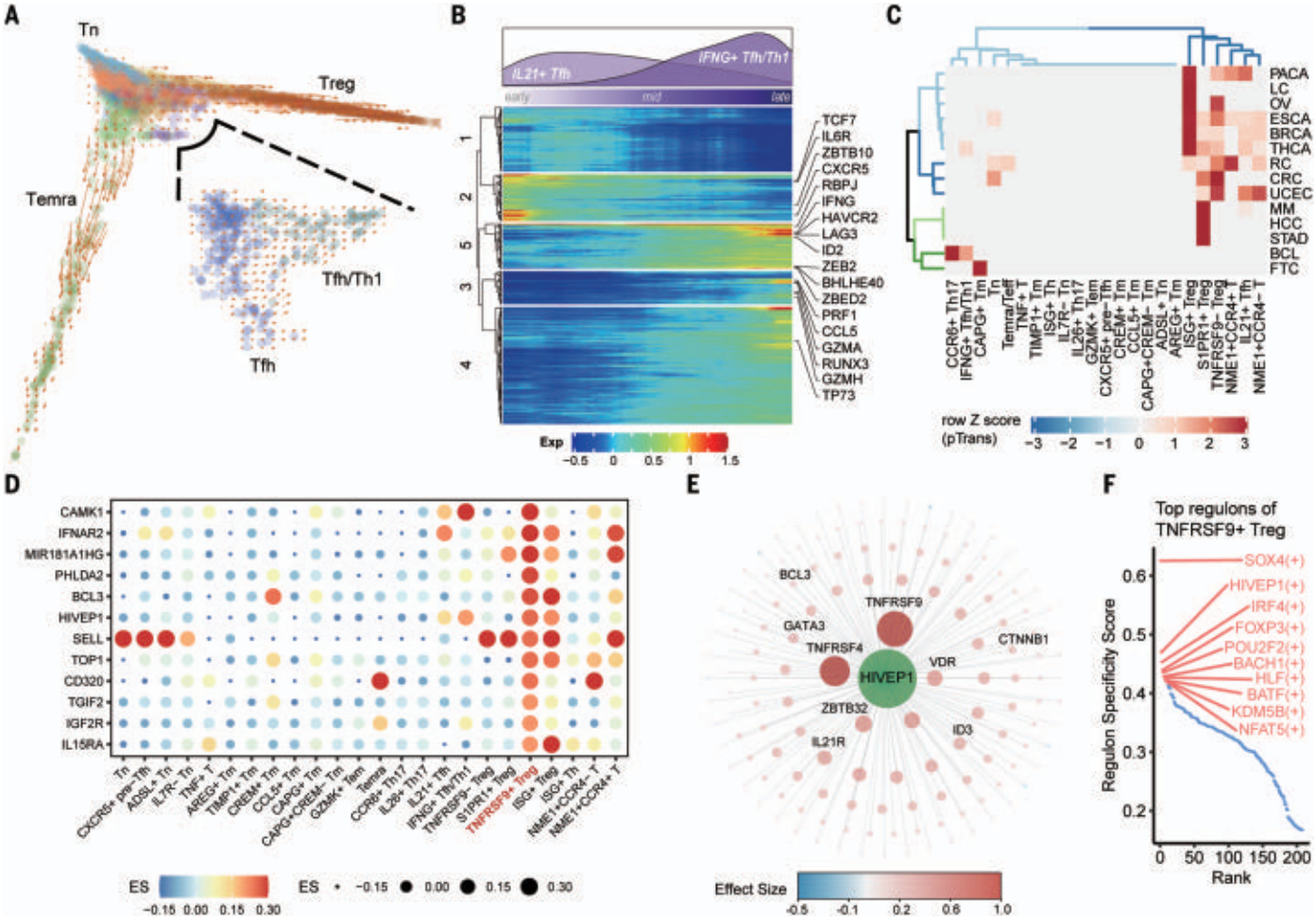


Fig. 3. Properties of potentially tumor-reactive CD4$^+$ T cells. (**A**) Diffusion map of CD4$^+$ T cells. Arrows on a grid show the RNA velocity field. (Inset) A similar diffusion map to zoom in the two T_{FH} cell metaclusters. (**B**) Heatmap showing genes with significant expression (absolute coefficient > 0.5 and FDR < 0.01, generalized additive model) changes along with the pseudotime. Color represents the *z*-score–scaled expression. The density plot of the distribution of the two T_{FH} cell metaclusters along the pseudotime is shown on top of the heatmap. (**C**) Heatmap showing the pTrans between *TNFRSF9*$^+$ T_{reg} cells and other metaclusters, stratified by cancer types. Color represents the *z*-score–scaled pTrans value; pTrans are whited out if the value < 0.01. (**D**) Dot plot showing expression of representative signature genes of *TNFRSF9*$^+$ T_{reg} cells. Both color and size indicate the effect size (ES). (**E**) The transcriptional regulatory network showing the target genes of TF *HIVEP1*. Color represents the effect size. (**F**) Scatter plot showing the specificity scores of regulons of *TNFRSF9*$^+$ T_{reg} cells. The top 10 regulons are highlighted.

generalized additive model) (Fig. 3B). Also, the TF *RUNX3* (runt-related transcription factor 3) exhibited elevated expression at a point at which a high density of T_{FH}/T_H1 cells emerged (fig. S25C), which is consistent with a previous report that *RUNX3* regulated the cytotoxic phenotype in CD4$^+$ cytotoxic T cells (*39*). TF *TP73* (tumor protein P73) appeared at an earlier point of pseudotime (fig. S25C) and was identified as the regulator of a regulon with high specificity in T_{FH}/T_H1 cells (fig. S25D). These observations suggested other key players in acquiring and maintaining the phenotype of T_{FH}/T_H1 cells.

For T_{reg} cells, a trajectory from *TNFRSF9*$^-$ T_{reg} cell to *TNFRSF9*$^+$ T_{reg} cell emerged (Fig. 3A and fig. S26, A to D), indicating a gradual transition from the resting state (*TNFRSF9*$^-$) to activated state (*TNFRSF9*$^+$). The ISG$^+$ T_{reg} cells were located at the center of the trajectory, suggesting that a fraction of T_{reg} cells responded to type I interferons in TME during activation, which is consistent with a recent report suggesting high ISG as a feature of intermediate state during CD4$^+$ T cell activation (*29*). Such a T_{reg} cell developmental trajectory was common across cancer types (Fig. 3C). At the pan-cancer level, we did not observe obvious induction of T_{reg} cells from non–T_{reg} cell conventional T_H cells, but *TNFRSF9*$^+$ T_{reg} cells exhibited certain state transition potentials with non–T_{reg} cells in a few cancer types (Fig. 3C). For example, the *TNFRSF9*$^+$ T_{reg} cells were mainly connected with *CCR6*$^+$ T_H17 and T_{FH}/T_H1 cells in B cell lymphoma, but instead had a connection with T_{FH} cells in the uterine corpus endometrial carcinoma and pancreatic cancer. Thus, various conventional CD4$^+$ T cell populations had conversion relationships with T_{reg} cells, but such conversion patterns were diverse and varied among cancer types.

We identified multiple *TNFRSF9*$^+$ T_{reg} cell signature genes that have not been previously found (fig. S27), including those encoding membrane proteins with kinase activities [*CAMK1* (calcium/calmodulin dependent protein kinase I) and *IGF2R* (insulin like growth factor 2 receptor)], cytokine receptor [*IL15RA* (interleukin 15 receptor α)], known drug targets [*IFNAR2* (interferon α and β receptor subunit 2) and *TOP1* (DNA topoisomerase I)], and TFs [*TGIF2* (TGFB-induced factor 2 protein) and *HIVEP1* (HIVEP zinc finger 1)] (Fig. 3D). *HIVEP1* was inferred as a key regulator of 143

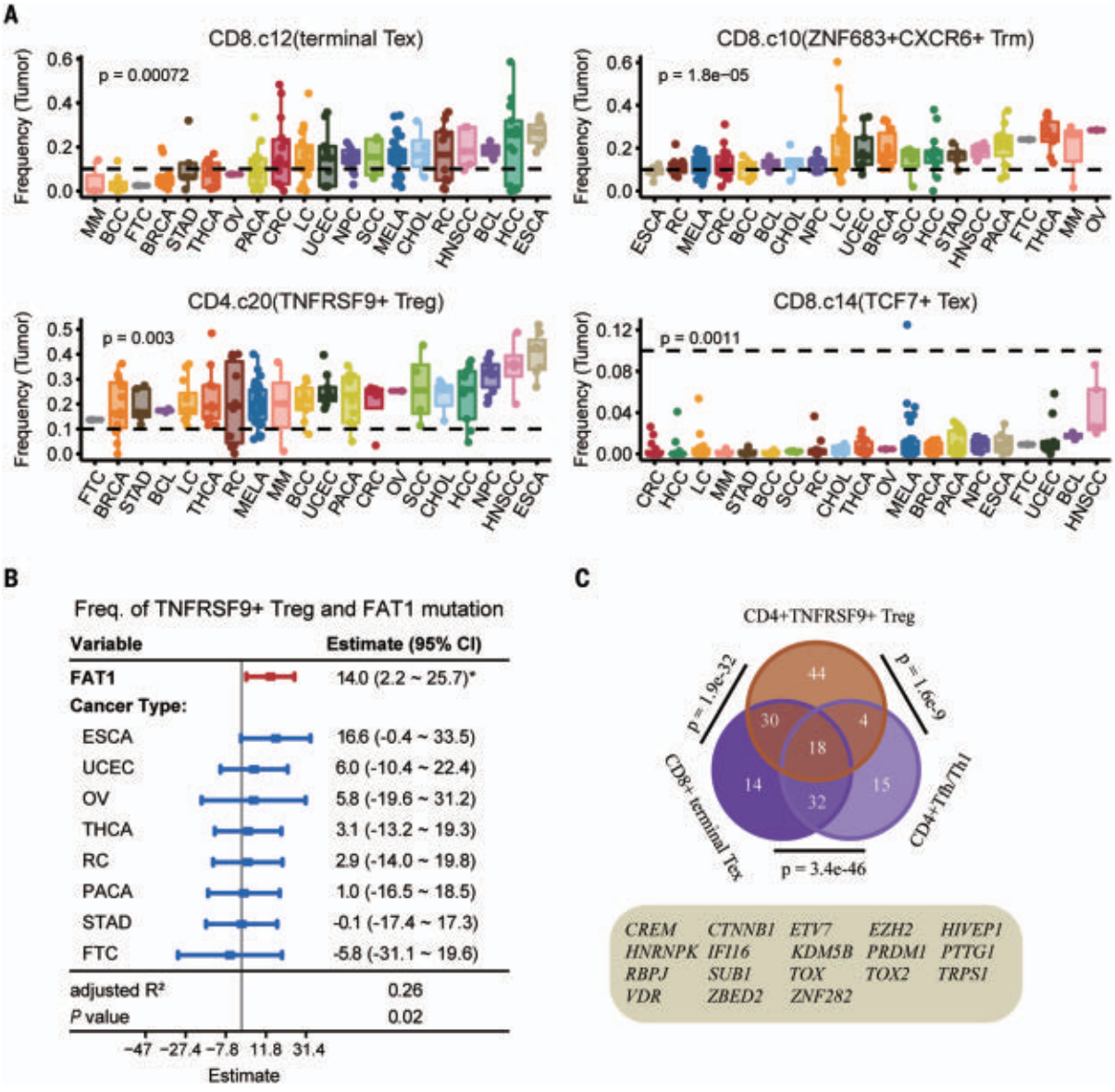


Fig. 4. TME shaping the landscape of tumor-infiltrating T cells. (**A**) Box plots showing the frequencies of metaclusters in tumors across cancer types. Only four metaclusters with significant differences (ANOVA, $P < 0.05$) among cancer types are shown (fig. S30). (**B**) Forest plot showing the association between *FAT1* mutation and the frequency of *TNFRSF9*$^+$ T_{reg} cells in the tumor. The estimated coefficients and their 95% confidence intervals, the goodness of fitting (adjusted R^2), and the significance of the model are reported. *$P < 0.05$. (**C**) Venn diagram illustrating the overlap of signature genes encoding TFs of three metaclusters. The *P* values are calculated with hypergeometric tests, and the 18 signature genes shared by all the three metaclusters are highlighted.

target genes, most of which were also signature genes of *TNFRSF9*$^+$ T_{reg} cells, including *TNFRSF4*, *TNFRSF9*, *ID3* (inhibitor of DNA binding 3), *IL21R*, and *VDR* (Fig. 3, E and F, and table S4). *HIVEP1* accelerated its expression in a late stage of the T_{reg} cell trajectory, which is distinct from the pattern of other known T_{reg} cell TFs such as *FOXP3* and *BATF* (fig. S26E) that were "turned on" at an early stage of the T_{reg} cell trajectory.

TME shaping the landscape of tumor-infiltrating T cells

We examined extrinsic factors associated with distinct T cell compositions in the tumor. Although we did not find a tight association between the frequency of each metacluster with age, gender (fig. S28A), or the clinical stages (fig. S28B), we detected a strong association between the frequency of *IL21*$^+$ T_{FH} cells and the body mass index (BMI) [proportion of variance explained (PVE) >20%] (fig. S28C). Additionally, certain metaclusters clearly exhibited tissue specificity, supporting the effect by host tissues, particularly the liver-enriched Tc17 cells (figs. S28D and S29).

Cancer types exert an extensive impact on the frequencies of T cell populations because we observed distinct T cell distribution patterns across cancer types (Fig. 4A and figs. S28E and S30). The median frequencies of terminal T_{ex} cells, for example, ranged from highly abundant (26.64% of CD8$^+$ T cells) in esophageal carcinoma to barely detectable in multiple myeloma (0.15%) [analysis of variance (ANOVA) $P = 0.00072$, PVE 13.1%]. Similarly, although the median frequencies of *TNFRSF9*$^+$ T_{reg} cells were high in all cancer types (>10%), the variability was still high across cancer types (ANOVA $P = 0.003$, PVE 13.2%), with esophageal carcinoma as well as head and neck cancer exhibiting approximately twofold higher frequencies than that of breast and stomach cancers (Fig. 4A). Additionally, although the *TCF7*$^+$ T_{ex} cell was a rare population, with

the head and neck squamous cell carcinoma showing the highest median value (2.63%), a subset of melanoma samples could reach a frequency of 12.5%. Further, although the $CD8^+$ T_{emra} cell was underrepresented in the tumor, a few tumor samples of lung cancer and melanoma and a large fraction of renal tumors showed much higher frequencies of such cells (fig. S30). Thus, therapeutic strategies targeting the above cell subtypes should consider the variability across cancer types.

The tumor mutation burden (TMB) has been associated with the efficacy of ICB (*40*). By partitioning tumors into TMB-high and TMB-low groups (fig. S31A), we found that only the frequency of $CD4^+$ T_{FH}/T_H1 cells showed a strong correlation with TMB (FDR < 0.001, PVE > 26.3%) (figs. S28F and S32A). This association could be validated in pan-cancer and multiple individual cancer types using the bulk TCGA (The Cancer Genome Atlas) data (fig. S32B). We also identified a positive association between *FAT1* (fatty acid translucase 1) mutations and *TNFRSF9*$^+$ T_{reg} cells (Fig. 4B and figs. S31B and S32C). Thus, the T cell composition in the tumor could be affected not only by the number of potential neoantigens, reflected by TMB but also by specific somatic mutations of cancer cells.

To reveal the overall pattern of T cell compositions across cancer types, we inspected the frequency correlations among metaclusters in the tumor and identified several highly correlated modules of metaclusters (fig. S33A). One module consisted of the three ISG^+ metaclusters, whereas the four $CD8^+$ T_{ex} cell populations, T_{FH}/T_H1 cells, and *TNFRSF9*$^+$ T_{reg} cells formed another module. Metaclusters with similar gene signatures but from different compartments ($CD8^+$ Tc, $CD4^+$ T_H, and $CD4^+$ T_{reg} cells) tended to cluster together (for example, $CD8^+$ Tc17 and $CD4^+$ T_H17 cells). The potentially tumor-reactive metaclusters were negatively correlated with certain metaclusters, which could be explained by several mechanisms, including the dynamic state transitions between metaclusters. For example, $CD8^+$ terminal T_{ex} metacluster was negatively correlated with $CD8^+$*ZNF683*$^+$ T_{rm} and $CD8^+$*GZMK*$^+$ T_{em} metaclusters (fig. S33B) and showed an aforementioned state transition relationship. The positively correlated metaclusters usually had significant overlap between signature genes ($P < 0.01$, hypergeometric tests) (Fig. 4C and fig. S34), suggesting that the same regulators induced similar transcriptional programs in different T cell populations.

Using those overlapped signature genes and the NicheNet algorithm (*41*) to find shared ligands, we found that *RORC* and other type-17 response-related genes were potentially induced in both T_H17 and Tc17 by ligands such as *IL23A* and *IL15* (fig. S35A). We identified shared ligands for the potentially tumor-reactive metaclusters. For the ISG^+ metaclusters, the type I interferon-encoding genes were expectedly ranked as the top ligands by the NicheNet analysis (fig. S35B). For terminal T_{ex}, T_{FH}/T_H1, and *TNFRSF9*$^+$ T_{reg} cells, a total of 325 shared signature genes were identified, 18 of which were TFs (for example, *TOX*, *TOX2*, *VDR*, *ZBED2*, *ETV7*, *ZNF282*, and *HIVEP1*) (Fig. 4C and fig. S34). Thus, similar transcriptional machineries were likely used by the three different cell populations to respond to TME stimuli. Further, both terminal T_{ex} and T_{FH}/T_H1 cells produced *CXCL13* and type 1 response cytokines, and *TGFB1* was inferred as one of the top potential ligands inducing their shared signature genes (fig. S35C). Meanwhile, *TNFRSF9*$^+$ T_{reg} cells highly express TGF-β–induced TFs, including *SOX4* and *TGIF2* (table S3). Moreover, *IFNB1* was inferred as one of the top potential ligands inducing the shared signature genes between *TNFRSF9*$^+$ T_{reg} cells and terminal T_{ex} or T_{FH}/T_H1 cells (fig. S35, D and E). These observations suggested that TGF-β and interferons may affect the transcriptional program and abundance of the potentially tumor-reactive T cells.

Immune types of pan-cancer defined by the composition of T cells

Next, using the frequencies of these correlated metaclusters, we found that tumor samples could be clustered into eight groups (C1 to C8) (Fig. 5A). The C1 and C2 harbored high frequencies of terminal T_{ex} cells, and C1 also had the highest frequency of *TNFRSF9*$^+$ T_{reg} cells. Tumors of C3 to C8 harbored a low frequency of terminal T_{ex} cells and high frequency of $CD8^+$*ZNF683*$^+$*CXCR6*$^+$ T_{rm} cells and could be further divided into groups dominated by naïve T cell (C7), enriched naïve T cell (C8), enriched T_{emra} cell (C6), enriched Tc17 or T_H17 cell (C4), and with a low frequency of *TNFRSF9*$^+$ T_{reg} cell (C5), respectively. On the basis of the linear model analysis, we found that our grouping could explain more variabilities of the T cell composition of the tumor than other factors (figs. S28 and S36). Although each immune type included mixed cancer types, certain cancer types exhibited clear preferences (fig. S37). For example, nearly half of esophageal and nasopharyngeal carcinoma tumors were of C1. By contrast, thyroid carcinoma and uterine corpus endometrial carcinoma were enriched in C3, suggesting that a large proportion of these two cancer types with high T cell suppression might still benefit from immunotherapy because of the presence of ISG^+ activating T cells and the low abundance of terminal T_{ex} cells. More than half of melanomas were of C2, which has high T_{ex} cell but lower *TNFRSF9*$^+$ T_{reg} cell frequency, which is consistent with their tendency to respond to ICB. Both basal cell carcinoma and hepatocellular carcinoma were enriched in C4, indicating that their tendencies are inflammatory through IL17-producing T cells. These T cell–based immune types provide a reference to understand the overall tumor-infiltrating T cell properties, which may help guide the development of newer therapies and patient stratification instead of the conventional cancer type metrics.

Such immune type classification may have clinical implications. Using immune type signatures to stratify the TCGA cancer patients, we found that the $T_{ex}^{lo}T_{rm}^{hi}$ (C3 to C8) tumors had better overall survival than that of $T_{ex}^{hi}T_{rm}^{lo}$ tumors (C1 and C2) across cancer types or in multiple individual cancer types, including lung adenocarcinoma, hepatocellular carcinoma, and renal papillary cell carcinoma (Fig. 5B and fig. S38A). Because T cells are the direct target for many immunotherapies, the T cell–based immune types could logically be associated with the treatment efficacies. Reanalysis of published data of PD-1 antibody treatment for melanoma (*42*) indicated that responsive tumors had a lower frequency of terminal T_{ex} cells and a higher frequency of naïve T cells (Fig. 5C and fig. S38B). The T_{ex} cell connection was reproduced in another dataset (*43*), showing that the responder group was enriched with more $T_{ex}^{lo}T_{rm}^{hi}$ tumors than that of the nonresponders (Fisher's exact test, $P = 0.025$) (fig. S38C). Pretreatment tumors in responders also had a higher frequency of Tc17 (Fig. 5C), implicating an important role of Tc17 in ICB treatment. Further investigation is needed to reveal how this finding is tied to the notion that Tc17 could also go into exhaustion in the tumor.

Discussion

We systematically characterized the T cells from various human cancers, investigating different aspects from gene expression signature and heterogeneity to state transitions and regulations. Multiple tumor-enriched metaclusters—including T_{ex}, T_{FH}/T_H1, and *TNFRSF9*$^+$ T_{reg} cells—deserve particular attention because of their potential tumor reactivities. Our analyses revealed diverse paths to T cell exhaustion and the cancer type preference of those paths (fig. S39). Such landscape depiction deepens our understanding of cancer immunity and will facilitate therapeutic development.

The T cell states and infiltration in tumors are affected by multifaceted elements, such as tumor-intrinsic and metabolic factors (*8*). In our data, the TMB shows a positive association with T_{FH}/T_H1 cells, whereas the BMI exhibits a positive association with T_{FH} cells. Because both TMB (*40*) and BMI (*44*) have been previously linked to ICB responses, our findings highlight the importance of T_{FH}-related cell populations in the antitumor response. Additionally, specific mutations could affect T cell compositions in the TME. *FAT1* mutations are positively correlated with *TNFRSF9*$^+$ T_{reg} cell

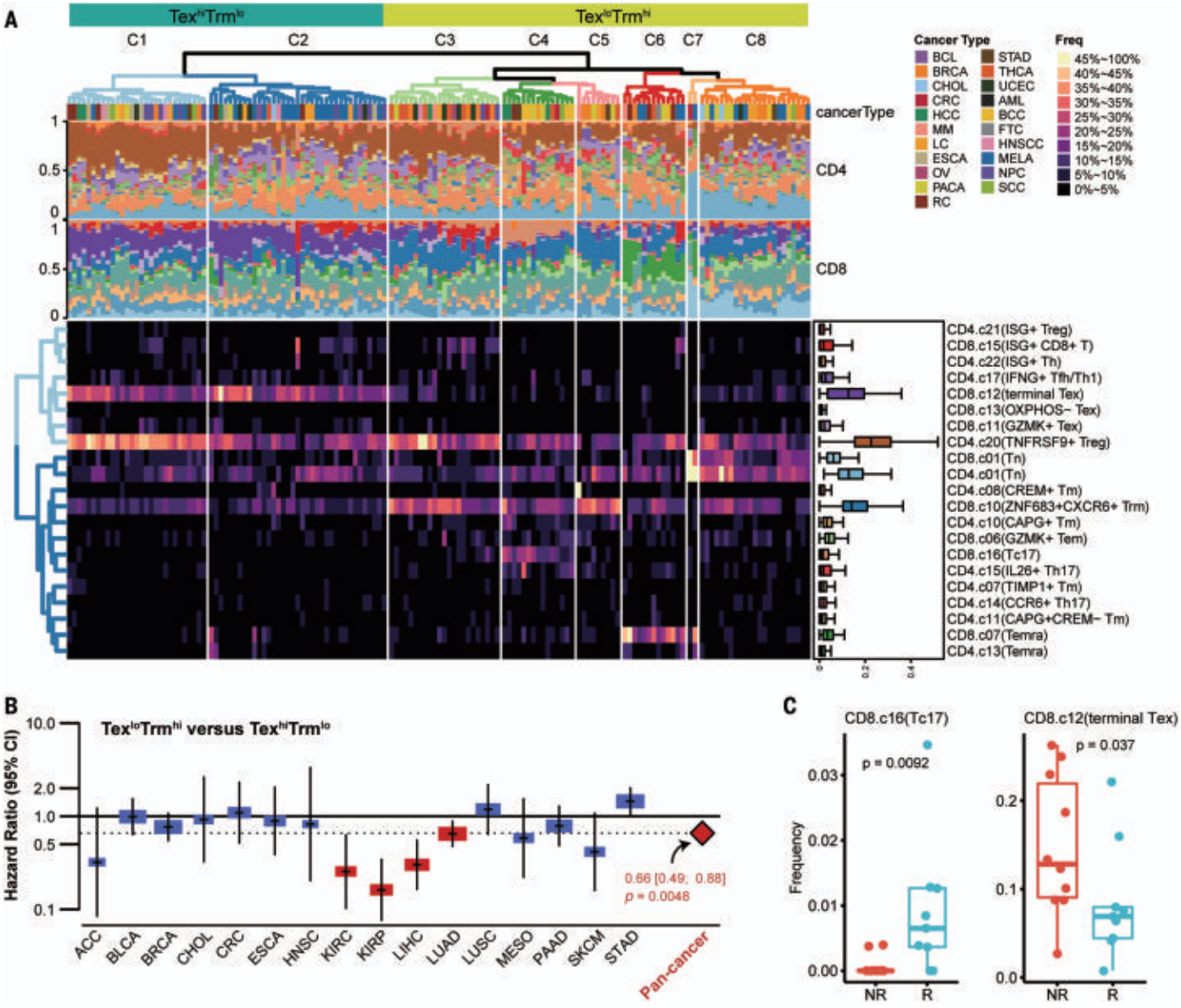


Fig. 5. Immune types of pan-cancer defined by the T cell composition. (**A**) Heatmap reporting the frequencies of metaclusters in the tumor. Only metaclusters that have a correlation of >0.35 with at least one other metacluster are shown. The rows for metaclusters, the columns for tumor samples. For columns, stacked bar plots illustrate the proportion of metaclusters; for rows, a box plot illustrates the distribution of the metacluster proportion across samples. Hierarchical clustering based on Euclidean distance is applied. (**B**) Forest plot reporting the effect of the two major T cell–based immune types on overall survival. The hazard ratios are calculated by using Cox regression models with the age, gender, and stage corrected. The black solid line for hazard ratio 1 (meaning no effect). Red for FDR < 0.05. (**C**) Boxplots comparing the frequencies of metaclusters in melanoma between nonresponders (NR) and responders (R) with antibody-to-PD-1 treatment. Published data from (*42*) is used. The *P* values by Wilcoxon tests are shown.

frequencies. This association might be mediated by the *FAT1*-Hippo-YAP1 (yes-associated protein 1) pathway dysfunction in cancer cells (*45*–*47*). The effect of tumor mutations on infiltrating immune cells is emerging, and mechanisms of such connection will likely be actively pursued in the future. Furthermore, our study demonstrates notable differences of T cell compositions in distinct TMEs of various cancer types and suggests an immune-typing scheme that leverages the overall tumor-infiltrating T cell properties. With further tuning of the single cell–based immune-typing that could faithfully recapitulate the complex tumor-infiltrating T cell properties, we will be better informed when developing future immunotherapies that can be personalized to achieve maximal clinical benefit.

Methods summary

The scRNA-seq data of T cells were collected from both newly sequenced and previously published datasets. For newly generated data, the cancer patients of origin were enrolled, pathologically diagnosed, and surgically biopsied at PKU Cancer Hospital and Institute, with approval of their Research and Ethical Committee. Written informed consents were obtained. The tumors and adjacent noncancer tissues were digested on the basis of gentleMACS and the related kit (Miltenyi, USA). $CD45^+$ living cells were sorted by means of a BD FACSAria III sorter (BD Biosciences, USA)

from single-cell suspensions. The libraries of single-cell transcriptome and single-cell TCR were prepared by means of a 10x Chromium Single-cell 5′ and VDJ library construction kit, then sequenced by means of a Hiseq X Ten sequencer (Illumina, USA).

We applied Cell Ranger (version 3.0) for gene expression quantification, TCR sequence assembly, and cell identification. Scrublet was used to remove potential doublets. Seurat v3 was used to identify T and NK cells. The $CD3^+CD8^+CD4^-$ and $CD3^+CD4^+CD8^-$ T cells were isolated according to computational gating and processed separately in downstream clustering and signature gene analysis.

To integrate heterogeneous data from different sources, a three-step procedure was applied. We first performed per-cell size-factor normalization and per-gene z-score scaling across cells for each dataset. Then, cells within each dataset were partitioned into small groups (miniclusters) to reduce noise. Subsequently, a batch effect correction algorithm, Harmony, was applied to further improve the integration. On the basis of the Harmony result, Seurat was applied to identify clusters, termed metaclusters. We used limma to identify differentially expressed genes among metaclusters. After estimating the moderated effect size of each dataset, the combined effect size was calculated by weighted averaging of the effect sizes. The Gene Set Enrichment Analysis (GSEA) (version 4.0.3) was performed to evaluate the pathway activities of metaclusters.

To characterize the metaclusters, using TCRs as markers, we applied STARTRAC to quantify the magnitude of T cell clonal expansion, migration potential, and state transition potential. A proliferation index, indicating the ongoing proliferation activity of a metacluster, was defined as the frequency of proliferating cells in a metacluster. The OR was used to characterize the tissue distribution of metaclusters.

To model the T cell state transition among metaclusters, we used multiple methodologies, including diffusion map, UMAP, monocle3, and RNA velocity. Specific clonotypes spanning different cell states with high likelihood ratios were also identified, providing direct and intuitive evidence for cell state transitions. We used SCENIC to construct the TF regulatory network. The NicheNet was applied to identify the potential ligands that induced the expression of genes of interest.

The bulk tumor and peripheral blood of patients were subjected to whole-exome sequencing for somatic mutation calling. TMB was calculated and tumors were divided as TMB-high and -low groups by using a cutoff of 10. Patient-matched tumors were also used for RNA-seq, and gene expression quantification was performed following the UCSC Xena Toil RNAseq pipeline.

REFERENCES AND NOTES

1. L. Chen, D. B. Flies, Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* **13**, 227–242 (2013). doi: 10.1038/nri3405; pmid: 23470321
2. H. Li *et al*., Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **181**, 747 (2020). doi: 10.1016/j.cell.2020.04.017; pmid: 32359441
3. X. Guo *et al*., Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018). doi: 10.1038/s41591-018-0045-3; pmid: 29942094
4. A. C. Scott *et al*., TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019). doi: 10.1038/s41586-019-1324-y; pmid: 31207604
5. O. Khan *et al*., TOX transcriptionally and epigenetically programs $CD8^+$ T cell exhaustion. *Nature* **571**, 211–218 (2019). doi: 10.1038/s41586-019-1325-x; pmid: 31207603
6. B. C. Miller *et al*., Subsets of exhausted $CD8^+$ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019). doi: 10.1038/s41590-019-0312-6; pmid: 30778252
7. D. R. Sen *et al*., The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016). doi: 10.1126/science.aae0491; pmid: 27789799
8. D. S. Thommen, T. N. Schumacher, T cell dysfunction in cancer. *Cancer Cell* **33**, 547–562 (2018). doi: 10.1016/j.ccell.2018.03.012; pmid: 29634943
9. L. Zhang *et al*., Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018). doi: 10.1038/s41586-018-0694-x; pmid: 30479382
10. O. Zavidij *et al*., Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nat. Cancer* **1**, 493–506 (2020). doi: 10.1038/s43018-020-0053-3; pmid: 33409501
11. T. D. Wu *et al*., Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* **579**, 274–278 (2020). doi: 10.1038/s41586-020-2056-8; pmid: 32103181
12. J. Qian *et al*., A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020). doi: 10.1038/s41422-020-0355-0; pmid: 32561858
13. I. Korsunsky *et al*., Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi: 10.1038/s41592-019-0619-0; pmid: 31740819
14. Materials and methods are available as supplementary materials.
15. G. X. Y. Zheng *et al*., Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: 10.1038/ncomms14049; pmid: 28091601
16. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014). doi: 10.1038/nprot.2014.006; pmid: 24385147
17. M. Dusseaux *et al*., Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17–secreting T cells. *Blood* **117**, 1250–1259 ((2011)). doi: 10.1182/blood-2010-08-303339; pmid: 21084709
18. M. St Paul, P. S. Ohashi, The roles of $CD8^+$ T cell subsets in antitumor immunity. *Trends Cell Biol.* **30**, 695–704 (2020). doi: 10.1016/j.tcb.2020.06.003; pmid: 32624246
19. C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948). doi: 10.1002/j.1538-7305.1948.tb01338.x
20. A. M. van der Leun, D. S. Thommen, T. N. Schumacher, $CD8^+$ T cell states in human cancer: Insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020). doi: 10.1038/s41568-019-0235-4; pmid: 32024970
21. M. H. Spitzer *et al*., Systemic immunity is required for effective cancer immunotherapy. *Cell* **168**, 487–502.e15 (2017). doi: 10.1016/j.cell.2016.12.022; pmid: 28111070
22. P. K. Gupta *et al*., CD39 expression identifies terminally exhausted $CD8^+$ T cells. *PLOS Pathog.* **11**, e1005177 (2015). doi: 10.1371/journal.ppat.1005177; pmid: 26485519
23. Z. Chen *et al*., TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. *Immunity* **51**, 840–855.e5 (2019). doi: 10.1016/j.immuni.2019.09.013; pmid: 31606264
24. I. Siddiqui *et al*., Intratumoral $Tcf1^+PD\text{-}1^+CD8^+$ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity* **50**, 195–211.e10 (2019). doi: 10.1016/j.immuni.2018.12.021; pmid: 30635237
25. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015). doi: 10.1093/bioinformatics/btv325; pmid: 26002886
26. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi: 10.1038/s41587-020-0591-3; pmid: 32747759
27. J. Cao *et al*., The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019). doi: 10.1038/s41586-019-0969-x; pmid: 30787437
28. E. Becht *et al*., Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018). doi: 10.1038/nbt.4314; pmid: 30531897
29. P. A. Szabo *et al*., Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019). doi: 10.1038/s41467-019-12464-3; pmid: 31624246
30. E. Stelekati *et al*., Bystander chronic infection negatively impacts development of $CD8^+$ T cell memory. *Immunity* **40**, 801–813 (2014). doi: 10.1016/j.immuni.2014.04.010; pmid: 24837104
31. T. Wu *et al*., The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci. Immunol.* **1**, eaai8593 (2016). doi: 10.1126/sciimmunol.aai8593; pmid: 28018990
32. S. J. Im *et al*., Defining $CD8^+$ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016). doi: 10.1038/nature19330; pmid: 27501248
33. D. T. Utzschneider *et al*., T cell factor 1-expressing memory-like CD8(+) T cells sustain the immune response to chronic viral infections. *Immunity* **45**, 415–427 (2016). doi: 10.1016/j.immuni.2016.07.021; pmid: 27533016
34. V. Larochette *et al*., IL-26, a cytokine with roles in extracellular DNA-induced inflammation and microbial defense. *Front. Immunol.* **10**, 204 (2019). doi: 10.3389/fimmu.2019.00204; pmid: 30809226
35. S. Aibar *et al*., SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017). doi: 10.1038/nmeth.4463; pmid: 28991892
36. A. T. Satpathy *et al*., Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019). doi: 10.1038/s41587-019-0206-z; pmid: 31375813
37. H. Yoshitomi *et al*., Human Sox4 facilitates the development of CXCL13-producing helper T cells in inflammatory environments. *Nat. Commun.* **9**, 3762 (2018). doi: 10.1038/s41467-018-06187-0; pmid: 30232328
38. M. C. Gerner *et al*., The TGF-b/SOX4 axis and ROS-driven autophagy co-mediate CD39 expression in regulatory T-cells. *FASEB J.* **34**, 8367–8384 (2020). doi: 10.1096/fj.201902664; pmid: 32319705
39. H. Cheroutre, M. M. Husain, CD4 CTL: Living up to the challenge. *Semin. Immunol.* **25**, 273–281 (2013). doi: 10.1016/j.smim.2013.10.022; pmid: 24246226
40. J. J. Havel, D. Chowell, T. A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019). doi: 10.1038/s41568-019-0116-x; pmid: 30755690
41. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020). doi: 10.1038/s41592-019-0667-5; pmid: 31819264
42. M. Sade-Feldman *et al*., Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **176**, 404 (2019). doi: 10.1016/j.cell.2018.12.034; pmid: 30633907
43. W. Hugo *et al*., Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **168**, 542 (2017). doi: 10.1016/j.cell.2017.01.010; pmid: 28129544
44. Z. Wang *et al*., Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. *Nat. Med.* **25**, 141–151 (2019). doi: 10.1038/s41591-018-0221-5; pmid: 30420753
45. D. Martin *et al*., Assembly and activation of the Hippo signalome by FAT1 tumor suppressor. *Nat. Commun.* **9**, 2372 (2018). doi: 10.1038/s41467-018-04590-1; pmid: 29985391
46. M. Shibata, K. Ham, M. O. Hoque, A time for YAP1: Tumorigenesis, immunosuppression and targeted therapy. *Int. J. Cancer* **143**, 2133–2144 (2018). doi: 10.1002/ijc.31561; pmid: 29696628
47. L. M. Francisco *et al*., PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J. Exp. Med.* **206**, 3015–3029 (2009). doi: 10.1084/jem.20090847; pmid: 20008522
48. L. Zheng, S. Qin, Codes for the paper "Pan-cancer single-cell landscape of tumor-infiltrating T cells." Zenodo (2021). doi: 10.5281/zenodo.5461803

ACKNOWLEDGMENTS

We thank Y. Yang, B. Zhang, Y. Gao, Y. Li, M. Gao, K. Wang, D. Lin, L. Zhou, and L. Jiang for clinical sample collection and F. Wang and X. Zhang for assistance with fluorescence-activated cell sorting (FACS). We thank the Computing Platform of the Center for Life Science and the National Center for Protein Sciences Beijing (Peking University). **Funding:** This project was supported by the Beijing Advanced Innovation Centre for Genomics at Peking University, the Beijing Municipal Science and Technology Commission (Z201100005320014), the National Natural Science Foundation of China (81988101, 91959000, 91942307, and 31991171), the Beijing Natural Science Foundation (7182075), Capital's Funds for Health Improvement and Research (2020-4-40916), and the Beijing Natural Science Foundation (7204327). W.S. was supported in part by the Postdoctoral Fellowship of Peking-Tsinghua Center for Life Sciences. **Author contributions:** Conceptualization: Z.Z., J.J., X.H., Z.B., and L.Z. Resources: A.W., X.W., J.Z., B.X., N.W., N.Z., H.Z., H.O., K.C., Z.B., and J.J. Methodology: Z.Z., X.H., L.Z., S.Q., W.S., and X.R. Investigation: R.G. and L.W. Formal analysis, L.Z. and S.Q. Writing, original draft: L.Z., S.Q., W.S., X.H., and Z.Z. Writing, review and editing: L.Z., S.Q., W.S., X.H., and Z.Z. **Competing interests:** Z.Z. is a founder of Analytical BioSciences and is a consultant for InnoCare Pharma and ArsenalBio. **Data and materials availability:** The data presented in this manuscript are tabulated in the supplementary materials. Sequencing data are available at Genome Sequence Archive (accession no. PRJCA001702), and processed gene expression data are deposited in Gene Expression Omnibus (accession no. GSE156728) and can be accessed through an online data browser (http://cancer-pku.cn:3838/PanC_T). All codes used for analysis are available from the Zenodo repository (*48*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abe6474
Materials and Methods
Figs. S1 to S39
References (*49*–*65*)
MDAR Reproducibility Checklist
Tables S1 to S4

4 September 2020; resubmitted 17 May 2021
Accepted 26 October 2021
10.1126/science.abe6474

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Science Advances
AAAS

GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant

Paul Elliott*, David Haw†, Haowei Wang†, Oliver Eales†, Caroline E. Walters, Kylie E. C. Ainslie, Christina Atchison, Claudio Fronterre, Peter J. Diggle, Andrew J. Page, Alexander J. Trotter, Sophie J. Prosolek, The COVID-19 Genomics UK (COG-UK) Consortium, Deborah Ashby, Christl A. Donnelly, Wendy Barclay, Graham Taylor, Graham Cooke, Helen Ward, Ara Darzi, Steven Riley*

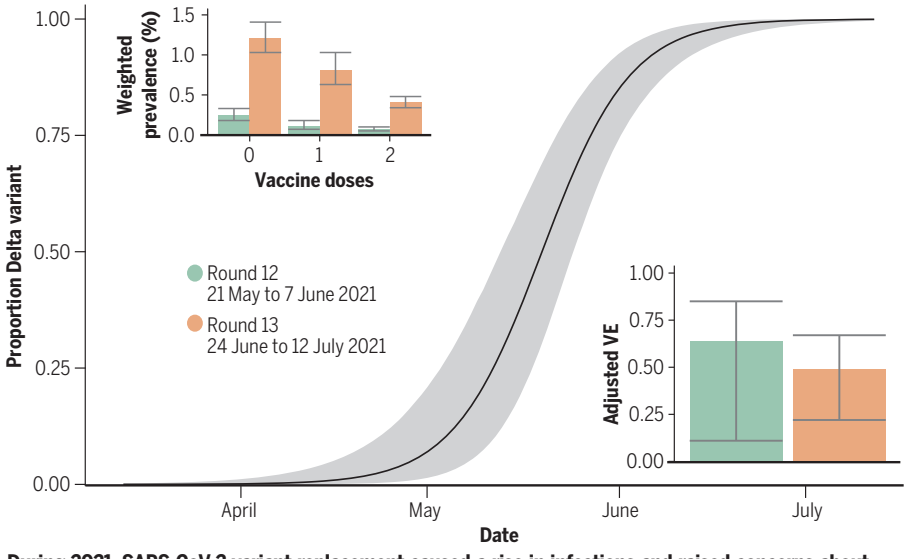
BACKGROUND: The prevalence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection continues to drive rates of illness and hospitalizations despite high levels of vaccination, with the proportion of cases caused by the Delta lineage increasing in many populations. As vaccination programs roll out globally and social distancing is relaxed, future SARS-CoV-2 trends are uncertain.

METHODS: The Real-time Assessment of Community Transmission–1 (REACT-1) study has been tracking the spread of the COVID-19 pandemic in England since May 2020. The study involves obtaining a self-administered throat and nose swab for reverse transcription polymerase chain reaction (RT-PCR) from ~100,000 or more people during 2 to 3 weeks each month, based on random samples of the population in England at ages 5 years and above. As well as information on swab positivity, we collect demographic and other data on potential risk factors and (since January 2021) vaccination history. Prevalence estimates are weighted to be representative of the population of England as a whole. Here, we analyzed prevalence trends and their drivers using RT-PCR swab positivity data from REACT-1 round 12 (between 20 May and 7 June 2021) and round 13 (between 24 June and 12 July 2021). Response rates, defined as the percentage of invitees from whom we received a valid swab result, were 20.4% across all rounds and 13.4% and 11.7% for rounds 12 and 13, respectively.

RESULTS: We observed sustained exponential growth as the third wave in England took hold, with reproduction number R estimated at 1.44 (95% credible interval 1.20, 1.73) in round 12 and 1.19 (1.06, 1.32) in round 13, corresponding to an average doubling time of 11 days (7, 23 days) in round 12 and 25 days (15, >50 days) in round 13. This resulted in an increase in average weighted prevalence from 0.15% (0.12%, 0.18%) in round 12 (based on 135 positives out of 108,911 valid swabs) to 0.63% (0.57%, 0.69%) in round 13 (527 positives out of 98,233). The rapid growth across and within rounds appears to have been driven by complete replacement of the Alpha variant by Delta, and by the high prevalence in younger, less-vaccinated age groups: Among those aged 13 to 17 years, we observed an increase in weighted prevalence by a factor of 9 between round 12 [0.16% (0.08%, 0.31%)] and round 13 [1.56% (1.25%, 1.95%)]. In round 13, weighted prevalence among those who reported being unvaccinated [1.21% (1.03%, 1.41%)] was greater than for those who reported having had two doses of vaccine [0.40% (0.34%, 0.48%)] by a factor of 3; however, 44% of infections occurred in doubly vaccinated individuals, reflecting imperfect vaccine effectiveness (VE) against infection after two doses despite high overall levels of vaccination.

Among participants aged 18 to 64 years, on the basis of self-reported vaccination status, we estimated VE against infection (adjusted for age, sex, region, ethnicity, and index of multiple deprivation) of 49% (95% confidence interval 22%, 67%) in round 13, rising to 58% (33%, 73%) when only strong positives [cycle threshold (Ct) values below 27] were considered. For the same age group, we estimated adjusted VE of 59% (23%, 78%) against symptomatic infection—that is, among those reporting one or more common COVID-19 symptoms in the month prior to testing (fever, loss or change of sense of smell or taste, new persistent cough). Ethnicity, household size, and local levels of deprivation, in addition to age, jointly contributed to the risk of higher prevalence of swab positivity.

CONCLUSION: From the end of May to the beginning of July 2021 in England, where there was a highly successful vaccination campaign with high vaccine uptake, infections were increasing exponentially—driven by the Delta variant—with high infection prevalence among younger, unvaccinated individuals. Despite slower growth (or level or declining prevalence) during summer 2021 in the Northern Hemisphere, increased mixing in the presence of the Delta variant likely explains renewed growth that occurred in autumn 2021, even in populations with high levels of vaccination. ■



During 2021, SARS-CoV-2 variant replacement caused a rise in infections and raised concerns about vaccine effectiveness (VE) against infection. Main and top left: Complete replacement of Alpha by the Delta variant from REACT-1 round 12 to round 13 and weighted prevalence of SARS-CoV-2 infection among a random sample of the population of England ages 5 years and above by self-reported vaccine status. Bottom right: VE adjusted for age, sex, index of multiple deprivation, region, and ethnicity.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: p.elliott@imperial.ac.uk (P.E.); s.riley@imperial.ac.uk (S.R.)
†These authors contributed equally to this work.
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as P. Elliott *et al.*, *Science* **374**, eabl9551 (2021). DOI: 10.1126/science.abl9551

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl9551

RESEARCH ARTICLE

CORONAVIRUS

Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant

Paul Elliott[1,2,3,4,5,6]*, David Haw[1,7]†, Haowei Wang[1,7]†, Oliver Eales[1,7]†, Caroline E. Walters[1,7], Kylie E. C. Ainslie[1,7,8], Christina Atchison[1], Claudio Fronterre[9], Peter J. Diggle[9], Andrew J. Page[10], Alexander J. Trotter[10], Sophie J. Prosolek[10], The COVID-19 Genomics UK (COG-UK) Consortium[11]‡, Deborah Ashby[1], Christl A. Donnelly[1,7,12], Wendy Barclay[13], Graham Taylor[13], Graham Cooke[2,3,13], Helen Ward[1,2,3], Ara Darzi[2,3,14,15], Steven Riley[1,7]*

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections were rising during early summer 2021 in many countries as a result of the Delta variant. We assessed reverse transcription polymerase chain reaction swab positivity in the Real-time Assessment of Community Transmission–1 (REACT-1) study in England. During June and July 2021, we observed sustained exponential growth with an average doubling time of 25 days, driven by complete replacement of the Alpha variant by Delta and by high prevalence at younger, less-vaccinated ages. Prevalence among unvaccinated people [1.21% (95% credible interval 1.03%, 1.41%)] was three times that among double-vaccinated people [0.40% (95% credible interval 0.34%, 0.48%)]. However, after adjusting for age and other variables, vaccine effectiveness for double-vaccinated people was estimated at between ~50% and ~60% during this period in England. Increased social mixing in the presence of Delta had the potential to generate sustained growth in infections, even at high levels of vaccination.

Despite the successful development, licensing, and distribution of effective vaccines against COVID-19 (*1*, *2*), the number of newly reported cases and deaths continued to rise globally into the Northern Hemisphere summer of 2021 (*3*). Prior trends of decreasing prevalence were being reversed in some populations where the Delta variant had become dominant, leading to estimates of a substantially higher transmissibility for Delta relative to Alpha (*4*). In addition, globally, as of July 2021, only 13% of the population were double-vaccinated and only 1% of people in low-income countries had received even one dose (*5*). Despite slower growth (or level or declining prevalence) during the Northern Hemisphere summer, many countries experienced a further large wave of infections in the autumn, driven by the Delta variant.

The vaccine rollout in England started with the oldest and most vulnerable groups, beginning in December 2020. Since then, there has been a strong correlation among age, vaccine type, and date of vaccination, with individuals receiving the same vaccine for first and second dose. Initially, health care workers and older adults received BNT162b2 before doses were switched to ChAdOx1 for many people between the ages of 40 and 80 and some younger people. The program then switched back to BNT162b2 for those below the age of 40 (also using small numbers of mRNA-1273 vaccine). Subsequently, from September 2021, the vaccination program was expanded to include children from the age of 12 years.

The incidence of reverse transcription polymerase chain reaction (RT-PCR)–confirmed cases of COVID-19 increased substantially in England after the Delta variant became established during April and May 2021 (*6*). Over the same period, the UK government proceeded with its gradual relaxation of social distancing (roadmap out of lockdown) (*7*) and the ending of almost all legal restrictions in England on 19 July 2021 (*8*). Although a much lower proportion of COVID-19 cases resulted in hospitalizations in England versus a comparable period of growth during autumn 2020, exponential growth in hospitalizations was still observed from mid-June 2021 (*6*).

With first data collection starting in May 2020, we established the Real-time Assessment of Community Transmission–1 (REACT-1) study to track the spread of the COVID-19 pandemic in England and improve situational awareness (*9*, *10*). The study involves obtaining a self-administered throat and nose swab for RT-PCR from ~100,000 or more people during 2 to 3 weeks each month, based on random samples of the population in England at ages 5 years and above (see materials and methods). As well as information on swab positivity, we collect demographic and contextual data including (since January 2021) on vaccination history. By July 2021, ~1.9 million people had taken part (table S1). Here, we describe the key patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections for round 12 (20 May to 7 June 2021) and round 13 (24 June to 12 July 2021) during the third wave of the epidemic in England. Valid RT-PCR results were obtained from 108,911 participants in round 12 and 98,233 participants in round 13 (table S1).

Prevalence and growth

Prevalence of infection with SARS-CoV-2 increased substantially in England between rounds 12 and 13 (Fig. 1) as the third wave took hold, linked to the rapid replacement of Alpha by the Delta variant. In round 13, between 24 June and 12 July 2021, we found 527 positives from 98,233 swabs, giving a weighted prevalence of 0.63% [95% credible interval (CrI) 0.57%, 0.69%], and, on average, a factor of >4 rise relative to the weighted prevalence in round 12 of 0.15% (CrI 0.12%, 0.18%) (table S1). The prevalence in round 13 was similar to that observed in early October 2020 and late January 2021 during, respectively, the rise and fall of the second wave (Fig. 1).

The Delta variant completely replaced Alpha during the period of our study, consistent with genomic data from outbreak investigation and routine surveillance (*11*). Of the 254 lineages determined for round 13, 100% were the Delta variant, compared with round 12 during which 36 of 46 (78.3%) were Delta and the remaining 10 were the Alpha variant. The growth of Delta against Alpha from round 10 (11 to 30 March 2021) to round 13 corresponded to a daily growth rate advantage of 0.14 (CrI 0.10, 0.20) for Delta, which, in turn, implied an additive R advantage of 0.86 (CrI 0.63, 1.23) (Fig. 1). This is consistent with estimates based on trends in the proportion of positive PCR assays where the S gene was not detected [presumed to be Alpha (*12*)] and on differences in household attack rate for households where Delta was identified rather than Alpha (*13*). Within the Delta variant, we did not detect the K417N mutation associated with the AY.1 and AY.2 lineages. Under the assumption that REACT-1 participants provide an unbiased

[1]School of Public Health, Imperial College London, London, UK. [2]Imperial College Healthcare NHS Trust, London, UK. [3]National Institute for Health Research Imperial Biomedical Research Centre, London, UK. [4]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [5]Health Data Research UK London at Imperial College London, London, UK. [6]UK Dementia Research Institute Centre at Imperial, London, UK. [7]MRC Centre for Global Infectious Disease Analysis and Jameel Institute, Imperial College London, London, UK. [8]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, Netherlands. [9]CHICAS, Lancaster Medical School, Lancaster University, and Health Data Research UK, Lancaster, UK. [10]Quadram Institute, Norwich, UK. [11]www.cogconsortium.uk. [12]Department of Statistics, University of Oxford, Oxford, UK. [13]Department of Infectious Disease, Imperial College London, London, UK. [14]Institute of Global Health Innovation at Imperial College London, London, UK. [15]Health Security Initiative, Flagship Pioneering UK Ltd., Bristol, UK.
***Corresponding author. Email: p.elliott@imperial.ac.uk (P.E.); s.riley@imperial.ac.uk (S.R.)**
†These authors contributed equally to this work.
‡The full list of consortium members and affiliations is provided in the supplementary materials.

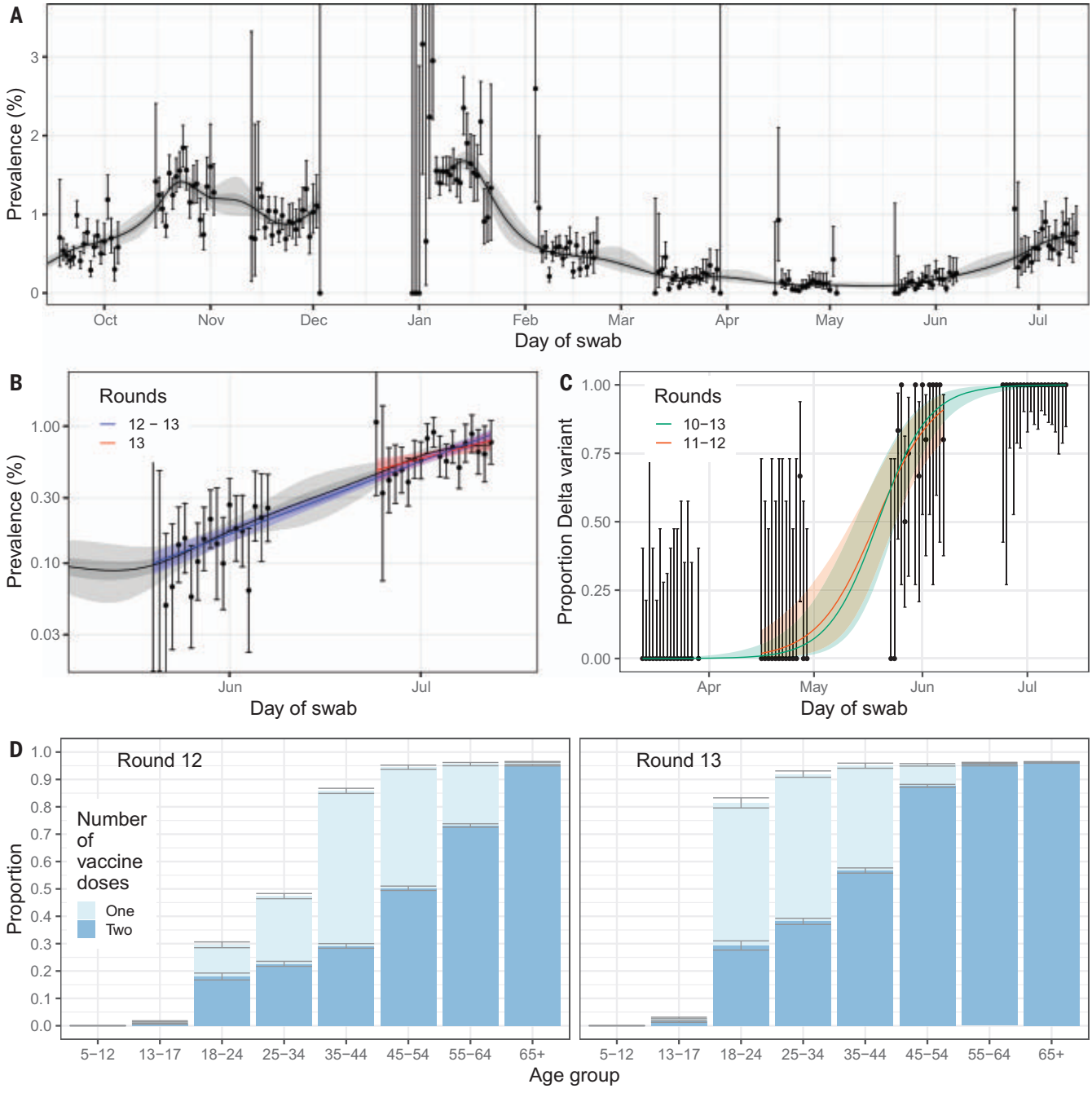


Fig. 1. Temporal trends in prevalence, proportion of positive cases determined to be the Delta variant, and vaccine coverage. (**A**) Prevalence of national swab positivity for England estimated using a P-spline for all 13 rounds with central 50% (dark gray) and 95% (light gray) posterior credible intervals. From round 5 of the study onward, weighted observations (black dots) and 95% binomial confidence intervals (vertical lines) are also shown. Note that the period between rounds 7 and 8 (December) of the model is not included, as there were no data available to capture the late December peak of the epidemic. (**B**) Comparison of the exponential model fit to round 12 and 13 (blue) and the exponential model fit to round 13 only (red). Also shown is the P-spline model fit from (A). Shown here only for rounds 12 and 13 of the study with a $\log_{10}$ y axis. (**C**) Proportion of Delta against Alpha over time. Points show raw data; error bars denote the 95% confidence interval. Shaded regions show best-fit Bayesian logistic regression models, fit to rounds 10 to 13 (green) and rounds 11 and 12 (orange), with 95% credible interval. (**D**) Proportion of individuals with known vaccine status who reported being vaccinated with one (light blue) or two (dark blue) doses. Error bars denote 95% binomial confidence intervals.

sample of infections, we can exclude, with 95% confidence, a population prevalence of non-Delta lineages greater than 0.004%, corresponding to 2350 infections in England on average during round 13.

Nationally, we observed an exponential trend in prevalence with sustained growth for rounds 12 and 13 (between 20 May and 12 July 2021) (Fig. 1 and table S2) despite England having one of the highest adult vaccination rates internationally (*5*). Averaging over the period of each of rounds 12 and 13 separately, we estimated the reproduction number R at 1.44 (CrI 1.20, 1.73) (round 12) and 1.19 (CrI 1.06, 1.32) (round 13), corresponding

to doubling times of 11 days (CrI 7, 23 days) and 25 days (15, >50 days) respectively. Across rounds 12 and 13, R was 1.28 (CrI 1.24, 1.31) with a doubling time of 17 days (CrI 15, 19 days). Patterns of growth for the period of the study were robust when considering alternative definitions of positivity, such as only nonsymptomatic individuals or positive samples with lower cycle threshold (Ct) values, corresponding to higher viral load (table S2).

Age

Alongside the rapid rise of the Delta variant, recent growth in England appears to have been driven by younger age groups (table S3 and fig. S1). For example, in 13- to 17-year-olds, weighted prevalence in round 13 [1.56% (CrI 1.25%, 1.95%)] was higher than in round 12 [0.16% (CrI 0.08%, 0.31%)] by a factor of 9. Similar patterns were observed in England for the same period in a longitudinal household study (*14*). In contrast, at ages 65 to 74 years, weighted prevalence increased from round 12 [0.07% (CrI 0.04%, 0.12%)] to round 13 [0.25% (CrI 0.19%, 0.34%)] by a factor of 3 to 4. More generally, participants aged between 5 and 24 years were overrepresented among infected people in our study, contributing 50% of infections (weighted age-standardized) while only representing 25% of the population of England aged 5 years or above (*15*). Therefore, whether because of mixing patterns, infectiousness or susceptibility, this group was driving transmission and, during a period of exponential growth, any vaccination targeted at the younger ages would have a disproportionate impact in slowing the epidemic (*16*).

Prevalence among vaccinated and unvaccinated

Participants who reported having received two doses of vaccine were at substantially reduced risk of testing positive relative to those who reported not being vaccinated. For round 13, the prevalence of swab positivity among unvaccinated participants [1.21% (CrI 1.03%, 1.41%)] was greater for all ages than among those who had received two doses of vaccine [0.40% (CrI 0.34%, 0.48%)] by a factor of 3 (table S3). The prevalence in unvaccinated relative to double-vaccinated individuals was similar for round 12, with a prevalence of 0.24% (CrI 0.18%, 0.33%) in those unvaccinated versus 0.07% (CrI 0.05%, 0.10%) in those reporting two doses (table S3).

However, these estimates conflate the effect of vaccination with other correlated variables such as age, which is strongly associated with likelihood of having been vaccinated and also acts as a proxy for differences in behavior across the age groups. Specifically, in England, few children and young people under the age of 18 years have been vaccinated twice, whereas few over the age of 65 years remain unvaccinated (Table 1 and Fig. 1). We therefore restricted the analyses to those aged 18 to 64 years (n = 64,415 in round 12, n = 57,457 in round 13), which permitted direct contrast of infection rates between double-vaccinated and unvaccinated groups (Table 1).

At these ages, we compared swab-negatives with (i) all swab-positives and (ii) the subset of swab-positives who were symptomatic [i.e., reporting one or more common COVID-19 symptoms in the month prior to testing (fever, loss or change of sense of smell or taste, new persistent cough)]. After adjusting for age, sex, region, ethnicity, and index of multiple deprivation (IMD) (*17*), for all swab-positives, we estimated vaccine effectiveness (VE) of 64% [95% confidence interval (CI) 11%, 85%] in round 12 and 49% (CI 22%, 67%) in round 13 among people who had received two doses of vaccine of any type. For those with symptoms, we estimated VE of 83% (CI 19%, 97%) in round 12 and 59% (CI 23%, 78%) in round 13 (Table 2).

Independent data on vaccination status was provided for 57,338 (89%) participants aged 18 to 64 in round 12 consenting to data linkage, and 49,923 (87%) in round 13 (materials and methods). Using these linked data, we estimated adjusted VE at 75% (CI 35%, 90%) in round 12 and 62% (CI 38%, 77%) in round 13. The apparently higher VE for the linked participants reflected differences in odds of infection among the linked and unlinked groups (table S4), suggesting possible bias introduced by consent to linkage, but also some misclassification of vaccine status in the self-reported data (table S5). Because reported dates of vaccination were more reliable in the linked data, we used those data to examine the effect of including a lag period of 14 days after the second vaccination and observed similar odds ratios for zero lag and 14-day lag following the second dose (Table 1). In addition, we observed a similar unweighted prevalence of swab positivity among double-vaccinated individuals who did and did not report prior infection more than 28 days before their swab (table S5), which suggests in our study that prior infection did not materially affect the estimate of VE. Moreover, the strong correlation among age, vaccine type, and time since vaccination in England, together with limited numbers, prevented us from being able to reliably assess the impact of vaccine type or time since infection independently of age.

Although vaccination was associated with lower prevalence of swab positivity, there remained potential for large numbers of people who had received two doses of vaccine to become infected. During the period of round 12, we extrapolated from our data that 29% of infections in England occurred in double-vaccinated people, rising to 44% during the period of round 13. These increases in prevalence in vaccinated individuals in round 13 could be driven by increased social mixing or by a higher proportion of infections being the Delta variant, or attributable to waning of protection from infection. Also, although lower than for unvaccinated individuals, nearly one in 25 double-vaccinated individuals [3.84% (CI 2.81%, 5.21%)] tested swab-positive if they reported contact with a known COVID-19 case (table S6).

Cycle threshold values

We analyzed Ct values associated with positive results among vaccinated and unvaccinated individuals as a measure of viral load. For all positives in round 13, at ages 18 to 64 years, median Ct value for vaccinated participants [27.6 (CI for median, 25.5, 29.7)] was higher than for unvaccinated ones [23.1 (CI 20.3, 25.8) (positive defined as N gene Ct below 37 or both N gene– and E gene–detected; see materials and methods) (Fig. 2 and table S7). The higher Ct values among vaccinated people may suggest lower infectiousness (*18*), consistent with transmission studies conducted when the Alpha variant was dominant, in which vaccinated individuals were at substantially lower risk of passing on infection (*19*). As a secondary analysis, we reduced the Ct threshold for positivity to capture strong positives, which resulted in a smaller difference in median Ct values between vaccinated and unvaccinated individuals (Fig. 2, C and D). At the same time, our estimate of VE for those who reported having received two doses of vaccine increased to 54% (CI 29%, 71%) for a Ct threshold of 35, plateauing between 57% (CI 32%, 72%) and 58% (CI 33%, 73%) for a Ct threshold of 33 and 27, respectively.

Time series of infections, hospital admissions, and deaths

We next investigated how swab positivity measured in REACT-1 related to daily hospital admissions and deaths in publicly available data (*6*), finding a best-fitting lag between swab positivity and hospitalizations of 20 days and between swab positivity and deaths of 26 days (Fig. 3). At these lags, from early February 2021, there was a clear divergence between swab positivity and deaths, coinciding with the rollout of England's mass vaccination campaign, with a smaller divergence between swab positivity and hospitalizations. However, as the Delta variant became dominant in mid-April 2021, the associations between infections and hospitalizations and deaths began to reconverge, both for people below and above 65 years (fig. S2).

Geographical variation

At the regional level, estimates of R were consistent with the overall trend within round 13.

Table 1. Self-reported and linked vaccination status and swab positivity in rounds 12 and 13 of REACT-1 shown for all participants (5 years and above) and for the subset aged 18 to 64 years.

Dataset	Age group	Vaccine status	Round 12			Round 13		
			Negative	Positive	Odds ratio	Negative	Positive	Odds ratio
Self-reported	All	Unvaccinated	22,709	51	Reference	14,957	178	Reference
		Vaccinated (1 dose)	18,654	20	0.48 (0.28, 0.80)	9,598	77	0.67 (0.52, 0.88)
		Vaccinated (2 or more doses)	48,383	30	0.28 (0.18, 0.43)	55,765	197	0.30 (0.24, 0.36)
		Vaccinated (unknown doses)	2,889	1	0.15 (0.02, 1.12)	3,314	11	0.28 (0.15, 0.51)
		Vaccine status not known	16,141	33	0.91 (0.59, 1.41)	14,072	64	0.38 (0.29, 0.51)
	18–64	Unvaccinated	9,012	16	Reference	2,574	28	Reference
		Vaccinated (1 dose)	18,307	19	0.58 (0.30, 1.14)	9,467	76	0.74 (0.48, 1.14)
		Vaccinated (2 or more doses)	25,248	17	0.38 (0.19, 0.75)	34,503	145	0.39 (0.26, 0.58)
		Vaccinated (unknown doses)	1,173	0	0.00 (0.00, NA)	1,517	9	0.55 (0.26, 1.16)
		Vaccine status not known	10,597	26	1.38 (0.74, 2.58)	9,089	49	0.50 (0.31, 0.79)
Linked	All	Unvaccinated	19,115	52	Reference	11,357	153	Reference
		Vaccinated (1 dose)	26,285	33	0.46 (0.30, 0.71)	11,885	93	0.58 (0.45, 0.75)
		Vaccinated (2 or more doses)	50,721	34	0.25 (0.16, 0.38)	61,202	206	0.25 (0.20, 0.31)
	18–64	Unvaccinated	8,099	21	Reference	1,553	25	Reference
		Vaccinated (1 dose)	25,657	32	0.48 (0.28, 0.83)	11,652	92	0.49 (0.31, 0.77)
		Vaccinated (2 or more doses)	23,511	18	0.30 (0.16, 0.55)	36,448	153	0.26 (0.17, 0.40)
	All	Unvaccinated	19,115	52	Reference	11,357	153	Reference
		Vaccinated (<14 days 2nd dose)	31,826	35	0.40 (0.26, 0.62)	13,425	102	0.56 (0.44, 0.73)
		Vaccinated (≥14 days 2nd dose)	45,180	32	0.26 (0.17, 0.40)	59,662	197	0.25 (0.20, 0.30)
	18–64	Unvaccinated	8,099	21	Reference	1,553	25	Reference
		Vaccinated (<14 days 2nd dose)	30,593	34	0.43 (0.25, 0.74)	13,170	101	0.48 (0.31, 0.74)
		Vaccinated (≥14 days 2nd dose)	18,575	16	0.33 (0.17, 0.64)	34,930	144	0.26 (0.17, 0.39)

Prevalence in round 13 was highest in London at 0.94% (CrI 0.76%, 1.16%), up from 0.13% (CrI 0.08%, 0.20%) in round 12 (table S3). There was a suggestion of a possible slowing of the rise in London in the most recent data, although with wide confidence intervals (table S8).

At the subregional level, there was a suggestion of prevalence of infection decreasing in some areas and increasing in others (fig. S3). For example, in the North West of England, high prevalence in a large urban area covering Greater Manchester and Lancashire during the first half of round 13 was less evident in the second half, whereas prevalence increased between the first and second halves in nearby south Yorkshire, part of the Yorkshire and The Humber region. These data are indicative of rapidly changing local spread of the virus within the context of the national exponential rise in infections.

Ethnicity, household size, and neighborhood deprivation

Ethnicity, household size, and area levels of deprivation jointly contributed to the risk of higher prevalence of swab positivity, in addition to age. Unadjusted prevalence (table S3) showed highest prevalence in people of Black ethnicity at 1.21% (CrI 0.75%, 1.93%) compared with 0.59% (CrI 0.53%, 0,65%) in people of white ethnicity; highest prevalence in those in the largest households of six or more people at 1.35% (CrI 0.90%, 2.01%) compared with 0.44% (CrI 0.32%, 0.61%) and 0.44% (CrI 0.36%, 0.53%) in single- and two-person households, respectively; and highest prevalence in participants living in the most deprived neighborhoods at 0.82% (CrI 0.65%, 1.04%) compared with the least deprived at 0.48% (CrI 0.39%, 0.59%). Prior rounds of REACT-1 have shown different ethnicities at increased prevalence at different times, consistently higher prevalence of infection in larger households, and usually increased prevalence in more deprived neighborhoods (*20*–*25*). In models including each of the above variables, similar patterns were observed in the odds of testing positive, although odds were reduced when all three of the above variables were considered jointly, together with age, sex, region, and key worker status (table S9). Age remained an important predictor of swab positivity in these mutually adjusted models. Also, in these analyses, women had lower odds of infection than men at 0.80 (CI 0.67, 0.96) in round 13, although not in round 12

Table 2. Unadjusted and adjusted estimates of vaccine effectiveness against infection for self-reported vaccine status and linked vaccine status for rounds 12 and 13 of REACT-1 for participants aged 18 to 64 years.

Vaccination data source (*n*)	Adjustment	Vaccine effectiveness (2 doses)	
		Round 12	Round 13
Self-report, all positives, 18 to 64 years	Age, sex	61% (2%, 84%)	47% (18%, 65%)
	Age, sex, IMD, region, ethnicity	64% (11%, 85%)	49% (22%, 67%)
Self-report, symptomatic only, 18 to 64 years	Age, sex	81% (5%, 96%)	56% (19%, 77%)
	Age, sex, IMD, region, ethnicity	83% (19%, 97%)	59% (23%, 78%)
Linked, all positives, 18 to 64 years	Age, sex	75% (33%, 90%)	61% (36%, 76%)
	Age, sex, IMD, region, ethnicity	75% (35%, 90%)	62% (38%, 77%)

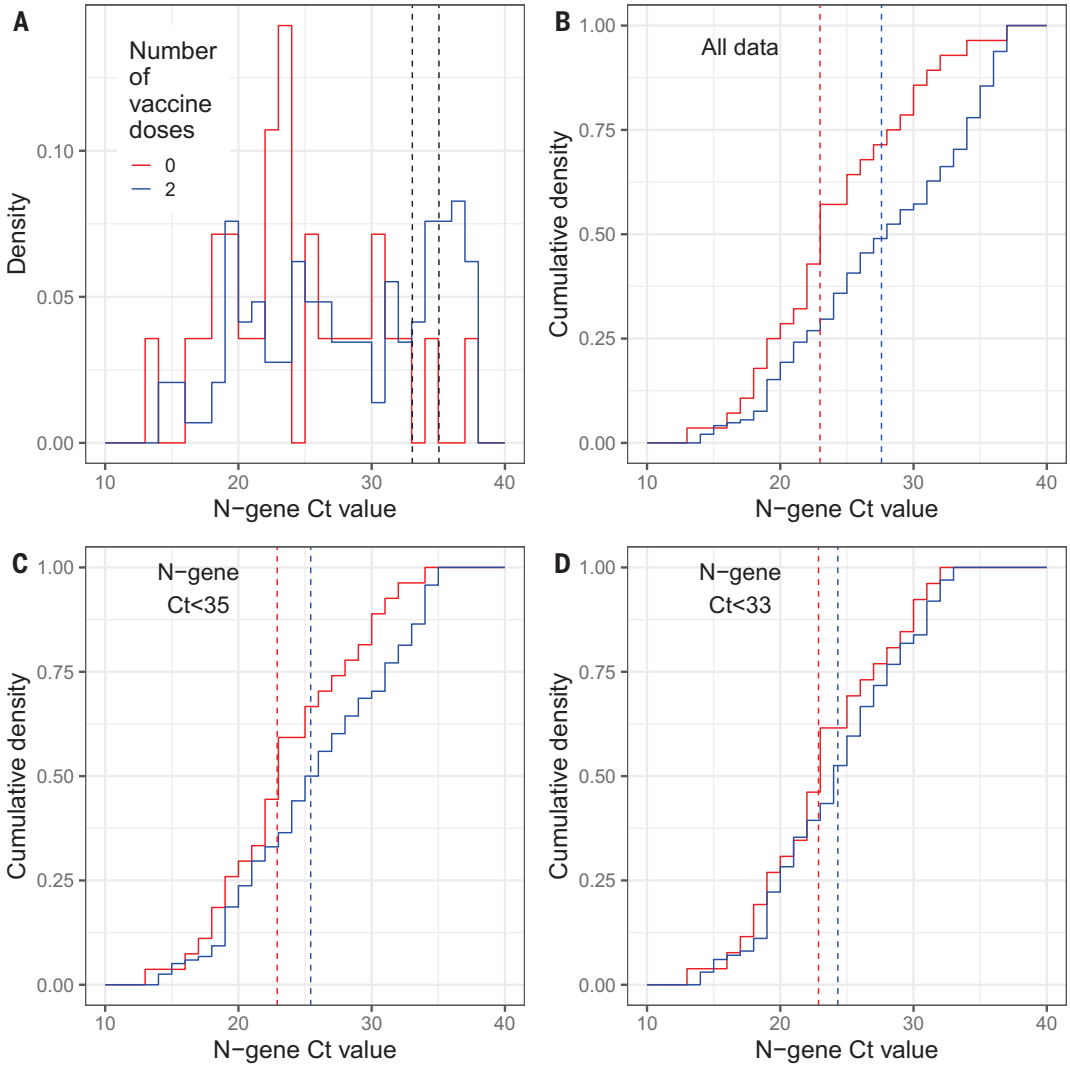


Fig. 2. Distribution of N-gene Ct values, by vaccine status, for positive samples obtained from individuals aged 18 to 64 years inclusive. (**A**) Distribution of all N-gene Ct values for those who are unvaccinated (red) and those who reported receiving two doses of a vaccine (blue). Also shown are two black dashed lines at N-gene Ct = 33 and 35; these show the threshold values for a sample to be classed as positive, used in sensitivity analyses. (**B**) Cumulative density of N-gene Ct values using all available data for unvaccinated individuals (red) and individuals who have had two doses of a vaccine (blue). (**C**) Cumulative density of N-gene Ct values using all data in which N-gene Ct is less than 35 for unvaccinated individuals (red) and individuals who have had two doses of a vaccine (blue). (**D**) Cumulative density of N-gene Ct values using all data in which N-gene Ct is less than 33 for unvaccinated individuals (red) and individuals who have had two doses of a vaccine (blue). In (B) to (D), red and blue vertical dashed lines show the median value for each distribution.

at 1.34 (CI 0.93, 1.92) (table S9); this difference may be related to increased social mixing associated with England's progression in the Euro 2020 football competition during June and July 2021, as was seen previously in Scottish data, reflecting their earlier exit from the competition (*26*).

Discussion

We report a rapidly rising prevalence of infection in England during 20 May to 12 July 2021 associated with the replacement of Alpha by the Delta variant in a highly vaccinated population. Our central estimate of VE against all SARS-CoV-2 infections for two doses of vaccine (self-report) was 49% in the most recent data, increasing to 58% when we defined effectiveness only for strong positives, and 62% in the linked data. These estimates are lower than some others (*19*, *27*, *28*) but consistent with more recent data from Israel (*29*).

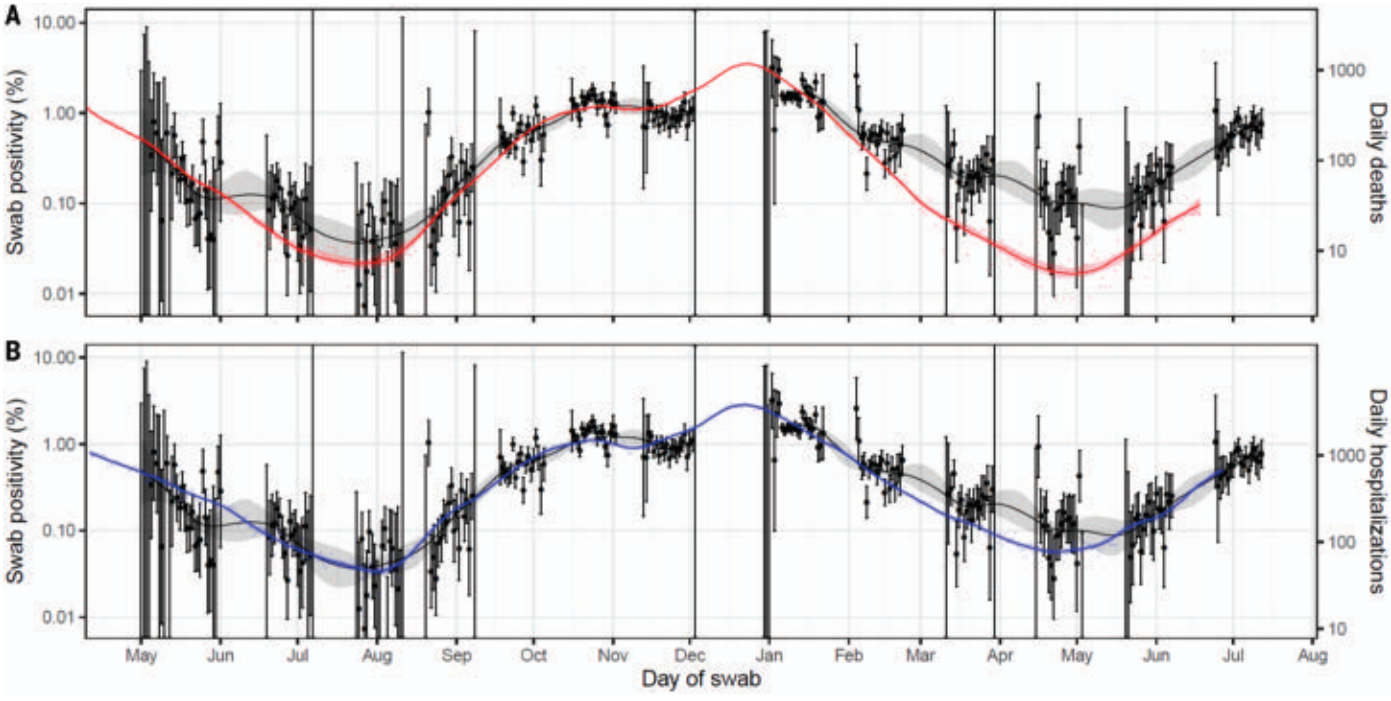


Fig. 3. Comparison of daily deaths and hospitalizations to swab positivity as measured by REACT-1. Daily swab positivity for all 13 rounds of the REACT-1 study (black points with 95% confidence intervals, left *y* axis) with P-spline estimates for swab positivity (solid black line; shaded area is 95% credible interval). (**A**) Daily deaths in England (red points, right *y* axis) and P-spline model estimates for expected daily deaths in England (solid red line, right *y* axis; shaded area is 95% credible interval). Daily deaths have been shifted by 26 (26, 26) days backward in time along the *x* axis. The two *y* axes have been scaled using the best-fit population adjusted scaling parameter 0.059 (0.058, 0.061). (**B**) Daily hospitalizations in England (blue points, right *y* axis) and P-spline model estimates for expected daily hospitalizations in England (solid blue line, right *y* axis; shaded area is 95% credible interval). Daily hospitalizations have been shifted by 20 (19, 20) days backward in time along the *x* axis. The two *y* axes have been scaled using the best-fit population adjusted scaling parameter 0.241 (0.236, 0.246).

Estimates of VE are not absolute but will vary depending on a variety of factors. Our estimates were higher when we restricted our analyses to people reporting symptoms of COVID-19 in the previous month and to those who consented to linkage of health records, although still lower than those from routine testing of symptomatic people presenting for RT-PCR in England (*27*). Unlike routine testing, our data are based on a random sample of the population and include asymptomatic people, as well as symptomatic individuals who may not present for testing; our results may therefore give a less biased representation of infection risk. Also, our estimated effectiveness was lower than that from a longitudinal household survey that included asymptomatic individuals but was conducted before the emergence of Delta, where vaccine status was based on a mix of self-reported and linked data (*19*). More generally, estimates of VE may depend on vaccine type, interval between doses, possible waning over time, and the extent of past natural infection among the comparator (unvaccinated) group.

We show that the third wave of infections in England was being driven primarily by the Delta variant in younger, unvaccinated people. This focus of infection offers considerable scope for interventions to reduce transmission among younger people, with knock-on benefits across the entire population. Also, given the rapid rise of the Delta variant that occurred in Europe, the US, South Asia and elsewhere, and its estimated increased transmissibility, patterns in England were informative of what was subsequently observed elsewhere. In our data, the highest prevalence of infection during June to July 2021 was among 13- to 24-year-olds. In the UK, the Joint Committee on Vaccinations and Immunizations recommended in August 2021 that vaccination should be offered to all 16- and 17-year-olds and then in September 2021 further extended the UK program to include children aged 12 to 15 years, as has been done in the US and some other countries. This expansion of the vaccination program to those at highest risk of infection had the potential to reduce transmission in the autumn and winter 2021 as levels of social mixing, including indoors, increased (*30*). Also, development of vaccines against Delta and other variants may be warranted in the light of evidence of antigenic change measured by neutralization (*31*) and the relationship between neutralization titer and protection from mild disease (*32*).

Estimates of VE against serious outcomes of greater than 90% have been reported for those who have received two doses of either BNT162b2 (*33*) or ChAdOx1-S (*34*) vaccines. This is in keeping with our observation of a weakening of the association between infections and hospitalizations and deaths from mid-February to early April 2021 when the Alpha variant was dominant. However, in our more recent data (since mid-April 2021), infections and hospitalizations began to reconverge, potentially reflecting the increased prevalence and severity of Delta compared with Alpha (*35*), a changing age mix of hospitalized cases to younger ages, and possible waning of protection (*29*, *36*).

Our study has limitations. One estimate of effectiveness was based on self-reported vaccine status, because we could only obtain linked vaccination data for the subset of participants who gave consent, with individuals who did and did not consent to linkage appearing to have different patterns of swab positivity across the vaccinated and unvaccinated groups. Because age, date of vaccination, and vaccine type are so strongly correlated in England, and with limitations in numbers, we were wary of introducing a time variable into the analyses to investigate the waning of VE explicitly. However, the design of the study—based on estimation of infection prevalence from independent

samples within (as well as across) separate rounds, conducted monthly—itself provides strong control for any time effects.

Over the course of the study since round 1 in May 2020, toward the end of the first lockdown in England, we observed a gradual reduction in response rates, from 30.5% in round 1 to 11.7% in round 13. These rates are conservative estimates because they are based on numbers of swabs with a valid RT-PCR result compared to the total number of letters of invitation sent out, some of which may have been returned, sent to the wrong address, or left unopened by the recipient. Nonetheless, the drop in response rates means that our sample may be becoming less representative, particularly in some groups such as young people (18 to 24 years) and those living in the most deprived areas where response rates by round 13 had fallen to 4.2% and 5.1%, respectively. Note, however, that these response rates have been achieved without the use of financial or other incentives.

Our method of sampling was designed initially to achieve sufficient numbers in each lower-tier local authority (LTLA) in England so that we could analyze subregional trends and also, by weighting the sample, provide estimates of prevalence that were representative of the population of England. Whereas previously we had aimed to achieve approximately equal numbers of people in our sample by LTLA, in rounds 12 and 13 we switched to sampling in proportion to population in order to capture greater resolution in inner-city areas, which were relatively underrepresented in our previous sampling regimen. In either case, as we reweight the sample according to the national population profile, weighted prevalence should be comparable across rounds, albeit with lower precision in later rounds because of the lower response rates.

Our data show that rapid exponential growth of SARS-CoV-2 prevalence occurred during the third wave in England at a time when the Delta variant became dominant. The rapid rollout of the vaccination program in England has so far limited the number of infections and serious cases relative to the unvaccinated population. Level or declining prevalence was observed during summer 2021 in the Northern Hemisphere, reflecting school vacations, greater time spent outdoors, and reduced social interactions. But without additional interventions, increased mixing (including indoors) in the presence of the Delta variant likely explains renewed growth that occurred in autumn 2021, even in populations with high levels of vaccination. Continued surveillance to monitor the spread of the epidemic is therefore required.

Materials and methods

The REACT-1 study methods have been described elsewhere (*9*). Briefly, at each round, we sent an invitation by post to named individuals from the list of patients registered with a National Health Service (NHS) general practitioner in England, obtained from NHS Digital, covering almost the entire population. We included all 317 LTLAs in England, and by combining the Isles of Scilly with Cornwall and the City of London with Westminster, we report results across 315 LTLAs overall.

For round 1 to round 11, we aimed to obtain approximately equal numbers of participants in each LTLA to be powered to provide local estimates of prevalence. From round 12 onward, we adjusted the sampling procedure to select the sample randomly in proportion to population at the LTLA level, thus obtaining more samples in LTLAs with higher population density in inner urban areas. However, we ensured that data were comparable across rounds as we reweighted the data at each round to be representative of England as a whole (see below).

For those registering to participate, we obtained age, sex, address, and residential postcode from the NHS register and collected additional information on demographics, health, and lifestyle via online or telephone questionnaire. This included information on ethnicity, smoking, household size, key worker status, contact with a known or suspected COVID-19 case, and whether, at time of survey, participants had experienced one or more of 29 symptoms in the past week or past month (participants not reporting symptoms may have developed symptoms later, but these were not captured). Participants were also asked for consent to longer-term follow-up through linkage to their NHS records including data from the national immunization program. The questionnaires are available on the study website (*37*).

Response rates have varied by age and over time and place, and are available for each round ["For Researchers: REACT-1 Study Materials" (*37*)]. Overall response rate was defined as the percentage of invitees from whom we received a valid swab result; this was 20.4% across all rounds, and 13.4% and 11.7% for rounds 12 and 13, respectively. In round 13, response rate varied by age from 4.2% at ages 18 to 24 years to 24% at ages 65 to 74 years and by IMD decile from 5.1% in the most deprived areas to 20.8% in the least deprived.

Participants were requested to provide a self-administered throat and nose swab (obtained by parent or guardian for children aged 5 to 12 years) following written and video instructions. Swabs were placed into a dry tube (no solution or preservative), refrigerated at home, picked up by courier, and then sent chilled to a single commercial laboratory for testing for SARS-CoV-2 by RT-PCR.

Ct threshold and laboratory calibration experiments

We tested two gene targets (the E and N genes) with Ct values used as a proxy for intensity of viral load. The RT-PCR test was considered positive if both gene targets were detected or if the N gene was detected with a Ct value less than 37. The Ct threshold used to determine positivity was set following three separate calibration experiments. First, 10 RNA extraction plates were sent from the commercial laboratory for blinded reanalysis in two laboratories accredited by the UK Accreditation Service (UKAS). We found concordant results for 919 negative samples and all 40 controls. We detected viral RNA in 11 of the 19 samples with a Ct value reported positive by the commercial laboratory (N gene Ct value ranging from 16.5 to 40.7); in 10 of these 11 samples, the N gene Ct value was <37. Second, in a serial dilution experiment of synthetic SARS-CoV-2 RNA, the commercial laboratory detected 2.5 copies at Ct 38; also while following serial dilution of known positive samples with low viral load, the commercial laboratory identified an N gene signal at Ct > 37 in most instances. Third, a Public Health England (PHE) reference laboratory reanalyzed a further 40 unblinded positive samples (on 19- × 96-well plates) with N gene Ct values > 35 (range 35.7 to 46.8) and without a signal for an E gene, detecting SARS-CoV-2 RNA in 15/40 (38%) samples (2/4 with N gene Ct value < 37). The results of all three calibration experiments were then consolidated to set the positivity criteria noted above, which have been used throughout each round of REACT-1.

Prevalence estimates and weighting

We obtained unweighted (crude) prevalence estimates for different sociodemographic and occupational groups by dividing counts of swab positivity (based on RT-PCR) by the number of swabs returned in that group. We then applied rim weighting (*38*) to provide prevalence weighted to be representative of the population of England as a whole, by age, sex, deciles of the IMD, LTLA counts, and ethnic group. We obtained the age by sex and LTLA counts from the Office for National Statistics mid-year population estimates (*39*) and counts by ethnic group from the Labour Force Survey (*40*), and calculated the IMD decile points from linkage of postcode to area-level IMD using the original sampling frame obtained from NHS Digital. Because of the different sources of population estimates, the rim weighting was based on proportions rather than population totals. We grouped age into nine categories: 5 to 12; 13 to 17; 18 to 24; 25 to 34; 35 to 44; 45 to 54; 55 to 64; 65 to 74; 75 years or above, giving 18 age-sex categories. Self-reported ethnicity was grouped into nine categories: white; mixed/multiple ethnic groups; Indian; Pakistani; Bangladeshi; Chinese; any other Asian background; Black African/Caribbean/other; and any other ethnic group or missing.

For the rim weighting, initially (first stage) the sample was weighted to LTLA counts and

age by sex groups only, adjusting the age and sex groups to ensure that the final weighted estimates were as close as possible to the population profile. Then, using the first-stage weights as starting weights, the rim weighting was adjusted for all four measures, with the adjustment factor between the first- and second-stage weights trimmed at the 1st and 99th percentiles to dampen the extreme weights and improve efficiency. The final weights were calculated as the first-stage weights multiplied by the trimmed adjustment factor for the second stage, with confidence intervals for weighted prevalence estimates calculated using the "survey" package in R (*41*).

Statistical analyses

Statistical analyses were carried out in R (*42*). To investigate the potential confounding effects of covariates on prevalence estimates, we performed logistic regression on swab positivity as the outcome, and sex, age, region, employment type, ethnicity, household size, and neighborhood deprivation as explanatory variables. We adjusted for age and sex, and mutually adjusted for the other covariates to obtain odds ratio estimates and 95% confidence intervals. We decided not to adjust for multiple testing to facilitate direct comparisons with other publications where only comparison-wise error rate (CER) has been controlled for (*43*).

We estimated adjusted VE as 1 – (odds ratio) where the odds ratio was obtained from comparing vaccinated and unvaccinated individuals in a logistic regression model with swab positivity as outcome and with adjustment for age and sex, and age, sex, IMD quintile, and ethnicity.

To estimate the underlying geographical variation in prevalence at the local (subregional) level, we used a neighborhood spatial smoothing method based on nearest neighbor up to 30 km. We calculated N_n, the median number of study participants within 30 km of each study participant for each round or subround. We then calculated the local prevalence for 15 members of each LTLA as an estimate of the smoothed neighborhood prevalence in that area.

To analyze trends in swab positivity over time, we used an exponential model of growth or decay with the assumption that the weighted number of positive samples (from the weighted total number of samples) each day arose from a binomial distribution. The model is of the form $(t) = I_0.e$, where $I(t)$ is the swab positivity at time t, I_0 is the swab positivity on the first day of data collection per round, and r is the growth rate. The binomial likelihood for P (out of N) positive tests on a given day is then $P \sim (N, I_0.e^{rt})$ based on day of swabbing or, if unavailable, day of sample collection. We used a bivariate No-U-Turn sampler to estimate posterior credible intervals assuming uniform prior distributions on I_0 and r (*44*). We estimated the reproduction number R assuming a generation time that follows a gamma distribution with a shape parameter, n, of 2.29 and a rate parameter, β, of 0.36 (corresponding to a mean generation time of 6.29 days) (*45*). R was estimated from the equation $R = (1 + r/\beta)\hat{}n$ (*46*) using data from two sequential rounds and separately per round. We carried out a range of sensitivity analyses including estimation of R for different thresholds of Ct values that determine swab positivity and for nonsymptomatic individuals (not reporting symptoms on the day of swab or month prior).

We fit a Bayesian penalized spline (P-spline) model (*47*) to the daily data using a No-U-Turn Sampler in logit space, segmenting the data into approximately 5-day sections by regularly spaced knots, with further knots beyond the study period to minimize edge effects. We defined fourth-order basis splines (b-splines) over the knots with the final model consisting of a linear combination of these b-splines. We guarded against overfitting by including a second-order random-walk prior distribution on the coefficients of the b-splines, taking the form $b_i = 2b_{i-1} - b_{i-2} + u_i$, where b_i is the ith b-spline coefficient and u_i is normally distributed with $u_i \sim N(0, \rho^2)$. This prior penalizes against changes in the growth rate unless supported by the data; the strength of the penalization is determined by the parameter ρ for which we assume an inverse gamma prior distribution, $\rho \sim IG(0.001, 0.001)$. We assume that the first two b-spline coefficients have uniform distribution (i.e., b_1 and $b_2 \sim$ constant).

We compared daily prevalence data from rounds 1 to 13 of REACT-1 with publicly available national daily hospital admissions and COVID-19 mortality data (deaths within 28 days of a positive test). To do this, we fit P-spline models as before to the daily hospital admissions and to the daily death data in order to obtain estimates for the expected number of outcomes on a given day. We then fit a simple two-parameter model consisting of a lag time between the posterior of the P-spline estimate for each of hospitalizations or deaths, the daily weighted prevalence calculated from REACT-1 data, and a scaling parameter, corresponding to the percentage of people who were swab-positive in the population on a particular day in comparison with future hospitalizations or deaths. Because of the time delay between the REACT-1 prevalence signal and daily hospitalizations and deaths, the model was only fit to rounds 1 to 12. We then compared round 13 data to the estimated trend in hospitalizations and deaths to visualize any alterations in the link between these parameters and infection prevalence as measured in REACT-1. We estimated these relationships for all ages and separately for those aged under 65 years, and those 65 years and above.

To visualize the trends of the REACT-1 data over time, we also fitted P-splines to all subsets of the REACT-1 data examined. For the REACT-1 data split by age (below 65 years and 65 years and above), we fit a mixed P-spline model in which a P-spline was fit separately to each age group but the smoothing parameter, ρ, was fit to both datasets simultaneously. Further changes in the first derivative were assumed to happen at the same time for both datasets, with the condition $u_{i,<65} - u_{i,65+} \sim N(0, \eta^2)$ and η given an uninformative prior distribution, $\eta \sim IG(0.001, 0.001)$.

Viral genome sequencing

RT-PCR positive swab samples where there was sufficient sample volume and with N gene Ct values of <32 were sent frozen from the laboratory to the Quadram Institute (Norwich, UK) for viral genome sequencing. Amplification of viral RNA used the ARTIC protocol (*48*) and sequencing libraries were prepared using CoronaHiT (*49*). Analysis of sequencing data used the ARTIC bioinformatic pipeline (*50*) with lineages assigned using PangoLEARN (*51*).

We fit a Bayesian logistic regression model to the proportion of lineages that were identified as the Delta variant from round 10 to round 13 to obtain a daily growth rate advantage between Delta and other circulating lineages, Δr. Assuming an exponential generation time of mean 6.29 days (*45*), the reproduction number, R, is given by $R = 1 + r \times \underline{g}$ (*46*). The estimate of growth rate advantage can thus be converted into an additive R advantage through the equation $\Delta R = \Delta r \times \underline{g}$, assuming the mean generation time is the same for all lineages. We chose not to estimate a multiplicative R advantage (*52*), because it relies on the assumption of a zero-variance discrete generation time interval, which is less consistent with estimates of an overdispersed serial interval (*45*).

As a sensitivity the model was also fit to data from only round 11 to round 12 to check that edge effects were not introducing bias. The upper bound of prevalence for non-Delta lineages (none of which were detected in round 13) was estimated by calculating the 95% Wilson upper bound on the proportion of non-Delta lineage detected, then multiplying by the weighted prevalence estimate for round 13. This was then multiplied by the population of England to get an estimate for the upper bound on the average number of people infected with a non-Delta lineage at any one time during round 13.

Data availability

Access to REACT-1 individual-level data is restricted to protect participants' anonymity. Summary statistics, descriptive tables, and code from the current REACT-1 study are available at https://github.com/mrc-ide/reactidd. REACT-1 study materials are available for each

round at www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/react-1-study-materials/.

Public involvement

A Public Advisory Panel provides input into the design, conduct, and dissemination of the REACT research program.

Ethics

We obtained research ethics approval from the South Central–Berkshire B Research Ethics Committee (IRAS ID: 283787).

REFERENCES AND NOTES

1. P. M. Folegatti *et al*., Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: A preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* **396**, 467–478 (2020). doi: 10.1016/S0140-6736(20)31604-4; pmid: 32702298
2. F. P. Polack *et al*., Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020). doi: 10.1056/NEJMoa2034577; pmid: 33301246
3. Johns Hopkins University Coronavirus Resource Center, https://coronavirus.jhu.edu/.
4. M. S. Dhar *et al*., Genomic characterization and Epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. medRxiv 21258076 [preprint] (2021). doi: 10.1101/2021.06.02.21258076
5. Coronavirus (COVID-19) Vaccinations. *Our World In Data*; https://ourworldindata.org/covid-vaccinations.
6. UK Government, Covid-19 Dashboard; https://coronavirus.data.gov.uk/.
7. UK Government, "Prime Minister sets out roadmap to cautiously ease lockdown restrictions"; www.gov.uk/government/news/prime-minister-sets-out-roadmap-to-cautiously-ease-lockdown-restrictions.
8. "Moving to step 4 of the roadmap." UK Government (2021); www.gov.uk/government/publications/covid-19-response-summer-2021-roadmap/moving-to-step-4-of-the-roadmap.
9. S. Riley *et al*., REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: Study protocol. *Wellcome Open Res.* **5**, 200 (2021). doi: 10.12688/wellcomeopenres.16228.1; pmid: 33997297
10. S. Riley *et al*., Resurgence of SARS-CoV-2: Detection by community viral surveillance. *Science* **372**, 990–995 (2021). doi: 10.1126/science.abf0874; pmid: 33893241
11. Public Health England, "SARS-CoV-2 variants of concern and variants under investigation in England: Technical Briefing 18" (2021).
12. N. Ferguson, "B.1.617.2 transmission in England: risk factors and transmission advantage" (2021); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993159/S1270_IMPERIAL_B.1.617.2.pdf.
13. A. H. Allen *et al*., "Increased household transmission of COVID-19 cases associated with SARS-CoV-2 Variant of Concern B.1.617.2: a national case- control study" (2021); https://khub.net/documents/135939561/405676950/Increased+Household+Transmission+of+COVID-19+Cases+-+national+case+study.pdf/7f7764fb-ecb0-da31-77b3-b1a8ef7be9aa.
14. Office for National Statistics, UK, "Coronavirus (COVID-19) Infection Survey, UK: 23 July 2021" (2021); www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/23july2021.
15. Office for National Statistics, UK, "Population estimates for the UK, England and Wales, Scotland and Northern Ireland mid-2020" (2021); www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020.
16. J. Wallinga, M. van Boven, M. Lipsitch, Optimizing infectious disease interventions during an emerging epidemic. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 923–928 (2010). doi: 10.1073/pnas.0908491107; pmid: 20080777
17. D. McLennan, S. Noble, M. Noble, E. Plunkett, G. N. Wright, "The English Indices of Deprivation 2019" (2019); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833951/IoD2019_Technical_Report.pdf.
18. A. Singanayagam *et al*., Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill.* **25**, (2020). doi: 10.2807/1560-7917.ES.2020.25.32.2001483; pmid: 32794447
19. E. Pritchard *et al*., Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat. Med.* **27**, 1370–1378 (2021). doi: 10.1038/s41591-021-01410-w; pmid: 34108716
20. S. Riley *et al*., REACT-1 round 9 interim report: downward trend of SARS-CoV-2 in England in February 2021 but still at high prevalence. MedRxiv 21251973 [preprint] (2021). doi: 10.1101/2021.02.18.21251973
21. S. Riley *et al*., REACT-1 round 8 interim report: SARS-CoV-2 prevalence during the initial stages of the third national lockdown in England. MedRxiv 21250158 [preprint] (2021). doi: 10.1101/2021.01.20.21250158
22. S. Riley *et al*., REACT-1 round 7 updated report: regional heterogeneity in changes in prevalence of SARS-CoV-2 infection during the second national COVID-19 lockdown in England. MedRxiv 20248244 [preprint] (2020). doi: 10.1101/2020.12.15.20248244
23. S. Riley *et al*., REACT-1 round 6 updated report: high prevalence of SARS-CoV-2 swab positivity with reduced rate of growth in England at the start of November 2020. MedRxiv 20233932 [preprint] (2020). doi: 10.1101/2020.11.18.20233932
24. S. Riley *et al*., High prevalence of SARS-CoV-2 swab positivity and increasing R number in England during October 2020: REACT-1 round 6 interim report. MedRxiv 20223123 [preprint] (2020). doi: 10.1101/2020.10.30.20223123
25. S. Riley *et al*., High and increasing prevalence of SARS-CoV-2 swab positivity in England during end September beginning October 2020: REACT-1 round 5 updated report. MedRxiv 20211227 [preprint] (2020). doi: 10.1101/2020.10.12.20211227
26. Public Health Scotland, "Public Health Scotland COVID-19 Statistical Report as at 28 June 2021" (2021); www.google.com/url?q=https://www.publichealthscotland.scot/media/8268/21-06-30-covid19-publication_report.pdf&sa=D&source=editors&ust=1627937430378000&usg=AOvVaw2Kwz_u0_KQraqrxqTW-xyX.
27. J. Lopez Bernal *et al*., Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N. Engl. J. Med.* **385**, 585–594 (2021). doi: 10.1056/NEJMoa2108891; pmid: 34289274
28. N. Dagan *et al*., BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *N. Engl. J. Med.* **384**, 1412–1423 (2021). doi: 10.1056/NEJMoa2101765; pmid: 33626250
29. Ministry of Health, Israel, "Vaccine Efficacy Among Those First Vaccinated" (2021); www.gov.il/BlobFolder/reports/vaccine-efficacy-safety-follow-up-committee/he/files_publications_corona_two-dose-vaccination-data.pdf.
30. S. Saxena, H. Skirrow, K. Wighton, Should the UK vaccinate children and adolescents against covid-19? *BMJ* **374**, n1866 (2021). doi: 10.1136/bmj.n1866; pmid: 34301635
31. D. Planas *et al*., Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280 (2021). doi: 10.1038/s41586-021-03777-9; pmid: 34237773
32. D. S. Khoury *et al*., Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211 (2021). doi: 10.1038/s41591-021-01377-8; pmid: 34002089
33. E. J. Haas *et al*., Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: An observational study using national surveillance data. *Lancet* **397**, 1819–1829 (2021). doi: 10.1016/S0140-6736(21)00947-8; pmid: 33964222
34. J. Stowe, N. Andrews, C. Gower, E. Gallagher, L. Utsi, R. Simmons, Effectiveness of COVID-19 vaccines against hospital admission with the Delta (B.1.617.2) variant (Public Health England Library); https://khub.net/web/phe-national/public-library/-/document_library/v2WsRK3ZlEig/view_file/479607329?_com_liferay_document_library_web_portlet_DLPortlet_INSTANCE_v2WsRK3ZlEig_redirect=https%3A%2F%2Fkhub.net%3A443%2Fweb%2Fphe-national%2Fpublic-library%2F-%2Fdocument_library%2Fv2WsRK3ZlEig%2Fview%2F479607266.
35. A. Sheikh, J. McMenamin, B. Taylor, C. Robertson, SARS-CoV-2 Delta VOC in Scotland: Demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* **397**, 2461–2462 (2021). doi: 10.1016/S0140-6736(21)01358-1; pmid: 34139198
36. S. J. Thomas *et al*., C4591001 Clinical Trial Group, Six month safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. bioRxiv 21261159 [preprint] (2021). doi: 10.1101/2021.07.28.21261159
37. Real-time Assessment of Community Transmission (REACT) study (Imperial College); www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/.
38. T. Sharot, "Weighting Survey Results" (1986); http://redresearch.com/wp/wp-content/uploads/2016/01/Weighting-Survey-Results.pdf.
39. N. Park, Population estimates for the UK, England and Wales, Scotland and Northern Ireland. Office for National Statistics (2020); www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest.
40. Office for National Statistics, UK, Annual Population Survey/Labour Force Survey; www.nomisweb.co.uk/sources/aps.
41. T. Lumley, Analysis of Complex Survey Samples. *J. Stat. Softw.* **9**, 1–19 (2004).
42. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020); www.R-project.org/.
43. R. Bender, S. Lange, Adjusting for multiple testing—When and how? *J. Clin. Epidemiol.* **54**, 343–349 (2001). doi: 10.1016/S0895-4356(00)00314-0; pmid: 11297884
44. M. D. Hoffman, A. Gelman, The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. arXiv 1111.4246 [stat.CO] (2011); http://arxiv.org/abs/1111.4246.
45. Q. Bi *et al*., Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020). doi: 10.1016/S1473-3099(20)30287-5; pmid: 32353347
46. J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604 (2007). doi: 10.1098/rspb.2006.3754; pmid: 17476782
47. S. Lang, A. Brezger, Bayesian P-Splines. *J. Comput. Graph. Stat.* **13**, 183–212 (2004). doi: 10.1198/1061860043010
48. J. Quick, nCoV-2019 sequencing protocol v3 (LoCost) (2020); www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye.
49. D. J. Baker *et al*., CoronaHiT: High-throughput sequencing of SARS-CoV-2 genomes. *Genome Med.* **13**, 21 (2021). doi: 10.1186/s13073-021-00839-5; pmid: 33563320
50. *A Nextflow Pipeline for Running the ARTIC Network's Field Bioinformatics Tools* (Github; https://github.com/connor-lab/ncov2019-artic-nf).
51. *Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN)* (Github; https://github.com/cov-lineages/pangolin).
52. N. G. Davies *et al*., Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021). doi: 10.1126/science.abg3055; pmid: 33658326
53. S. Riley *et al*., reactidd R package with data, Version 0.92, Zenodo (2021); http://doi.org/10.5281/zenodo.5574472.

ACKNOWLEDGMENTS

We thank key collaborators on this work—Ipsos MORI: K. Beaver, S. Clemens, G. Welch, N. Gilby, K. Ward, G. Pantelidou, and K. Pickering; School of Public Health, Imperial College London: E. Johnson, R. Elliott, G. Blakoe; Institute of Global Health Innovation at Imperial College: G. Fontana, S. Satkunarajah, D. Thompson, and L. Naar; North West London Pathology and Public Health England for help in calibration of the laboratory analyses; Patient Experience Research Centre at Imperial College and the REACT Public Advisory Panel; Quadram Institute, Norwich, UK: T. L. Viet, N.-F. Alikhan, L. M. Jackson, C. Ludden; NHS Digital for access to the NHS register; the Department of Health and Social Care for logistic support; and the COVID-19 Taskforce of the Royal Statistical Society (UK) for helpful comments. S.R. acknowledges helpful discussion with attendees of meetings of the UK Government Scientific Pandemic Influenza–Modelling (SPI-M) committee. **Funding:** The study was funded by the Department of Health and Social Care in England. Sequencing was provided through funding from the COVID-19 Genomics UK (COG-UK) Consortium. P.E. is director of the Medical Research Council (MRC) Centre for Environment and Health (MR/L01341X/1, MR/S019669/1) and was supported by Health Data Research UK (HDR UK), the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre, NIHR Health Protection Research Unit

(HPRU) in Chemical and Radiation Threats and Hazards, NIHR HPRU in Environmental Exposures and Health, the British Heart Foundation Centre for Research Excellence at Imperial College London (RE/18/4/34215), and the UK Dementia Research Institute at Imperial (MC_PC_17114). Also supported by the MRC Centre for Global Infectious Disease Analysis, NIHR HPRU in Modelling and Health Economics, Wellcome Trust (200861/Z/16/Z, 200187/Z/15/Z), and US Centers for Disease Control and Prevention (U01CK0005-01-02) (S.R. and C.A.D.);. an NIHR Professorship (G.C.); and an NIHR Senior Investigator Award and the Wellcome Trust (205456/Z/16/Z) (H.War.). We thank the Huo Family Foundation for their support of our work on COVID-19. Quadram authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); their research was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10352. We thank members of the COVID-19 Genomics Consortium UK (COG-UK) for their contributions to generating the genomic data used in this study. COG-UK is supported by funding from the MRC, part of UK Research & Innovation (UKRI), NIHR and Genome Research Ltd., operating as the Wellcome Sanger Institute. **Author contributions:** P.E. and S.R. conceptualized and designed the study and drafted the manuscript. S.R., D.W., H.Wan., O.E., C.E.W., and K.E.C.A. undertook the data analysis. P.J.D., C.F., D.A., and C.A.D. provided statistical advice. A.J.P., A.J.T., and S.J.P. undertook the viral genome sequencing analysis. W.B., G.T., C.A., G.C., H.War., and A.D. provided study oversight. A.D. and P.E. obtained funding. All authors critically reviewed the manuscript and read and approved the final version of the manuscript. P.E. is the guarantor for this paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted, had full access to all the data in the study, and had final responsibility for the decision to submit for publication. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Code and additional data to support the figures are freely available at (*53*). 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.org/doi/10.1126/science.abl9551
Materials and Methods
Tables S1 to S9
Figs. S1 to S3
Data S1
COG-UK Consortium member list
MDAR Reproducibility Checklist

16 August 2021; accepted 29 October 2021
Published online 2 November 2021
10.1126/science.abl9551

RESEARCH ARTICLES

CARBON CAPTURE

A scalable metal-organic framework as a durable physisorbent for carbon dioxide capture

Jian-Bin Lin[1]†, Tai T. T. Nguyen[2], Ramanathan Vaidhyanathan[1,3], Jake Burner[4], Jared M. Taylor[1,5], Hana Durekova[4], Farid Akhtar[6], Roger K. Mah[1,5], Omid Ghaffari-Nik[7], Stefan Marx[8], Nicholas Fylstra[1], Simon S. Iremonger[1] , Karl W. Dawson[1] , Partha Sarkar[2], Pierre Hovington[7]*, Arvind Rajendran[2]*, Tom K. Woo[4]*, George K. H. Shimizu[1,5]*

Metal-organic frameworks (MOFs) as solid sorbents for carbon dioxide (CO_2) capture face the challenge of merging efficient capture with economical regeneration in a durable, scalable material. Zinc-based Calgary Framework 20 (CALF-20) physisorbs CO_2 with high capacity but is also selective over water. Competitive separations on structured CALF-20 show not just preferential CO_2 physisorption below 40% relative humidity but also suppression of water sorption by CO_2, which was corroborated by computational modeling. CALF-20 has a low enthalpic regeneration penalty and shows durability to steam (>450,000 cycles) and wet acid gases. It can be prepared in one step, formed as composite materials, and its synthesis can be scaled to multikilogram batches.

Capture of CO_2 after fossil fuel combustion requires CO_2 removal from a localized emission source but also regeneration and recycling of the capture system. Major challenges for the capture stage span materials design and development through to process engineering (*1*, *2*). Flue gas has a low concentration of CO_2 diluted in mostly N_2 along with water and acid gases (*3*). Amine and solvent systems (*4*, *5*) rely on contacting flue gas with a liquid that absorbs the CO_2 through a combination of chemical and physical absorption. Although CO_2 removal is effective, regeneration is energy intensive and can lead to chemical decomposition (*6*).

Solid sorbents represent a step-change technology for carbon capture (*7*–*10*) and have been demonstrated at smaller scales (*11*). Solids can bind CO_2 through either chemical or physical sorption (*3*, *7*–*10*). In most cases, chemisorptive materials have higher capacity and selectivity for CO_2 (*12*). However, factors that enhance CO_2 binding often proportionally increase the energy needed to regenerate the sorbent and can enhance binding of competing gases. For the absolute CO_2 uptake, the relevant parameter is working capacity under the operational cycling conditions to regenerate the solid sorbent (*3*). Selectivity over N_2 is typically reported, but sorption of CO_2 in the presence of water vapor is much less reported, especially for physisorptive capture systems (*12*–*14*). A physisorptive CO_2 capture solid would offer much lower regeneration costs, but it must have sufficient working capacity and selectivity in an actual flue stream in which gases are present with stronger intermolecular attractive forces than those of CO_2. Moreover, to translate to process productivity, the kinetics of sorption and release are as important as capacity.

Nearly all classes of porous solids have potential as solid sorbents for CO_2 capture (*1*–*3*, *7*–*10*), including metal-organic frameworks (MOFs) (*2*, *3*, *9*, *12*–*14*), in which chemical building blocks, pore sizes and shapes, surface functionalities, and even degrees of order can be varied to optimize CO_2 capture ability. More robust MOFs (*15*, *16*), including ones that are stable in the presence of water (*17*–*19*) and steam (*20*), have been reported, although stability to wet acid gases is less common (*21*–*23*). For sorbent powder to be a usable material, it must be capable of formation in macroscopic shape for rapid mass transfer and thermal management, be durable in that form, and be available at scale (hundreds of thousands of tonnes) and reasonable cost (*24*, *25*).

Solid sorbents optimized in an adsorption process have the potential to substantially decrease the CO_2 capture cost compared with traditional amine absorption processes because of lower regeneration energy, less chemical decomposition versus the solvent capture system, extensive use of stainless-steel owing to the corrosivity of amine solvents, and large plant footprint (*4*–*6*). Optimization of the solid sorbent process must include high volumetric productivity in the presence of water (present in the flue gas) and the lowest regeneration energy. For regeneration, several processes are under evaluation, including vacuum swing, pressure swing, and temperature swing (*26*). Although cycling performance per sorbent volume or productivity is one of the main drivers of final CO_2 capture cost, there are several other parameters that affect operating and capital expenses of CO_2 capture. For solid sorbents, unlike solvent-based absorption, it is not feasible to continuously replace deactivated sorbents with fresh ones.

Here we present Calgary Framework 20 (CALF-20), a MOF with high capacity and selectivity for CO_2 despite a physisorptive mechanism and modest heat of adsorption. Its selectivity extends beyond N_2 to capture CO_2 in a wet gas. CALF-20 is exceptionally robust and stable to steam, wet acid gases, and even prolonged exposure to direct flue gas from natural gas combustion. Its single-step synthesis from commercially available components is highly scalable. The origin of the CO_2 philicity, despite CALF-20 being highly water resistant, was studied by simulation. Structuring of CALF-20 was performed, as well as competitive breakthrough experiments in wet gas streams that aligned with pure-component isotherms, heats of adsorption, and molecular modeling. In particular, not only can CALF-20 physisorb CO_2 up to and beyond 40% RH, but the presence of CO_2 actually suppresses water sorption. Finally, we present durability and CO_2 capture data on the MOF that are based on industrial testing.

Synthesis, structure, and gas sorption

CALF-20, [Zn_2(1,2,4-triazolate)$_2$(oxalate)], was initially prepared solvothermally and single crystals obtained through the in situ degradation of a dihydroxybenzoquinone derivative (see supplementary materials). CALF-20 is composed of layers of 1,2,4-triazolate-bridged zinc(II) ions pillared by oxalate ions to form a three-dimensional (3D) lattice and 3D pore structure (Fig. 1, A to C). Channels of 2.73 Å by 2.91 Å, 1.94 Å by 3.11 Å, and 2.74 Å by 3.04 Å along [100], [011], and [0$\bar{1}$1], respectively (factoring van der Waals radii), that permeate the MOF result in a ~38% void volume. The one crystallographically unique Zn center is five-coordinate with a distorted trigonal bipyramidal geometry [Zn-O = 2.022(2), 2.189(3) Å; Zn-N = 2.007(2), 2.016(3), 2.091 (3) Å]. The N atoms in the 1,2 positions of the triazolate bridge Zn dimers are linked to the next dimer by the N atom in the 4-position. The Zn coordination is completed by two oxygen atoms of a chelating oxalate group, and there are no open coordination sites. The bulk powder shows the same phase (Fig. 1D). Detailed structural analyses on pillared zinc triazolates have shown that layers can exist

[1]Department of Chemistry, University of Calgary, Calgary, Alberta, Canada. [2]Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, Canada. [3]Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pashan, Pune, Maharashtra, 411008, India. [4]Department of Chemistry and Biomolecular Science, University of Ottawa, Ottawa, Ontario, Canada. [5]ZoraMat Solutions Inc., Calgary, Alberta, Canada. [6]Department of Materials Engineering, Luleå University of Technology, Luleå, Sweden. [7]Svante Inc., Vancouver, British Columbia, Canada. [8]BASF SE, Ludwigshafen am Rhein, Germany.
*Corresponding author. Email: phovington@svanteinc.com (P.H.); arvind.rajendran@ualberta.ca (A.R.); twoo@uottawa.ca (T.K.W.); gshimizu@ucalgary.ca (G.K.H.S.)
†Present address: C-CART, CREAIT Network, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada.

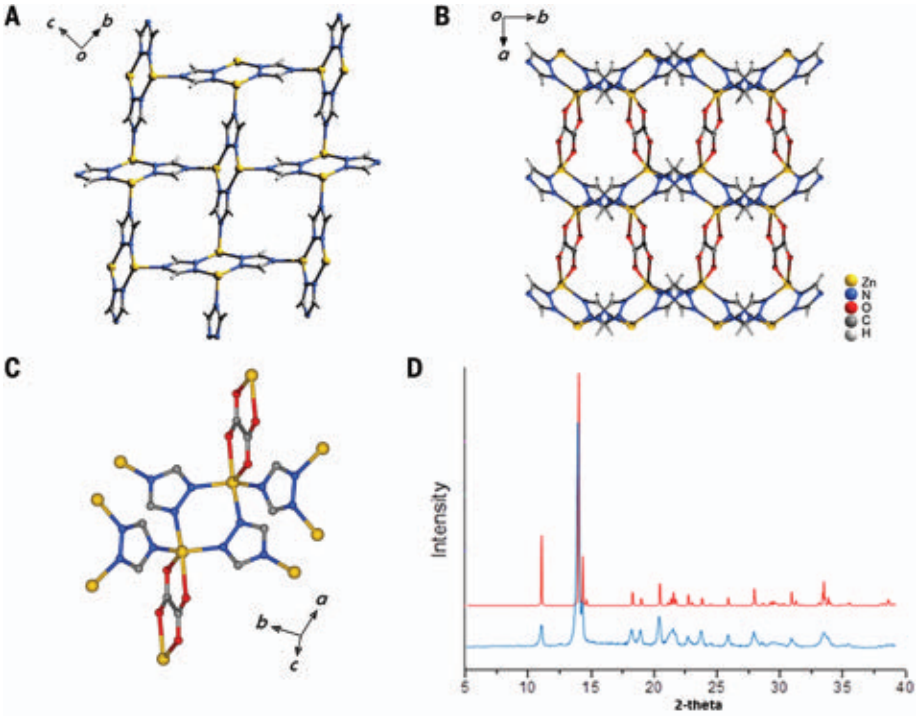


Fig. 1. Single-crystal structure of CALF-20. (**A**) View of the two-dimensional zinc triazolate grid. (**B**) View orthogonal to (A) showing the pillaring of the zinc triazolate layers by oxalate anions. (**C**) View of the zinc coordination sphere (H atoms removed). (**D**) Powder x-ray pattern simulated from the single-crystal structure (top) and obtained experimentally.

in different manifestations with varying degrees of buckling (*27*, *28*). Indeed, since a provisional patent application was filed in 2014, a hydrated form of [Zn_2(1,2,4-triazolate)$_2$(oxalate)] has been reported (*29*). This structure has the same connectivity but slightly different unit cell and pore dimensions. The specific pore structure affects sorption properties, and modeling was carried out with our obtained crystal data.

Gas adsorption experiments were performed for CO_2 and N_2 (Fig. 2A). The Langmuir surface area calculated from the N_2 isotherm at 77 K was 528 $m^2\ g^{-1}$, and the uptake for CO_2 was 4.07 mmol g^{-1} at 1.2 bar and 293 K. The zero-loading heat of adsorption for CO_2 was −39 kJ mol^{-1} (fig. S6), and the calculated selectivity for CO_2/N_2 by ideal adsorbed solution theory was 230 for a 10:90 CO_2/N_2 mixture. CALF-20 structured readily as a 20% polysulfone composite and retained the expected porosity (Fig. 2, B and C, and fig. S5). For CO_2 capacity and selectivity over N_2 as metrics, there are numerous other materials with noteworthy performance (*30*–*36*). The water sorption profile of CALF-20 was unusual in that, for a solid with good physisorptive capacity for CO_2, it exhibited poor water uptake at low partial pressures (Fig. 2, D and E). Comparisons to zeolite 13X (*37*), as well as two other water-resistant MOFs, CAU-10 (*38*) and Al fumarate (*39*), are included in Fig. 2 and fig. S16. Moreover, higher-temperature water isotherms showed that water uptake decreased more readily at higher temperatures than did the corresponding CO_2 isotherms.

Binding-site modeling

To gain insights into the nature of CO_2 binding in CALF-20 and its unusual water sorption behavior, we performed atomistic grand canonical Monte Carlo (GCMC) simulations (see supplementary materials). The experimental and simulated CO_2 and N_2 isotherms were in excellent agreement (see supplementary materials). Probability distributions of the guest molecules within the MOF allowed us to identify binding sites. The most probable CO_2 binding, which lies in the middle of the CALF-20 pore (Fig. 3A), had a binding energy of −34.5 kJ mol^{-1} based on the GCMC force field; the density functional theory (DFT) value with dispersion corrections was −36.5 kJ mol^{-1}. The interatomic distances displayed were consistent with physisorption; the shortest distance was 3.03 Å between the CO_2 oxygen and a hydrogen of the triazole (fig. S8). Analysis of the binding energy revealed that the CO_2-MOF interaction was dominated by attractive dispersion interactions (85%), with electrostatics contributing the balance.

Water adsorption isotherms are more challenging to simulate given the polar nature of water, which enables potentially strong interactions with the framework and with itself. The experimental water isotherm had a general S-shape, where the water uptake was initially low until ~10% relative humidity (RH), at which point there was a steep rise until ~30% RH (Fig. 3B). These features indicated that water condensed in the pores, and they were reproduced in the simulated isotherm. After the initial steep rise in water uptake beyond 30% RH, the experimental isotherm showed a more gradual increase in adsorption until reaching a saturation limit at ~11 mmol g^{-1}. However, for the simulated isotherm, the steep rise continued until full saturation at 40% RH and then flattened. The general S-shape and the saturation capacity of ~11 mmol g^{-1} were reproduced by the simulation.

A snapshot from the pure water simulation at 20% RH, where the water uptake was roughly half the saturation limit (Fig. 3C), revealed that the pores were either full of water molecules, forming a hydrogen-bonded network, or completely empty. In comparison, at 60% RH, where uptake had fully saturated, all the pores were full of hydrogen-bonded water molecules (fig. S9). The equilibrium distribution at 20% RH, where partially filled pores were not observed, suggested rapid condensation or evaporation of water. We extracted the water binding sites at 20% RH with the highest probability from the GCMC simulations, and the top three binding sites, in order, are labeled i, ii, and, iii in Fig. 3D. The binding energies with the framework of the sites, −17.5, −8.9, and −29.1 kJ mol^{-1}, respectively, were calculated by placing a single water molecule in the site with no other guest molecule present. The two most probable binding sites had a relatively low binding energy and were oriented away from the framework such that there were no hydrogen-bonding interactions with the oxalate linkers. Water molecules in these sites were poised to form hydrogen-bonding interactions with other water molecules, which suggested that the main driver for the initial water uptake was the interaction with other water molecules. This result was consistent with the experimentally observed water-uptake properties of CALF-20 at low RH.

Breakthrough studies

The intriguing CO_2 and water isotherms prompted a series of dynamic breakthrough studies (Fig. 4) on the CALF-20–polysulfone composite (see supplementary materials). Competitive CO_2/N_2 studies, with CO_2/N_2 mixtures of 5/95, 15/85, and 30/70 , respectively (Fig. 4, A and B), confirmed the selectivity suggested by the pure-component isotherms. In the N_2 profiles, a sharp front, indicating complete breakthrough of N_2, was observed at dimensionless time (ratio of experimental time to the time taken by a nonadsorbing tracer to travel through the column) $\bar{t}$~4 in all three cases. The "roll-up" effect of N_2, whereby the outlet composition of N_2 was higher than its inlet value, until CO_2 broke through is clearly

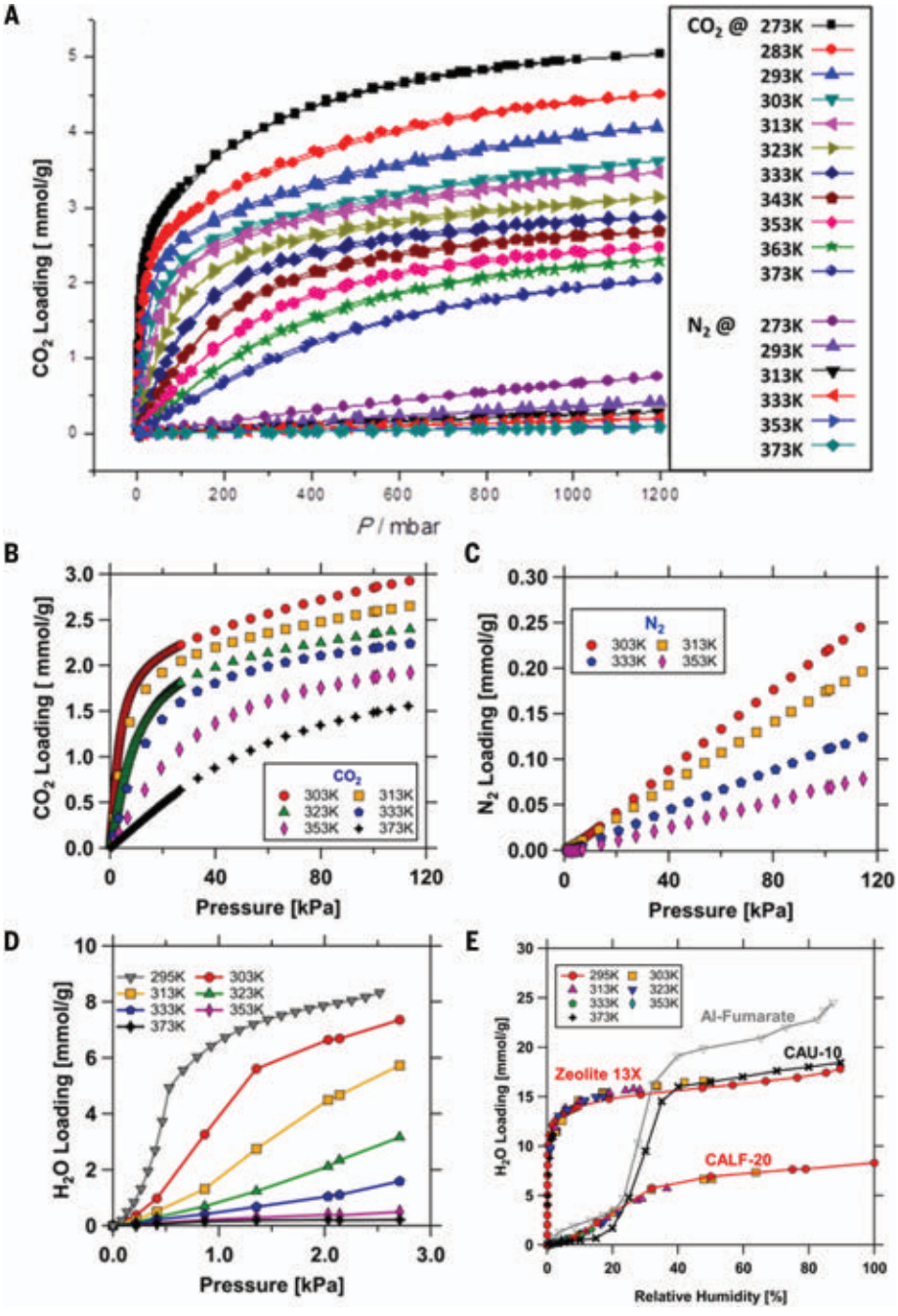


Fig. 2. Equilibrium gas uptake data on pure CALF-20. (**A**) CO_2 and N_2 isotherms from 273 to 373 K on pure CALF-20. (**B** to **D**) Structured CALF-20 (80% MOF:20% polysulfone). (B) CO_2 isotherms from 303 to 373 K. (C) N_2 isotherms from 303 to 353 K. (D) H_2O isotherms from 295 K to 373 K. (**E**) A comparison of H_2O isotherm on zeolite 13X (*37*), CAU-10 (*38*), Al fumarate (*39*) and structured CALF-20 at 295 K. The isotherms of CO_2 and N_2 were measured by volumetry, and that of H_2O was measured by gravimetry.

visible. The CO_2 concentration profiles showed different breakthrough times for various CO_2/N_2 compositions (Fig. 4A). Higher CO_2 composition in the feed led to shorter breakthrough times

We measured the competitive adsorption of CO_2 and H_2O using a combination of gravimetry, a study that measured the loading of CO_2 + H_2O by subjecting the sample to a moist stream of CO_2 whose RH was controlled, and a breakthrough experiment that provided the competitive loading of H_2O in the presence of CO_2 (Fig. 4, B and C). The difference between the total loading from the gravimetry and the H_2O loading from the breakthrough provided the competitive CO_2 loading. Up to a value of 30% RH, the CO_2 loading was nearly unaffected (Fig. 4F), which was unexpected for a physisorptive material but corroborated by the atomistic simulations. The CO_2 loading gradually decreased until it became negligible at RH > 80%. Additionally, the distinct shift in the H_2O isotherm in the presence of CO_2, compared to its pure-component isotherm, also confirmed the suppression of water sorption by CO_2.

To further demonstrate the physisorption of CO_2 by CALF-20 in wet environments, we measured the water breakthrough curves in air or CO_2 at two different RHs (Fig. 4, D and E). With the air experiment as a background, the water breakthrough was actually accelerated in CO_2, providing definitive support for the physisorptive preference of CALF-20 for CO_2 over water below 40% RH. The difference in water loading, exemplified by the area behind the breakthrough curve, between the two curves was pronounced. A comparison of both CO_2 and H_2O loading in competitive experiments (Fig. 4F and figs. S15 and S16) corroborated not only the sustained CO_2 capacity in wet gas but also the ability to suppress water sorption.

The nature of the water and CO_2 binding from the single-component water simulations and dry CO_2/N_2 simulations presented in Fig. 3 was consistent with the preferential binding of CO_2 over water observed at low RHs. Namely, CO_2 has strong binding sites in the center of the CALF-20 pores that precluded the formation of a hydrogen-bonded network that was responsible for the large uptake of water at high RHs. To corroborate this model, we performed multicomponent simulations of CO_2, N_2, and water at varying RH. Figure S10 shows the comparison of simulated water uptake at various RHs from a single-component water simulation to that of multicomponent simulations with 0.20 bar of CO_2 and 0.80 bar of N_2. The results were in good agreement with the experimental competitive isotherms shown in Fig. 4F. The simulations showed that without CO_2, water uptake at 20% RH is 6 mmol/g, whereas in the presence of CO_2, it was negligible. Only at 40% RH did water uptake reach 6 mmol/g when CO_2 was present. Calculated binding energies of most probable CO_2 and H_2O binding sites taken from the multicomponent $CO_2/N_2/H_2O$ simulations give −17.5 kJ mol^{-1} for H_2O and −33.5 kJ mol^{-1} for CO_2 (table S6). Calculated heats of adsorption, at zero loading and high loading (table S7), suggest that partially water-loaded pores were more attractive for subsequent water sorption than empty pores, and CO_2 had a stronger zero-loading heat of adsorption than water. A binding site analysis of the mixed $CO_2/N_2/H_2O$ simulations is presented in the supplementary materials.

The low water-affinity yet CO_2-phillic behavior of CALF-20 was enabled by its pore structure. Although a pore that is ideal for CO_2 is, of course, targeted in carbon capture, it is much less a focus that a pore be nonideal for water. Notably, a key feature of CALF-20 was the absence of any

strongly interacting functionality with CO_2. Although this property would be expected to moderate the affinity of the MOF for CO_2, the less specific dispersion interactions cumulatively compensate. As previously mentioned, dispersion interactions account for >85% of the binding energy in the most favorable CO_2 binding site. The boiling point of H_2O is 157°C higher than that of CO_2, so it is not expected that CO_2 would preferentially physisorb, but we can connect the competitive sorption in Fig. 4F with modeling in fig. S10. Interactions between guest molecules (*40*, *41*), or in this case, the lack thereof as CO_2 blocked cooperative H_2O binding, could tip subtle balances in binding enthalpies. The pore itself is the critical element in performing a sorptive function (*42*). Other MOFs with low water affinity such as CAU-10 (*38*) and Al fumarate (*39*) have been reported and studied for CO_2 capture from wet gas. These MOFs have good CO_2/N_2 selectivity and reasonably low water affinity, as indicated by stepped water isotherms. However, CAU-10 loses CO_2 capacity above a RH value of 20%, and aluminum fumarate loses 17% CO_2 capacity at 14% RH (fig. S16). However, CALF-20 has a higher CO_2 capacity and retains it up to and beyond 40% RH. Also, neither shows the suppression of water sorption by CO_2 that is observed with CALF-20.

Flue gas sorption and scaling

Industrially, materials must absorb CO_2 from postcombustion flue gases at 100°C containing water vapor and acid gases, and endure stresses during regeneration as the sorbent goes through a temperature swing, pressure swing, or vacuum swing process. CALF-20 has been run through stability assessments from multiple academic, government, and industry partners and shows robust performance, as confirmed by retention of structure and gas adsorption properties. The retention of CO_2 capacity after being repeatedly heated to dry air at 150°C in the thermal gravimetric analysis (Fig. 5A) showed excellent stability (*6*, *43*). This feature is key to high sorbent lifetime, as there is residual O_2 in the flue gas and during conditioning of the bed where air can oxidize reactive groups.

Powder x-ray diffraction (Fig. 5B) and N_2 sorption isotherms (Fig. 5C) are shown after a week of exposure to 150°C steam. CALF-20 was also tested for retention of structure and porosity (figs. S12 and S13) after treatment with 20 parts per million (ppm) SO_2 and 100 ppm NO_x at 20°C in separate experiments. We subjected CALF-20 to a real flue gas stream (50°C, flow of 100 $cm^3\ min^{-1}$) from natural gas combustion containing 7.3% H_2O, 7.1% O_2, 147 ppm CO, 78 ppm NO, and 13 ppm NO_2 (see supplementary materials, fig. S3, and table S4). Under these flowing flue gas conditions, powdered CALF-20 lost only 1.3% of its capacity after 6 days, as confirmed by gravimetric CO_2 uptake in a 15/85 mixture of CO_2/N_2. A process demonstration unit using the Svante VeloxoTherm process was built on the basis of rotating beds and fast cycles (~1 min) at 0.1 tonne per day CO_2 capacity, and it was deployed to test the CALF-20 lifetime with simulated cement flue gas. This simulated flue gas was generated by enriching real flue gas from a natural gas boiler with pure CO_2 and air to bring CO_2, water, and O_2 concentrations to cement kiln flue gas composition (17% CO_2, 10% O_2, 5% H_2O, balance N_2, at 45°C). The gas analyzer recorded around 60 ppm NO and 12 ppm NO_2 in the generated flue gas that was fed to the CALF-20 beds. The process was continuously tested for over 2000 hours with expected key performance indicators and no appreciable performance loss, as can be seen in fig. S17 (*44*). Furthermore, the process was able to achieve US Department of Energy target CO_2 purity of 95%.

For large-scale applications, it is important that the scale-up be feasible from an economic and technical viewpoint (*45*). Analysis of the cost driver for different MOF syntheses reveals that the costs of the raw materials, especially for the linker and less commonly for the metal, are often prohibitive. In addition, synthetic process conditions can have a substantial impact on the economics—for example, the necessity of high-pressure equipment is not only expensive but also results in costly safety precautions to protect employees and the environment. For CALF-20, none of these disadvantageous conditions apply (*46*). The raw materials are commercially available on a large scale from qualified vendors. Both linkers are low-cost bulk chemicals with large global production capacity (*47*): 200,000 metric tonnes per annum (MTPA) for oxalic acid, found mainly in pharmaceutical, textile and mining industries; 10,000 MTPA for triazoles, used mainly in the agricultural sector as a building block for azole-based fungicides. In addition, the reaction could be carried out in a water/methanol mixture, where the organic solvent represents <25 wt % with ongoing improvements. These conditions are particularly advantageous from a safety and environmental aspect. Further, in large-scale

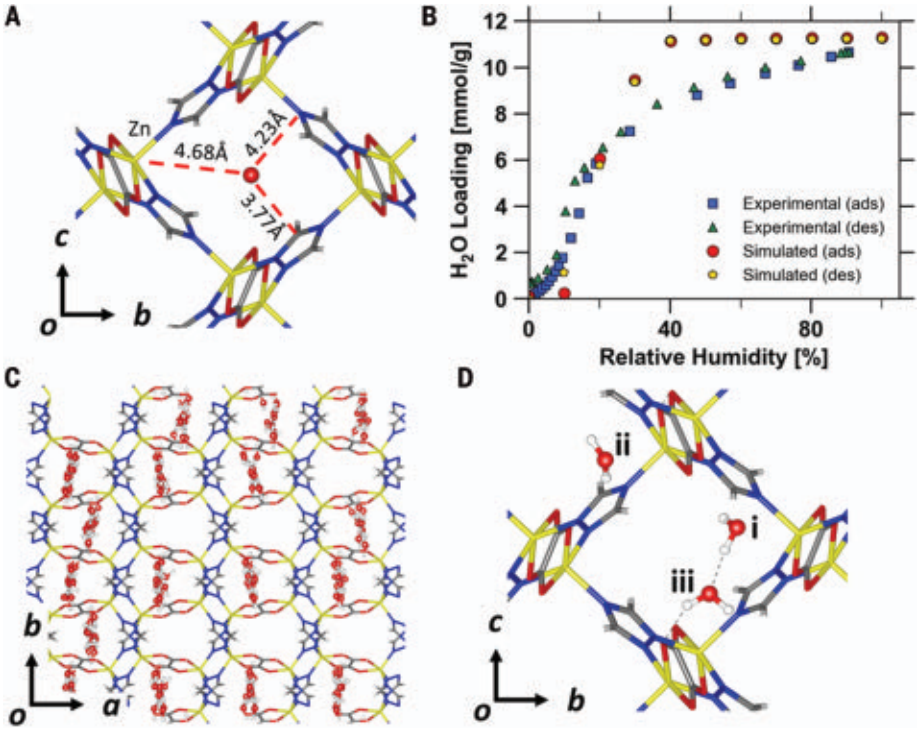


Fig. 3. Most probable CO_2 binding site determined from the single-component CO_2 GCMC simulation at 0.15 atm. (**A**) Select distances between heavy atoms of the CALF-20 framework and the CO_2 atoms are highlighted. These are the shortest distance between atoms of the framework and any atom of the CO_2 molecule. (**B**) Experimental and simulated single-component water isotherms at 293 K. Simulated adsorption results refer to the values obtained starting from empty pores, whereas the desorption results refer to the values obtained by starting the simulation with the pores saturated with water. (**C**) A snapshot from a 20% RH simulation of water in CALF-20. (**D**) The three most probable H_2O binding sites determined from the single-component water simulation at 20% RH. For (A), (C), and (D), ball-and-stick representations are used for the guest, whereas a tube representation is used for the framework. Atom colors are the same as shown in Fig. 1B.

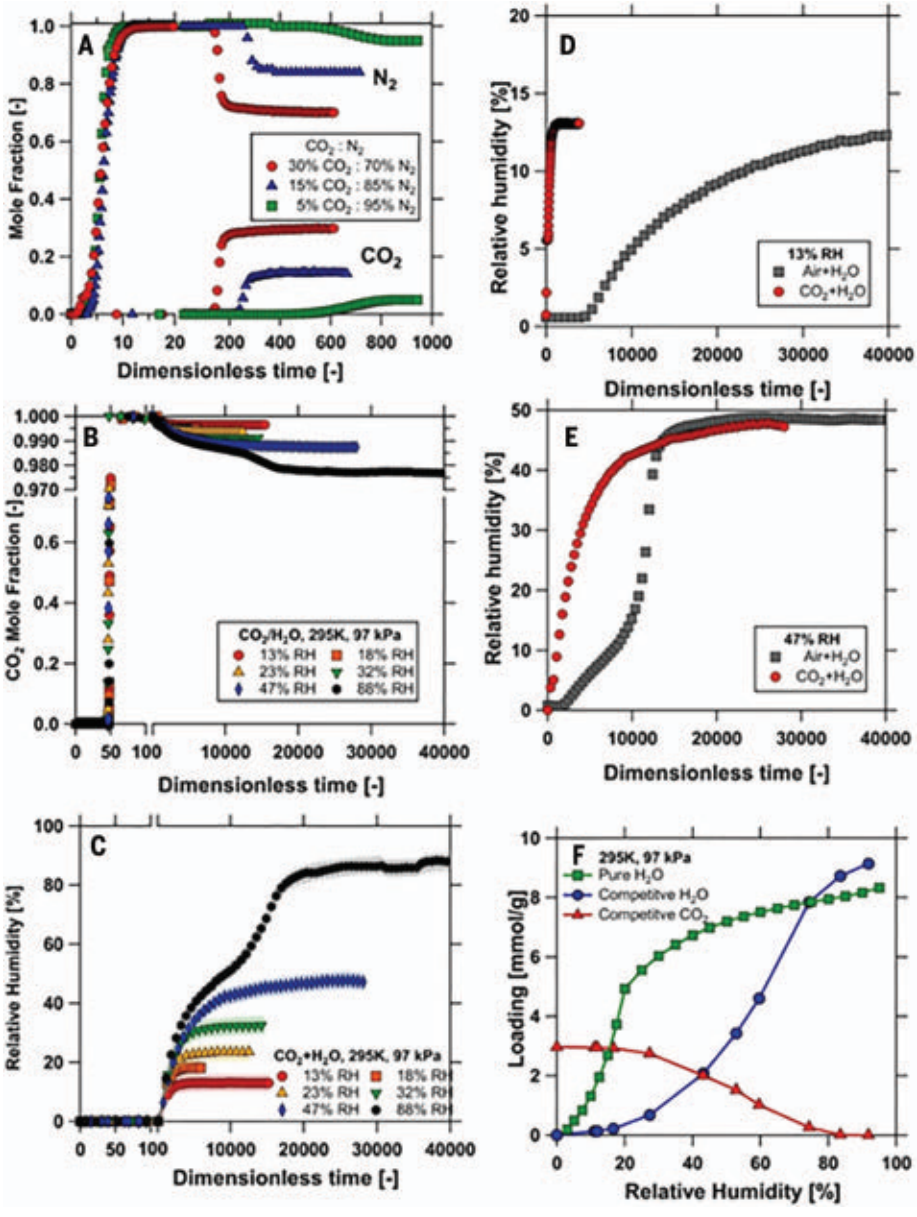


Fig. 4. Competitive dynamic column breakthrough (DCB) and equilibrium measurements on structured CALF-20 at 295 K and 97 kPa. (**A**) Competitive DCB of CO_2 and N_2 at different compositions. (**B**) Competitive CO_2 breakthrough curves measured at various RH values. (**C**) Competitive H_2O breakthrough curves at various RH values corresponding to the curves shown in (B). (**D**) A comparison of breakthrough curves obtained from experiments with Air + H_2O and that with CO_2 + H_2O at 13% RH. (**E**) A comparison of breakthrough curves obtained from experiments with Air + H_2O and that with CO_2 + H_2O at 47% RH. (**F**) Competitive CO_2 loadings (red triangle) and competitive H_2O loadings (blue circle) at various RH values. The loading of pure H_2O isotherm (green square) is shown as a reference. The breakthrough curves are plotted in dimensionless time, which is the ratio of the actual time to the average retention time taken of a nonadsorbed component. Also, there is a break in the abscissa of breakthrough curves.

batch synthesis, CALF-20 can obtain an unusually high solid content (total amount of dried MOF per total amount of solvents used) of >35%. The high yield of >90%, the reasonable reaction time, and the very high solid content result in an exceptional space-time yield (STY) for the precipitation step of 550 kg/m^3 day. In comparison, the STYs for zeolites are in the range of 50 to 150 kg/m^3 day (*48*). Critically, the CO_2 uptake of CALF-20 was retained through a wide range of scaling and structuring. Figure 5D shows a 3 million–fold difference in scale with matching CO_2 isotherms.

Outlook

An ideal adsorbent for the postcombustion CO_2 capture should exhibit several properties, including (i) high CO_2 adsorption capacity; (ii) fast adsorption/desorption kinetics; (iii) high CO_2 selectivity over N_2, O_2, and ability to function in wet gas; (iv) mild regeneration conditions; (v) the ability to be formed into structures, e.g., beads, laminates, or monoliths; (vi) chemical, mechanical, and thermal stability during adsorption-desorption cycling; and (vii) low cost and scalability of production. We have shown that CALF-20 can meet all of these criteria and help make industrial-scale CO_2 capture cost effective and reliable (*44*). Other MOFs have better reported properties in one or more of the aforementioned criteria, but not in all of them. For example, most reported MOFs cannot tolerate even ambient moisture or steam despite having very high CO_2 capacity or high CO_2/N_2 selectivity. The other important factor to consider is cost and scalability of synthesis. Most MOFs need aprotic solvents (such as dimethyl formamide or diethyl formamide) or contain expensive and noncommercial-grade organic linkers. With CALF-20, the components are commercially available at low cost and large volume, and water and methanol are the solvents used to synthesize this MOF.

In terms of gas separations, there is an increasing body of evidence showing that simple metrics such as selectivity and working capacity correlate poorly with ultimate process performance (*49*–*53*). A recent study that screened >5000 MOFs (*54*) showed that sorbent screening should include detailed process modeling and optimization. The Svante VeloxoTherm capture process used direct steam to rapidly desorb all the captured CO_2. In comparison to a traditional temperature swing process, the steam regeneration step of the VeloxoTherm process provided concentration swing in addition to heat, which allowed the extraction of the entire quantity of physisorbed CO_2, through its cyclic working capacity. Beyond steam stability, key aspects of the CALF-20 adsorbent synergizing with this process are its low water affinity and its ability to rapidly physisorb CO_2 in a wet gas, facilitating faster cycling, higher productivity, and ultimately resulting in a smaller plant footprint. In the Svante process with CALF-20, less energy is required to remove moisture in the drying cycle, and it also bears a higher moisture tolerance, which allows the capture cycle to recommence more rapidly.

Although materials can have one or more exceptional features, the key point is to merge those properties with process engineering conditions that best exploit them, such as capture conditions and available waste energy or heat for regeneration. The high uptakes at lower partial pressures of CO_2 make CALF-20

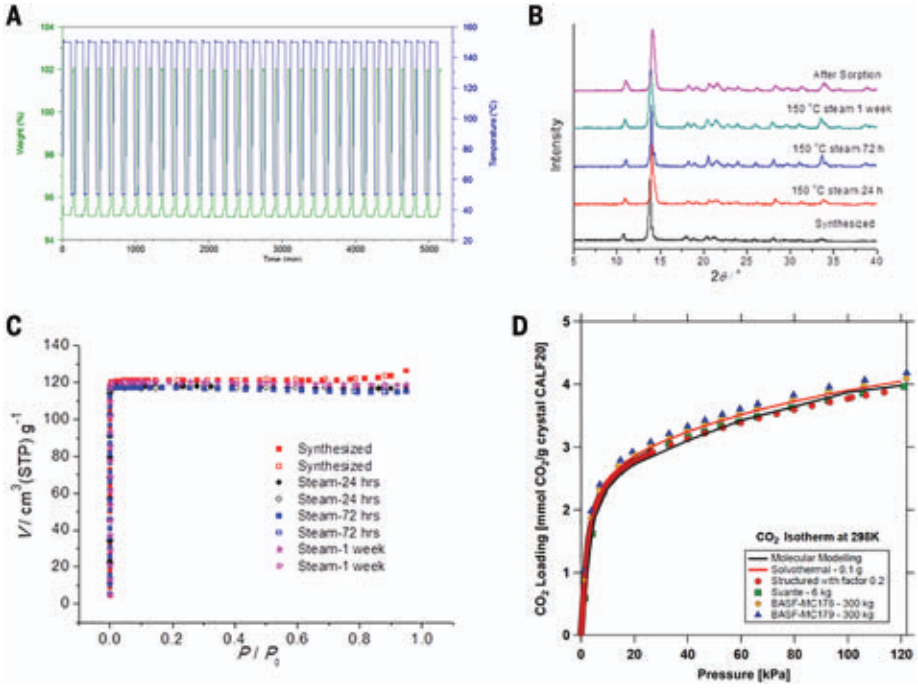


Fig. 5. CALF-20 scalability and stability. (**A**) Cycling of heating and introduction of CO_2 showing 30 cycles heated to 150°C. The left *y* axis is truncated to show the CO_2 mass gain on each cycle. CALF-20 survived more than 450,000 steam treatments in another test, but CO_2 uptake was only measured on the terminal sample. (**B**) Powder x-ray diffractograms and treatment with steam and running gas sorption shown in (**C**) N_2 isotherms at 77 K run on steam-treated samples and compared to pristine CALF-20. (**D**) CO_2 isotherms on 3 million–fold different scale batch preparations of CALF-20, showing retention of the CO_2 capacity. Comparisons with simulated uptake from the crystal structure and the structured CALF-20 scaled by a factor of 0.2 to account for 20% polysulfone are also shown.

a more suitable sorbent for a temperature or concentration swing process or potentially a pressure or vacuum swing at elevated temperatures. Independent of the regeneration process, the competitive nature of CO_2 and H_2O will enhance sorbent efficiency as coadsorption of water is reduced. Factoring its scalable preparation and durability, CALF-20 should derisk the use of MOFs for large-scale gas separation in industrial settings, and in particular, the challenge of postcombustion CO_2 capture (*55*, *56*). In terms of carbon capture and climate change, efficient capture is only a step, albeit a very important one, in reducing greenhouse gases. With the integration of better materials into advanced processes, this derisking should lead to additional and larger demonstration projects for critical testing of MOFs in CO_2 capture and other strategic challenges.

REFERENCES AND NOTES

1. "Accelerating Breakthrough Innovation in Carbon Capture, Utilization, and Storage" (U.S. Department of Energy, 2017).
2. Z. Hu, Y. Wang, B. B. Shah, D. Zhao, *Adv. Sustainable Syst.* **3**, 1800080 (2018).
3. K. Sumida *et al*., *Chem. Rev.* **112**, 724–781 (2012).
4. B. Dutcher, M. Fan, A. G. Russell, *ACS Appl. Mater. Interfaces* **7**, 2137–2148 (2015).
5. M. Wang, A. S. Joel, C. Ramshaw, D. Eimer, N. M. Musa, *Appl. Energy* **158**, 275–291 (2015).
6. C. Gouedard, D. Picqa, F. Launay, P.-L. Carrette, *Int. J. Greenh. Gas Control* **10**, 244–270 (2012).
7. M. Pardakhti *et al*., *ACS Appl. Mater. Interfaces* **11**, 34533–34559 (2019).
8. L. A. Darunte, K. S. Walton, D. S. Sholl, C. W. Jones, *Curr. Opin. Chem. Eng.* **12**, 82–90 (2016).
9. C. A. Trickett *et al*., *Nat. Rev. Mater.* **2**, 17045 (2017).
10. A. E. Creamer, B. Gao, *Environ. Sci. Technol.* **50**, 7276–7289 (2016).
11. J. C. Abanades *et al*., *Int. J. Greenh. Gas Control* **40**, 126–166 (2015).
12. D. G. Madden *et al*., *Philos. Trans. A, Math. Phys. Eng. Sci.* **375**, 20160025 (2017).
13. P. G. Boyd *et al*., *Nature* **576**, 253–256 (2019).
14. E. González-Zamora, I. A. Ibarra, *Mater. Chem. Front.* **1**, 1471–1484 (2017).
15. J. H. Cavka *et al*., *J. Am. Chem. Soc.* **130**, 13850–13851 (2008).
16. A. J. Howarth *et al*., *Nat. Rev. Mater.* **1**, 15018 (2016).
17. J. M. Kolle, M. Fayaz, A. Sayari, *Chem. Rev.* **121**, 7280–7345 (2021).
18. S. Yuan *et al*., *Adv. Mater.* **30**, e1704303 (2018).
19. J. Duan, W. Jin, S. Kitagawa, *Coord. Chem. Rev.* **332**, 48–74 (2017).
20. E. J. Kim *et al*., *Science* **369**, 392–396 (2020).
21. D. F. Sava Gallis, D. J. Vogel, G. A. Vincent, J. M. Rimsza, T. M. Nenoff, *ACS Appl. Mater. Interfaces* **11**, 43270–43277 (2019).
22. S. Bhattacharyya *et al*., *Chem. Mater.* **30**, 4089–4101 (2018).
23. G. W. Peterson, J. J. Mahle, J. B. DeCoste, W. O. Gordon, J. A. Rossin, *Angew. Chem. Int. Ed.* **55**, 6235–6238 (2016).
24. P. A. Julien, C. Mottillo, T. Friscic, *Green Chem.* **19**, 2729–2747 (2017).
25. S. Wang, C. Serre, *ACS Sustain. Chem.& Eng.* **7**, 11911–11927 (2019).
26. M. Bui *et al*., *Energy Environ. Sci.* **11**, 1062–1176 (2018).
27. R. B. Lin, D. Chen, Y. Y. Lin, J. P. Zhang, X. M. Chen, *Inorg. Chem.* **51**, 9950–9955 (2012).
28. Y. Y. Lin, Y. B. Zhang, J. P. Zhang, X. M. Chen, *Cryst. Growth Des.* **8**, 3673–3679 (2008).
29. X.-F. Wei, J. Miao, L.-L. Shi, *Synth. React. Inorg. Met.-Org. Nano-Met. Chem.* **46**, 365–369 (2016).
30. S. R. Caskey, A. G. Wong-Foy, A. J. Matzger, *J. Am. Chem. Soc.* **130**, 10870–10871 (2008).
31. J. An, S. J. Geib, N. L. Rosi, *J. Am. Chem. Soc.* **132**, 38–39 (2010).
32. P. Nugent *et al*., *Nature* **495**, 80–84 (2013).
33. T. M. McDonald *et al*., *Nature* **519**, 303–308 (2015).
34. P. M. Bhatt *et al*., *J. Am. Chem. Soc.* **138**, 9301–9307 (2016).
35. S. Nandi *et al*., *J. Am. Chem. Soc.* **139**, 1734–1737 (2017).
36. S. Xiang *et al*., *Nat. Commun.* **3**, 954 (2012).
37. N. S. Wilkins, J. A. Sawada, A. Rajendran, *Adsorption* **26**, 765–779 (2020).
38. V. B. López-Cervantes *et al*., *Polyhedron* **155**, 163–169 (2018).
39. J. A. Coelho *et al*., *Ind. Eng. Chem. Res.* **55**, 2134–2143 (2016).
40. K. S. Walton *et al*., *J. Am. Chem. Soc.* **130**, 406–407 (2008).
41. R. Vaidhyanathan *et al*., *Science* **330**, 650–653 (2010).
42. S. Kitagawa, R. Matsuda, *Coord. Chem. Rev.* **251**, 2490–2509 (2007).
43. M. Jahandar Lashaki, S. Khiavi, A. Sayari, *Chem. Soc. Rev.* **48**, 3320–3405 (2019).
44. P. Hovington, O. Ghaffari-Nik, L. Mariac, A. Liu, B. Henkel, S. Marx, Rapid Cycle Temperature Swing Adsorption Process Using Solid Structured Sorbent for CO_2 capture from Cement Flue Gas, Proceedings of the 15th Greenhouse Gas Control Technologies Conference, 15–18 March 2021.
45. M. Rubio-Martinez *et al*., *Chem. Soc. Rev.* **46**, 3453–3480 (2017).
46. J. M. Taylor, R. K. Mah, G. K. H. Shimizu, Synthesis of Zinc MOF Materials, PCT/CA2019/050530, filed 24 April 2019.
47. S. N. Bizzari, M. Blagoev, CEH Marketing Research Report, *Chemical Economics Handbook*, SRI Consulting, April 2010.
48. A. U. Czaja, N. Trukhan, U. Müller, *Chem. Soc. Rev.* **38**, 1284–1293 (2009).
49. A. K. Rajagopalan, A. M. Avila, A. Rajendran, *Int. J. Greenh. Gas Control* **46**, 76–85 (2016).
50. J. Park *et al*., *Ind. Eng. Chem. Res.* **59**, 7097–7108 (2020).
51. K. T. Leperi, Y. G. Chung, F. You, R. Q. Snurr, *ACS Sustain. Chem.& Eng.* **7**, 11529–11539 (2019).
52. M. Khurana, S. Farooq, *Ind. Eng. Chem. Res.* **55**, 2447–2460 (2016).
53. A. H. Farmahini, S. Krishnamurthy, D. Friedrich, S. Brandani, L. Sarkisov, *Chem. Rev.* **121**, 10666–10741 (2021).
54. T. D. Burns *et al*., *Environ. Sci. Technol.* **54**, 4536–4544 (2020).
55. A. Samanta, A. Zhao, G. K. H. Shimizu, P. Sarkar, R. Gupta, *Ind. Eng. Chem. Res.* **51**, 1438–1463 (2012).
56. R. L. Siegelman, E. J. Kim, J. R. Long, *Nat. Mater.* **20**, 1060–1072 (2021).

ACKNOWLEDGMENTS

Funding: This research was undertaken thanks in part to funding from Alberta Innovates Technology Futures (Strategic Research Grant), the Natural Sciences and Engineering Research Council (NSERC) of Canada (CREATE Grant), the US Department of Energy's (DOE) office of Fossil Energy (FE) DE- FOA-0001792, GreenSTEM from Alberta Jobs, Economy, and Innovation, Carbon Management Canada's Carbon Capture and Conversion Institute, MITACS, Innovate Calgary, the Canada First Research Excellence Fund (Global Research Initiative in Sustainable Low Carbon Unconventional Resources), and a Parex Innovation Fellowship to GKHS. We also thank Compute Canada for computing resources. **Author contributions:** Methodology and Investigation: J.-B.L., T.T.T.N., R.V., J.M.T., N.J.F., R.K.M., O.G.-N., S.S.I., K.W.D., P.S., S.M.; Formal Analysis: J.-B.L, T.T.T.N., J.B., H.D., O.G.-N., A.R., T.K.W., G.K.H.S.; Funding acquisition/Supervision/Project administration: A.R., T.K.W., P.H., G.K.H.S.; Writing: J.-B.L, T.T.T.N., A.R., T.K.W., and G.K.H.S. wrote the first draft. All authors contributed to the final draft. **Competing interests:** Two patents (CA2904546A1 and EP3784824A1) related to CALF-20 are licensed to Svante Inc. and ZoraMat Solutions Inc. for different fields of use. J.-B.L., R.V., R.K.M., J.M.T., S.S.I., K.W.D., and G.K.H.S. receive royalties from the license. T.T.T.N., H.D., J.B., F.A., S.M., N.J.F., P.S., A.R., T.K.W. have no competing interests. **Data and materials availability:** The CIF file for CALF-20 is available at the Cambridge Crystallographic Data Centre with deposition number CCDC 2084733. All other data, excepting the large-scale synthesis of CALF-20, which is patent-pending, are available in the manuscript or the supplementary materials. Samples of CALF-20 are available for data reproduction purposes from BASF/Svante under a material transfer agreement via P.H. (phovington@svanteinc.com).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi7281
Materials and Methods
Figs. S1 to S17
Tables S1 to S7
References (*57*–*74*)

26 March 2021; accepted 19 October 2021
10.1126/science.abi7281

SPIN CHEMISTRY

Readout of spin quantum beats in a charge-separated radical pair by pump-push spectroscopy

David Mims[1], Jonathan Herpich[1], Nikita N. Lukzen[2], Ulrich E. Steiner[3]*, Christoph Lambert[1,4]*

Spin quantum beats prove the quantum nature of reactions involving radical pairs, the key species of spin chemistry. However, such quantum beats remain hidden to transient absorption–based optical observation because the spin hardly affects the absorption properties of the radical pairs. We succeed in demonstrating such quantum beats in the photoinduced charge-separated state (CSS) of an electron donor–acceptor dyad by using two laser pulses—one for pumping the sample and another one, with variable delay, for further exciting the CSS to a higher electronic state, wherein ultrafast recombination to distinct, optically detectable products of singlet or triplet multiplicity occurs. This represents a spin quantum measurement of the spin state of the CSS at the time instant of the second (push) pulse.

Chemical reactions of radical pairs (RPs) are exceptional in that they can exhibit pronounced nonclassical kinetics, characterized by quantum beats of intermediate populations, as opposed to monotonic processes for elementary chemical reactions in classical kinetics. Such behavior of RPs is intimately related to what is known as spin chemistry (*1*, *2*), a field dealing with the interrelation between electron spin motion and chemical reactivity—a phenomenon that has been accounted for by the so-called radical pair mechanism (RPM). The RPM has been suggested independently by Kaptein and Oosterhoff (*3*) and by Closs (*4*) when explaining the occurrence of unusual nuclear spin polarization during certain chemical reactions involving RP intermediates by chemically induced dynamic nuclear polarization (CIDNP). Spin chemistry is the physical basis for many magnetic field–sensitive chemical reactions, including biophysical processes (*5*)—in particular, magnetoreception in migratory birds (*6*), where the RPM is supposed to function as a molecular mechanism for the magnetic compass sensor in the avian retina (*7*, *8*), representing a prominent example of quantum biology (*9*).

In RPs, the two electron spins may be aligned antiparallel to result in a spin singlet, with a total spin of zero, or parallel—i.e., a spin triplet with a total spin of 1, for which three distinct spatial orientations are possible. According to the Wigner-Witmer rules (*10*, *11*), electron spin is conserved during chemical reactions of the RP. Thus, RPs with singlet spin can only react to singlet products, and RPs with triplet spin can only react to triplet products. This rule represents the key principle of the RPM. Notably, the spin selectivity of a reaction has also been interpreted in terms of a quantum measurement of the spin state (*12*, *13*). As long as a RP stays together, either because of a solvent cage or by virtue of a chemical link (*14*), the electron spin state can change as a result of magnetic interactions—e.g., the hyperfine interaction with magnetic nuclei and/or the Zeeman interaction with external magnetic fields. The latter allows for direct control of the reaction behavior of the RP.

Because the change of electron spin under magnetic interactions is a genuine quantum process, it can exhibit quantum oscillations that, by virtue of the spin selectivity rule, may be transmitted to the chemical reaction kinetics. The first studies on the observation of quantum oscillations in the recombination of RPs came from the field of radioluminescence (*15*–*17*). In radioluminescence, the primary ionizing processes occur with the solvent, and charge and spin are transferred to fluorescent solvent molecules in secondary events. However, in most studies of spin chemistry, the RPs of interest are directly produced by photochemical reactions, such as photoinduced electron transfer or photocleavage of bonds. In such cases, experimental demonstrations of quantum oscillations in the kinetics of RPs are rather rare. To our knowledge, there is only one study (*18*) on time-resolved transient absorption (TA) experiments that shows the intrinsic kinetic quantum oscillations in RPs. This is because in many experiments, only the yields into the various reaction channels are recorded, and therefore spin oscillations are wiped out in a time-integrated type of observable. On the other hand, in time-resolved experiments of RP recombination kinetics, the spin motion is hidden because the recombination kinetics is usually too slow to follow the spin motion. There are some examples showing that spin motion in RPs can be affected in a time-resolved manner by applying resonant microwave pulses [reaction yield–detected magnetic resonance (RYDMR)] (*19*–*21*) or by pulsed switching of steady magnetic fields. Another technique, whereby quantum beats in RPs have been shown by Kothe *et al.* (*22*), is time-resolved electron paramagnetic resonance (EPR). With time-resolved EPR, the combined effect of an external magnetic field and resonant microwave radiation turns initial singlet-triplet coherences into oscillating magnetization.

In this paper, we report a method to follow the intrinsic, unperturbed motion of the RP spin system. The proposed method can be characterized as an optical readout technique with a pump pulse for the photochemical creation of a charge-separated state (CSS) type of RP and a second, so-called push pulse (*23*) of variable delay probing the spin state. However, in contrast to the familiar pump-probe method, where the probe pulse is of much weaker intensity than the pump pulse and does not appreciably change the state populations of the system, the second pulse in our method excites a considerable fraction of the populations to a higher excited state. Thus, a fast quantum measurement of the spin state takes place by virtually prompt irreversible recombination into the singlet or triplet product channel or by nonradiative deactivation to the initial CSS-RP state. The method exploits the fact that the rate of charge recombination (CR) is drastically enhanced by electronic excitation of the CSS. There have been several examples of this observation in the literature (*24*–*27*); however, they have been restricted to a single push pulse at a fixed delay time. Furthermore, spin aspects have only been considered in a paper from the Wasielewski group (*28*), where the second laser pulse induced charge transport to a nearby acceptor, and the conservation of zero quantum coherence was proved by EPR.

In our method, the delay time of the push pulse is systematically varied, and therefore information on the intrinsic spin evolution can be collected in a quasi-continuous manner as a function of time. Thus, by investigating the photoinduced CSS state of a molecular dyad TAA-An-PDI—consisting of a triarylamine electron donor (TAA) and a perylene diimide acceptor (PDI) linked by a dihydroanthracene (An) bridge—we succeeded in directly tracking the quantum oscillations between singlet and triplet spin states by an optical absorption method.

Pump-push TA experiments

In a laser flash TA experiment, excitation of the TAA-An-PDI dyad (Fig. 1A) in anisole solution at 18,800 cm^{-1} with 7-ns laser pulses populated the PDI singlet state, which induced charge separation to yield the CSS. In the CSS, the PDI was reduced, and the TAA was oxidized ($TAA^{\cdot+}$-An-$PDI^{\cdot-}$) (Fig. 1C), as proved by the characteristic excited state absorption (ESA)

[1]Institute of Organic Chemistry, University of Würzburg, 97074 Würzburg, Germany. [2]International Tomography Center and Novosibirsk State Universit, Novosibirsk 630090, Russia. [3]Department of Chemistry, University of Konstanz, 78464 Konstanz, Germany. [4]Center for Nanosystems Chemistry, University of Würzburg, 97074 Würzburg, Germany.
*Corresponding author. Email: ulrich.steiner@uni-konstanz.de (U.E.S.); christoph.lambert@uni-wuerzburg.de (C.L.)

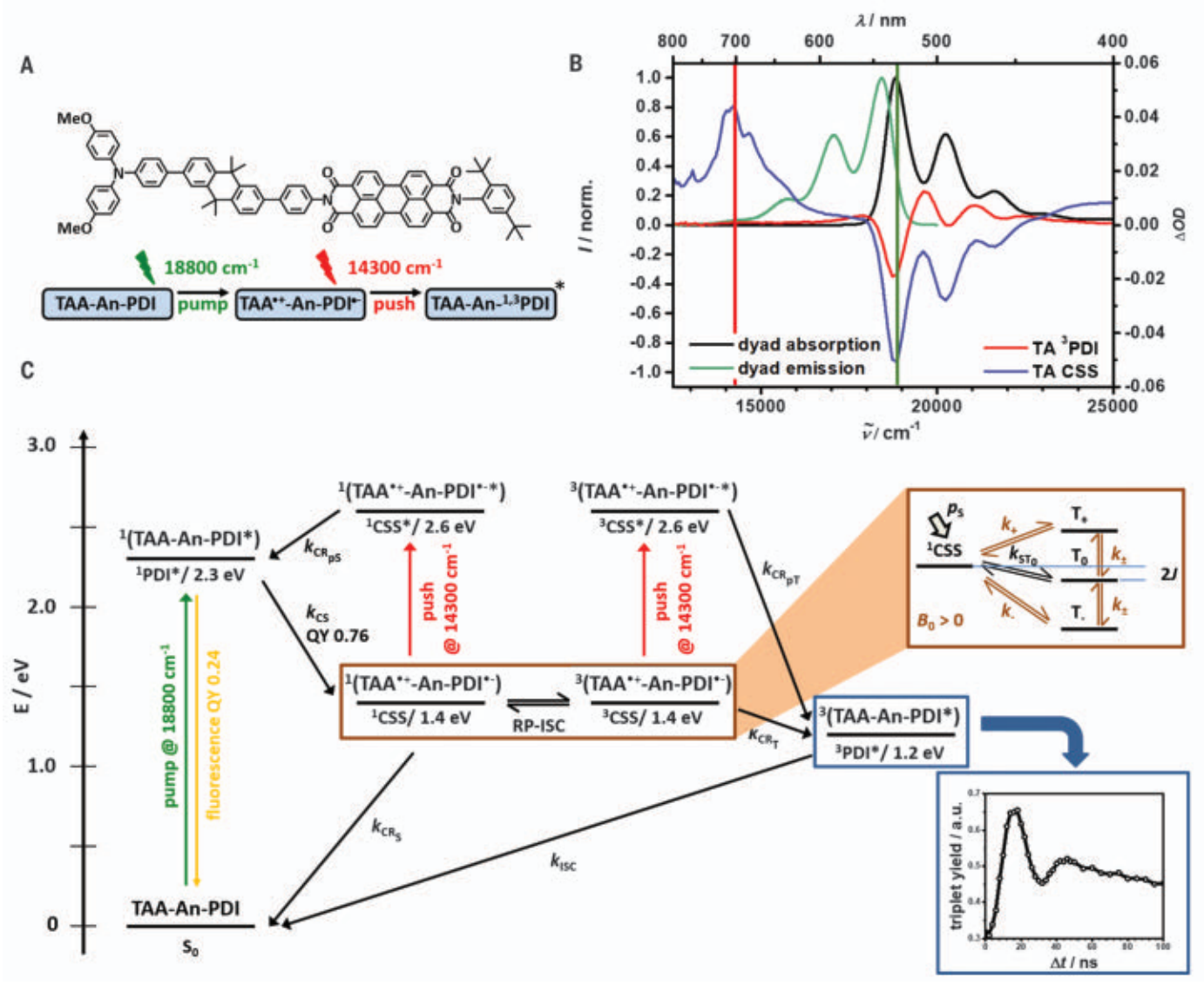


Fig. 1. Scheme of pump-push experiment. (**A**) Molecular structure of dyad TAA-An-PDI. (**B**) Absorption spectrum (black), emission spectrum (green), TA spectrum at 40 ns (blue), TA spectrum at 1000 ns (red), and spectral position of push pulse (red vertical line) corresponding to a relative excitation ratio of 4 to 1 favoring the radical anion over the cation. (**C**) Energy-level scheme with state energies and photophysical processes, including their rate constants. The processes depicted in the brown box show transitions between the spin levels under Zeeman splitting of the triplet CSS. Magnetic field–dependent processes are marked in brown. The inset in the blue frame shows the final ^{3}PDI yield (at B = 500 mT), reflecting the ^{3}CSS amplitude at the instant of the delayed push pulse. a.u., arbitrary units.

signals for the TAA radical cation and PDI radical anion between 13,000 and 16,000 cm^{-1} along with the ground state bleaching (GSB) between 18,000 and 23,000 cm^{-1} (blue spectrum in Fig. 1B). The formation of the CSS [quantum yield (QY) = 0.76] caused a substantial quenching of the ^{1}PDI* fluorescence (QY = 0.24; green spectrum in Fig. 1B).

The CSS can be conceived as a spin-correlated RP that was born in the singlet state as it derived from the singlet ^{1}PDI*. As outlined above, this ^{1}CSS underwent a time-dependent spin interconversion with the almost-degenerate ^{3}CSS. The tiny energy difference between both CSS states is indicated by the small value of ~1 mT of the resonance field B_{max} (figs. S5 to S7), which is usually assigned to twice the exchange energy J. The decay of the TA spectra was multi-exponential on a time scale of ~1 μs (fig. S4), which is a consequence of the complex decay kinetics. Although the ^{1}CSS could directly undergo CR to the ground state S_0, the ^{3}CSS decayed to a local ^{3}PDI* triplet state, whose energy (1.2 eV) was below that of the CSS (1.4 eV; for the evaluation of the state energies, see supplementary text, section I). At later times (>1.5 μs), where all CSS has decayed, only the broad ^{3}PDI* ESA between 17,000 and 24,000 cm^{-1} was visible (red spectrum in Fig. 1B), again superimposed by GSB. The ^{3}PDI* then slowly decayed to S_0 by intersystem crossing (ISC). This decay scenario has also been observed and analyzed by us in detail for analogous dyads with the same donor and acceptor moieties but other bridging units (*29*). In these dyads, we also demonstrated that the spin interconversion was magnetic field dependent, which influenced the overall decay kinetics of the CSS substantially. On the other hand, analysis of the magnetic field effect allowed us to extract all relevant kinetic parameters in detail. Notably, ^{1}CSS and ^{3}CSS, although differing in their reactivity, are completely identical in their optical properties.

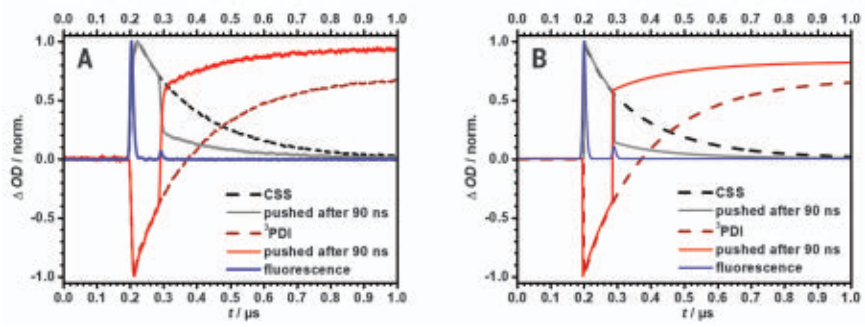


Fig. 2. Transient traces of the CSS, ^{3}PDI, and fluorescence in the pump-push experiment. (**A**) Time traces of TA signals of the CSS (at 14,100 cm^{-1}) and of the ^{3}PDI* (at 19,600 cm^{-1}) without push pulse (dashed lines in black and red) and with push pulse at Δt = 90 ns (solid lines in gray and red). Fluorescence signals (at 17,200 cm^{-1}) are given in blue. They appear synchronously with the pump and push pulses, respectively. At t = 0, the fluorescence and CSS signals are normalized to 1, and the ^{3}PDI/GSB signal is normalized to −1. (**B**) Simulation of the signals by the quantum theoretical model, assuming delta-shaped pump and push pulses and parameter values p_{SS} = 0.5 and p_{TT} = 0.9 (see the text).

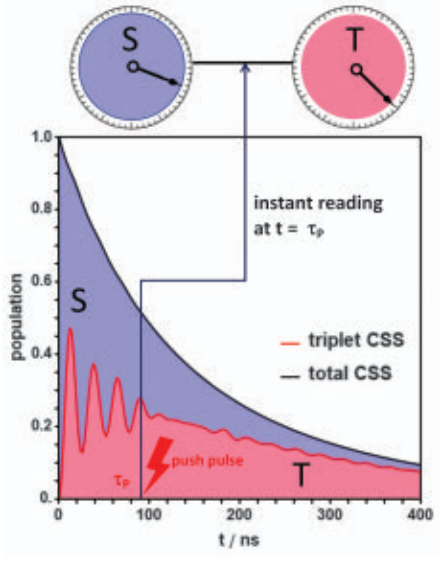


Fig. 3. Schematic of push pulse triggered quantum measurement. Although the total population of the CSS can be followed by optical TA, the singlet and triplet subpopulations are not immediately apparent but can be read out through the signals elicited by the push pulse, as shown in Fig. 2.

Thus, so far, the S/T composition of the CSS at any instant of time has not been easily accessed experimentally.

In this work, we have applied a second strong laser pulse of 14,300 cm^{-1}, time-delayed relative to the pump laser pulse of 18,800 cm^{-1}. This, so-called push pulse excited ^{1}CSS and ^{3}CSS with equal probability, creating a CSS* state with a ^{1}CSS*/ ^{3}CSS* ratio exactly reproducing their ratio in the CSS. Excited doublet states of radical ions are known to be much stronger electron donors and acceptors than their ground states and exhibit typical lifetimes of a few hundred picoseconds (*30*). Thus, CR from the excited CSS* state is extremely fast—i.e., essentially immediate on the time scale of normal CSS recombination. Recombination of ^{3}CSS*, just like the unexcited ^{3}CSS, yielded the locally excited ^{3}PDI; whereas recombination of ^{1}CSS*, unlike ^{1}CSS, which is too low in energy, mainly yielded the excited ^{1}PDI* state. The excited ^{1}PDI* formed this way decayed nonradiatively, by fluorescence, and by renewed charge separation to CSS. Thus, a push pulse induced three immediate responses: a jump in CSS population, a jump in ^{3}PDI population, and a delayed fluorescence signal. This situation is exemplified experimentally in Fig. 2A for a push pulse at 90-ns delay time.

Quantum dynamical simulation

In Fig. 2B, we show a simulation based on a quantum theoretical kinetic model. For dyads of the present type, involving the TAA donor and PDI acceptor, the push-free kinetics of CR comprising the channels to singlet ground state from ^{1}CSS and to ^{3}PDI* from ^{3}CSS has been successfully simulated in our previous work (*29*, *31*–*33*). Briefly, the reaction kinetics involving quantum mechanical spin motion was treated by a stochastic Liouville equation for the CSS electron–nuclear spin density matrix $\rho(t)$ of the general form

$$\dot{\rho}(t) = -i[H, \rho] + \hat{K}\rho + \hat{R}\rho \qquad (1)$$

where H represents the spin Hamiltonian accounting for Zeeman interaction, exchange interaction, and isotropic hyperfine interaction; $\hat{K}$ represents the spin-selective reaction superoperator; and $\hat{R}$ represents the relaxation superoperator (*34*), accounting for the effect of rotational modulation of anisotropic hyperfine interaction. For the details of parametrization, refer to supplementary text, section IV, and (*29*). Specific parameter values in the present case refer to the reaction rate constants of singlet recombination ($k_S = 9.5 \times 10^6$ s^{-1}) and triplet recombination ($k_T = 0.35 \times 10^6$ s^{-1}).

In modeling the push process, it was assumed that the push laser pulse excited all CSS population to CSS*, keeping the ratio of S/T unchanged. Recombination from ^{1}CSS* occurred promptly with a probability of p_{SS} of fluorescence plus nonradiative decay to the ground state and a probability of p_{TT} for recombining from ^{3}CSS* to ^{3}PDI*. The complementary parts, namely $(1 - p_{SS})$ of ^{1}CSS* and $(1 - p_{TT})$ of ^{3}CSS*, were assumed to return to the CSS with no other decoherence than that induced by the spin-selective recombinations in CSS*, as described by the Haberkorn (*35*) type of reaction operator. Thus, the spin density matrix after returning from CSS* was assumed to be given by

$$\rho_{postp} = \rho_{prep} - \frac{1}{2}p_{SS}\left(Q_S\rho_{prep} + \rho_{prep}Q_S\right) - \frac{1}{2}p_{TT}\left(Q_T\rho_{prep} + \rho_{prep}Q_T\right) \qquad (2)$$

where ρ_{postp} is the density matrix immediately after the push pulse, ρ_{prep} is the density matrix immediately before the push pulse, and Q_S and Q_T are the projection operators onto the singlet and triplet manifolds, respectively. Comparing Fig. 2, A and B, shows that the model described the essential features of the processes well.

From a general, theoretical point of view, the push pulses can be conceived as triggering a prompt Neumann-Wigner–type quantum measurement of the spin state of the RP (Fig. 3). By exciting the CSS to CSS*, the recombination rates to singlet or triplet products became so fast that the quantum state reductions to pure singlet or triplet were practically instantaneous on the time scale of normal CSS recombination kinetics or spin dynamics. It should be kept in mind, however, that this type of measurement differs from the notion of quantum measurements applied in the literature to spin-selective reactions in undisturbed, normal RP kinetics (*12*, *36*). In those cases, the decoherence effect of the quantum measurement on the surviving RPs has been the focus of interest (*37*).

Quantum beats and magnetic field dependence

The potential of the pump-push technique for revealing the details of spin dynamics underlying the recombination of the CSS is demonstrated in Fig. 4. Figure 4A shows a series of TA pump-push signals taken at short intervals of delay of the push pulse versus the pump pulse. Clearly, the amplitudes of the signal

jump after the push pulse exhibited oscillatory behavior—i.e., spin quantum beats. The observed pump-push signals of the triplet product and their magnetic field dependence were simulated quantum theoretically as mentioned above (Fig. 4C). To obtain a realistic representation of the observed signals, the finite width [full width at half maximum (FWHM); 9.5 ns total extension with fringes up to 40 ns; fig. S11] of the laser pulses has been considered. The width of the pump pulse was considered as well. In the region of overlap of pump and push pulse, the initial bleaching by the pump pulse and the rise partially compensated each other, making it difficult to separate the contribution of the positive jump because of the push pulse within the first ~20 ns. However, this situation can be treated with a theoretical simulation of the pump-push signals, such as shown in Fig. 4C (for more examples at higher fields, see fig. S17).

Electron spin motion is determined by internal and external magnetic fields. To extract the essential information about the spin quantum beats, we plotted the signal intensities at the end of each push pulse as a function of the delay time in Fig. 4B. In addition to zero field, the behavior at fields of 1, 7, and 500 mT—representing the cases of level crossing, middle field, and high field—has also been determined (see fig. S17 for the signal representations). The pump-push time profiles obtained, although somewhat noisy, directly revealed the oscillatory structure of spin motion in the first ~50 ns. In Fig. 4D, the theoretically simulated pump-push time profiles are shown. The positions of the oscillation maxima and minima as well as the general absolute scaling of the effects were well reproduced, although for 1 mT, the amplitude of the first maximum is obtained higher than in experiments. In fitting the theoretical results, it was found that the sign of the exchange interaction J caused a significant difference in the field-dependent pump-push profiles at low fields. Without carrying out the fitting of all parameters in the quantum calculation to a final optimum—because in this work we wanted to focus on a proof of principle of the pump-push spectroscopy—it can be concluded that only a positive exchange interaction (i.e., ^{3}CSS below ^{1}CSS) led to satisfactory results (figs. S18 to S20). For small absolute values of the exchange energy parameter J on the order of the intrinsic hyperfine couplings of the system, the usual rule of a magnetic field effect resonance at a field of $B_{max} = |2J|$ is no longer valid. A diagram of the resonance field B_{res} versus the value of J for positive and negative values is not symmetric around $J = 0$, where the so-called low-field effect is dominated by the hyperfine coupling constants. So far, the sign of J could only be determined for spin-correlated RPs by EPR (*38*–*41*) or by field-dependent

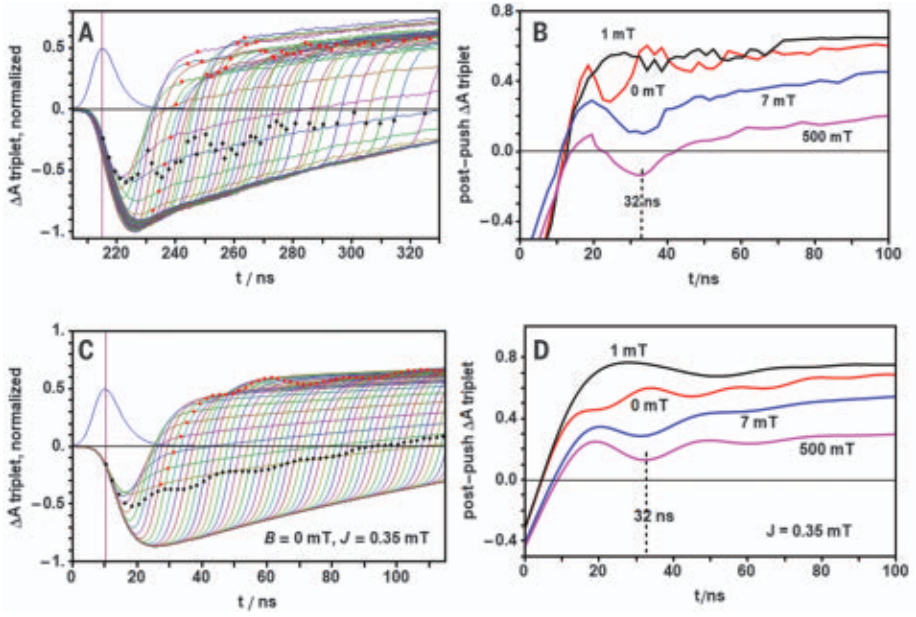


Fig. 4. Measured and simulated spin quantum beats. (**A**) Experimental pump-push TA signals taken at 19,600 cm^{-1} for the formation of ^{3}PDI* product at 0 mT. (**B**) Experimental postpush signal profiles at magnetic fields of 0, 1, 7, and 500 mT. (**C**) Theoretical pump-push signals at zero field, calculated with J = 0.35 mT. (**D**) Theoretical results simulating the curves in (B); for details, see the text. The vertical line in (A) marks the position of the maximum of the pump pulse. The black dots in (A) and (C) represent the positions of the maxima of the push pulses for the pertinent signals, and the red points represent the positions of the ends of the push pulses, taken 17 ns later. The end of that push pulse with the same maximum as the pump pulse is taken as time zero in (B) and (D).

CIDNP experiments (*42*). The pump-push spectroscopic method, which has also been exploited for the fluorescence signal yielding complementary results (figs. S13 to S16), offers a new approach with a much better time resolution.

Discussion and conclusions

In this paper, we have established pump-push spectroscopy as an optical technique for real-time observation of electron spin motion in CSSs. Although the spin motion in such states is usually hidden for optical probing absorption techniques, the proposed method allows us to look behind the scenes. This realization was achieved through push pulses triggering quasi-instantaneous spin measurements by two spin-selective CR channels exhibiting drastically enhanced rates in the electronically excited CSS. This method has the potential to fully map the coherent and incoherent spin dynamics in a CSS, including eventual quantum beats, as in the present case.

As has been shown in the work of Molin and co-workers (*16*), the best conditions for quantum beats prevail if radicals have either (i) simple hyperfine structure (i.e., small numbers of preferably equivalent nuclear spins, as is the case for the TAA-An-PDI dyad studied here) or (ii) small hyperfine coupling (e.g., because of perdeuteration) and large Δg in combination with strong external magnetic fields. These rules apply for any method of observation. However, Molin's method requires nonpolar solvents, high energy excitation, and efficient fluorophores with short fluorescence lifetimes. The pump-push method with absorption detection introduced here requires neither of these conditions and therefore is more generally applicable. However, our method is best suited for RPs with fixed distance. In flexibly linked or freely diffusing RPs, the fluctuating distance modulates the recombination efficiency after the push pulse to a large extent, which will probably attenuate the push pulse effect substantially.

Compared with the detection of quantum beats by transient EPR, our method provides the advantage that it can be applied at any magnetic field, including the notable J-resonant fields. The EPR technique requires microwaves and a magnetic field resonant with the microwaves, so only a discrete set of fields is accessible for which microwave frequency bands are available.

As far as the time resolution of the pump-push method is concerned, the laser pulse length and the recombination time in the CSS* state could be limiting. In the present case, the recombination time is much shorter than the laser pulse because the time derivative of the signal jumps induced by the push

pulses closely matches the laser profile. Here, the time resolution was limited by the ~10-ns width of the laser pulses, which already led to considerable damping of the quantum beats. Shorter pulse widths of 0.1 to 1 ns would allow better resolution and would be desirable to follow faster oscillations occurring, for example, because of higher values of J or because of fast Zeeman mixing at large Δg values and higher fields.

The pump-push technique reported here represents an extension of the arsenal of spin chemical techniques. It can be also expected to open applications for molecular electronics and quantum information science.

REFERENCES AND NOTES

1. U. E. Steiner, T. Ulrich, *Chem. Rev.* **89**, 51–147 (1989).
2. B. Brocklehurst, *Chem. Soc. Rev.* **31**, 301–311 (2002).
3. R. Kaptein, J. L. Oosterhoff, *Chem. Phys. Lett.* **4**, 195–197 (1969).
4. G. L. Closs, *J. Am. Chem. Soc.* **91**, 4552–4554 (1969).
5. Y. Kim *et al*., *Quantum Rep.* **3**, 80–126 (2021).
6. T. Ritz *et al*., *Biophys. J.* **96**, 3451–3457 (2009).
7. K. Schulten, C. E. Swenberg, A. Weller, *Z. Physik. Chem.* **111**, 1–5 (1978).
8. J. Xu *et al*., *Nature* **594**, 535–540 (2021).
9. M. Tiersch, H. J. Briegel, *Phil. Trans. R. Soc. A* **370**, 4517–4540 (2012).
10. E. Wigner, E. E. Witmer, *Eur. Phys. J. A* **51**, 859–886 (1928).
11. Y.-N. Chiu, M. Gong, *Chem. Phys.* **145**, 397–412 (1990).
12. J. A. Jones, P. J. Hore, *Chem. Phys. Lett.* **488**, 90–93 (2010).
13. I. K. Kominis, *Phys. Rev. E* **83**, 056118 (2011).
14. H. M. Hoang, V. T. B. Pham, G. Grampp, D. R. Kattnig, *ACS Omega* **3**, 10296–10305 (2018).
15. V. A. Bagryansky, V. I. Borovkov, Y. N. Molin, *Russ. Chem. Rev.* **76**, 493–506 (2007).
16. O. A. Anisimov, V. L. Bizyaev, N. N. Lukzen, V. M. Grigoryants, Y. N. Molin, *Chem. Phys. Lett.* **101**, 131–135 (1983).
17. D. V. Stass, B. M. Tadjikov, Y. N. Molin, *Chem. Phys. Lett.* **235**, 511–516 (1995).
18. P. Gilch, F. Pöllinger-Dammer, C. Musewald, M. E. Michel-Beyerle, U. E. Steiner, *Science* **281**, 982–984 (1998).
19. M. R. Wasielewski, C. H. Bock, M. K. Bowman, J. R. Norris, *Nature* **303**, 520–522 (1983).
20. Y. Sakaguchi, A. W. Astashkin, B. M. Tadjikov, *Chem. Phys. Lett.* **280**, 481–488 (1997).
21. H. Murai, *J. Photochem. Photobiol. C Photochem. Rev.* **3**, 183–201 (2003).
22. G. Kothe, S. Weber, E. Ohmes, M. C. Thurnauer, J. R. Norris, *J. Am. Chem. Soc.* **116**, 7729–7734 (1994).
23. G. M. Paternò *et al*., *Adv. Funct. Mater.* **29**, 1805249 (2019).
24. R. T. Hayes, M. R. Wasielewski, D. Gosztola, *J. Am. Chem. Soc.* **122**, 5563–5567 (2000).
25. R. T. Hayes, C. J. Walsh, M. R. Wasielewski, *J. Phys. Chem. A* **108**, 3253–3260 (2004).
26. L. Hammarström, *Acc. Chem. Res.* **48**, 840–850 (2015).
27. S. Neumann, C. Kerzig, O. S. Wenger, *Chem. Sci.* **10**, 5624–5633 (2019).
28. L. Kobr *et al*., *J. Am. Chem. Soc.* **134**, 12430–12433 (2012).
29. D. Mims *et al*., *J. Chem. Phys.* **151**, 244308 (2019).
30. D. J. Gosztola, M. P. Niemczyk, W. A. Svec, A. S. Lukas, M. R. Wasielewski, *J. Phys. Chem. A* **104**, 6545–6551 (2000).
31. N. N. Lukzen, J. H. Klein, C. Lambert, U. E. Steiner, *Z. Physik. Chem.* **231**, 197–223 (2017).
32. J. Schäfer, M. Holzapfel, A. Schmiedel, U. E. Steiner, C. Lambert, *Phys. Chem. Chem. Phys.* **20**, 27093–27104 (2018).
33. U. E. Steiner, J. Schäfer, N. N. Lukzen, C. Lambert, *J. Phys. Chem. C* **122**, 11701–11708 (2018).
34. T. P. Fay, L. P. Lindoy, D. E. Manolopoulos, *J. Chem. Phys.* **151**, 154117 (2019).
35. R. Haberkorn, *Mol. Phys.* **32**, 1491–1493 (1976).
36. I. K. Kominis, *Phys. Rev. E* **80**, 056115 (2009).
37. K. Maeda *et al*., *Chem. Commun.* **47**, 6563–6565 (2011).
38. C. D. Buckley, C. N. Hunter, P. J. Hore, K. A. McLauchlan, *Chem. Phys. Lett.* **135**, 307–312 (1987).
39. G. L. Closs, M. D. E. Forbes, J. R. Norris, *J. Phys. Chem.* **91**, 3592–3599 (1987).
40. P. J. Hore, D. A. Hunter, C. D. McKie, A. J. Hoff, *Chem. Phys. Lett.* **137**, 495–500 (1987).
41. M. D. E. Forbes, S. R. Ruberu, *J. Phys. Chem.* **97**, 13223–13233 (1993).
42. I. Zhukov *et al*., *J. Chem. Phys.* **152**, 014203 (2020).
43. C. Lambert, D. Mims, J. Herpich, N. Lukzen, U. Steiner, Data for spin quantum beats in pump push spectroscopy, *Dryad* (2021); https://doi.org/10.5061/dryad.4mw6m90b1.

ACKNOWLEDGMENTS

This work is dedicated to the memory of our late colleague Konstantin Ivanov, who died 5 March 2021 of COVID-19 at the age of 44. C.L. and D.M. thank A. Schmiedel for assisting with the pump-push setup. **Funding:** This work was supported by Deutsche Forschungsgemeinschaft grant LA991/20-1 (C.L. and D.M.); the SolTech Initiative of the Bavarian State Ministry of Education, Culture, Science, and the Arts; and Ministry of Science and Higher Education of the Russian Federation grant 075-15-2020-779 (N.N.L.). **Author contributions:** Conceptualization: C.L., U.E.S., and D.M. Formal analysis: U.E.S. and N.N.L. Funding acquisition: C.L. and N.N.L. Investigation: D.M., J.H., and U.E.S. Methodology: D.M. and U.E.S. Project administration: C.L. Software: N.N.L. and U.E.S. Supervision: C.L. and U.E.S. Writing – original draft: D.M. and U.E.S. Writing – review and editing: C.L., D.M., U.E.S., and N.N.L. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All experimental data shown in the main text or the supplementary materials are accessible at the Dryad repository (*43*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4254
Materials and Methods
Supplementary Text
Figs. S1 to S31
Table S1
Synthesis Protocols
References (*44*–*61*)

14 July 2021; accepted 19 October 2021
10.1126/science.abl4254

REPORTS

QUANTUM SIMULATION

Many-body–localized discrete time crystal with a programmable spin-based quantum simulator

J. Randall[1,2]†, C. E. Bradley[1,2]†, F. V. van der Gronden[1,2], A. Galicia[1,2], M. H. Abobeih[1,2], M. Markham[3], D. J. Twitchen[3], F. Machado[4,5], N. Y. Yao[4,5], T. H. Taminiau[1,2]*

The discrete time crystal (DTC) is a nonequilibrium phase of matter that spontaneously breaks time-translation symmetry. Disorder-induced many-body localization can stabilize the DTC phase by breaking ergodicity and preventing thermalization. Here, we observe the hallmark signatures of the many-body–localized DTC using a quantum simulation platform based on individually controllable carbon-13 nuclear spins in diamond. We demonstrate long-lived period-doubled oscillations and confirm that they are robust for generic initial states, thus showing the characteristic time-crystalline order across the many-body spectrum. Our results are consistent with the realization of an out-of-equilibrium Floquet phase of matter and introduce a programmable quantum simulator based on solid-state spins for exploring many-body physics.

A time crystal spontaneously breaks time-translation symmetry (*1*–*3*). Understanding the consequences and limitations of this idea has motivated intense exploration across a multitude of physical settings (*4*–*8*). A particularly fruitful case has been that of isolated periodically driven (Floquet) quantum systems, in which the stability of time crystals is connected to fundamental questions in nonequilibrium statistical mechanics (*4*, *5*, *9*–*11*).

[1]QuTech, Delft University of Technology, PO Box 5046, 2600 GA Delft, Netherlands. [2]Kavli Institute of Nanoscience Delft, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, Netherlands. [3]Element Six Innovation, Fermi Avenue, Harwell Oxford, Didcot, Oxfordshire OX11 0QR, UK. [4]Department of Physics, University of California, Berkeley, CA 94720, USA. [5]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
***Corresponding author. Email: t.h.taminiau@tudelft.nl**
†These authors contributed equally to this work.

In a Floquet system, the spontaneous breaking of the time-translation symmetry associated with the periodic drive results in a discrete time crystal (DTC) (*9*–*11*). The DTC is a new phase of matter that locks onto a period that is a multiple of that of the drive and is stable against perturbations. Experiments have revealed signatures of discrete time-crystalline behavior in a range of systems including trapped ions (*12*, *13*), spin ensembles (*14*–*17*), ultracold atoms (*18*, *19*), and superconducting qubits (*20*).

A key challenge is that the stability of the DTC, as well as any type of order in a Floquet many-body system, requires a mechanism to avoid thermalization and heating from the periodic drive. Although there exist a number of strategies for exponentially delaying thermalization (*7*, *13*, *21*–*23*), the only known mechanism for breaking ergodicity in a generic interacting system is disorder-induced

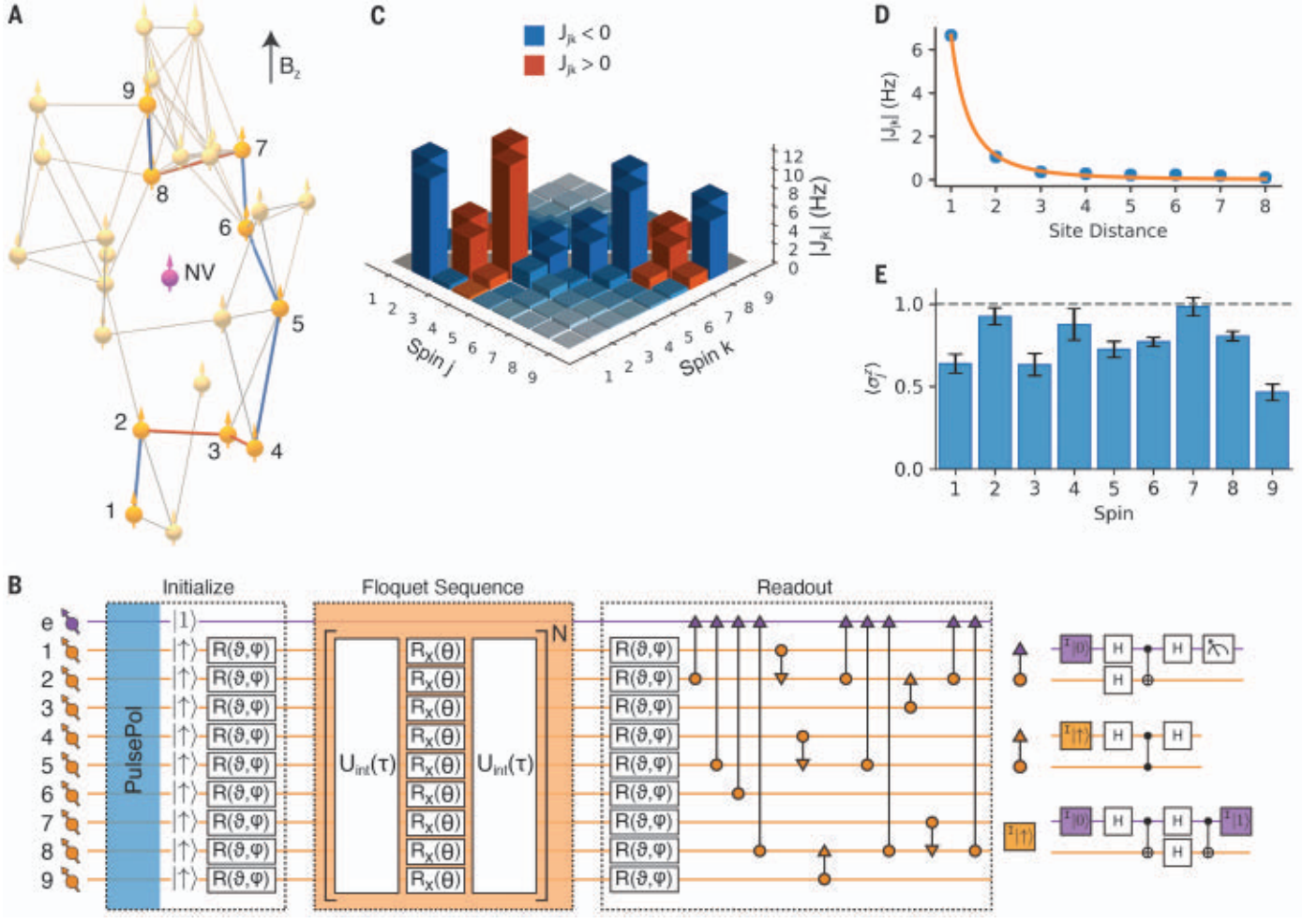


Fig. 1. Programmable spin-based quantum simulator. (**A**) We program an effective 1D chain of nine spins in an interacting cluster of 27 ^{13}C nuclear spins (orange) close to a single NV center. Connections indicate nuclear-nuclear couplings $|J_{jk}| > 1.5$ Hz, and blue (red) lines represent negative (positive) nearest-neighbor couplings within the chain (*29*). Magnetic field: $B_z \sim 403$ G. (**B**) Experimental sequence: The spins are initialized by applying the PulsePol sequence (*34*), followed by rotations of the form $R(\vartheta,\varphi) = \exp\left[-i\frac{\vartheta}{2}(\sin(\varphi)\sigma^x + \cos(\varphi)\sigma^y)\right]$. After evolution under N cycles of the Floquet unitary $U_F = U_{int}(\tau)\cdot U_x(\theta)\cdot U_{int}(\tau)$, the spins are sequentially read out through the NV electronic spin using electron-nuclear and nuclear-nuclear two-qubit gates (see text). Colored boxes with "I" denote re-initialization into the given state. (**C**) Coupling matrix for the nine-spin chain. (**D**) Average coupling magnitude as a function of site distance across the chain. Orange line: least-squares fit to a power-law function $J_0/|j-k|^\alpha$, giving $J_0 = 6.7(1)$ Hz and $\alpha = 2.5(1)$, confirming that the chain maps to an effective 1D system. (**E**) Measured expectation values $\langle\sigma_j^z\rangle$ after initializing the state $|\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\rangle$. The data are corrected for measurement errors (*27*).

many-body localization (MBL). Specifically in this context, DTC order is present across the full eigenspectrum of the Floquet system, translating into time-crystalline dynamics for generic initial states (*4*, *5*, *9*–*11*, *24*, *25*). The demonstration of such robust DTC order has remained an outstanding challenge (*4*, *26*).

Here, we present an observation of the hallmark signatures of the many-body–localized DTC phase. We develop a quantum simulator based on individually controllable and detectable ^{13}C nuclear spins in diamond, which can be used to realize a range of many-body Hamiltonians with tunable parameters and dimensionalities. We implement a Floquet sequence in a one-dimensional (1D) chain of $L = 9$ spins and observe the characteristic period doubling associated with the DTC. By combining the ability to prepare arbitrary initial states with site-resolved measurements, we observe the DTC response for a variety of initial states up to $N = 800$ Floquet cycles. This robustness for generic initial states provides a key signature to distinguish the many-body–localized DTC phase from prethermal responses, which only show a long-lived response for particular states (*13*, *21*, *22*, *26*).

Our experiments are performed on a system of ^{13}C nuclear spins in diamond close to a nitrogen-vacancy (NV) center at 4 K (Fig. 1A). The nuclear spins are well-isolated qubits with coherence times up to tens of seconds (*27*, *28*). They are coupled via dipole-dipole interactions and are accessed through the optically addressable NV electronic spin (*28*, *29*). With the electronic spin in the $m_s = -1$ state, the electron-nuclear hyperfine interaction induces a frequency shift h_j for each nuclear spin, which—combined with an applied magnetic field B_z in the z direction—reduces the dipolar interactions to Ising form (*27*). We additionally apply a radio-frequency (rf) driving field to implement nuclear-spin rotations. The nuclear-spin Hamiltonian is then given by $H = H_{int} + H_{rf}$, where H_{int} and H_{rf} describe the interaction and rf driving terms respectively:

$$H_{int} = \sum_j (B + h_j)\sigma_j^z + \sum_{j<k} J_{jk}\sigma_j^z\sigma_k^z$$
$$H_{rf} = \sum_j \Omega(t)\sigma_j^x \quad (1)$$

Here, σ_j^β ($\beta = x, y, z$) are the Pauli matrices for spin j, $B = \gamma_c B_z/2$ is the magnetic field

splitting, γ_c is the ^{13}C gyromagnetic ratio, J_{jk} is the zz component of the dipole-dipole interaction between spins j and k, and $\Omega(t)$ is the applied time-dependent rf field, and we set $\hbar = 1$. The system has previously been characterized in detail (*27*, *29*); for 27 ^{13}C spins, the hyperfine shifts h_j, the spatial coordinates, and the 351 interaction terms J_{jk} are known.

To investigate the DTC phase, we apply a Floquet unitary consisting of free evolution $U_{int}(\tau) = \exp(-iH_{int}\tau)$, interleaved with global spin rotations $U_x(\theta) = \exp\left(-i\theta\sum_j^L \sigma_j^x/2\right)$. To realize the global rotations, we develop multifrequency rf pulses that simultaneously rotate a chosen subset of spins (H_{rf} in Eq. 1) (*27*). We symmetrize the Floquet unitary such that $U_F = U_{int}(\tau)\cdot U_x(\theta)\cdot U_{int}(\tau)$, and apply N cycles of this basic sequence (Fig. 1B). For $\theta \sim \pi$, this decouples the targeted spins from their environment while preserving the internal interactions (*27*).

To stabilize MBL, the Floquet sequence $[U_F]^N$ should satisfy two requirements. First, the system should be low-dimensional and short-range interacting (*26*, *30*–*33*). This requirement is not naturally met in a coupled 3D spin system (Fig. 1A). To resolve this discrepancy, we program an effective 1D spin chain using a subset of nine spins [Fig. 1, A, C, and D (*27*)]. Second, because the periodic rotations approximately cancel the on-site disorder terms h_j, the system must exhibit Ising-even disorder to stabilize MBL in the Floquet setting (*4*, *11*, *26*). This is naturally realized in our system because the Ising couplings, J_{jk}, inherit the positional disorder of the nuclear spins. The disorder in the magnitude of the nearest-neighbor couplings is distributed over a range $W \sim 10$ Hz. The ratio of disorder to average nearest-neighbor coupling

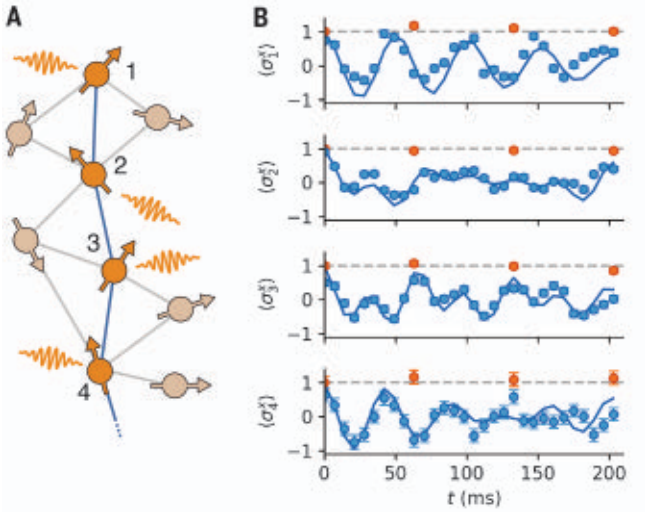


Fig. 2. Isolating spin chains. (**A**) We test the programming of interacting spin chains for the first four spins of the nine-spin chain (Fig. 1, A, C, and D). For $\theta \sim \pi$, the Floquet sequence $[U_F]^N$ decouples the spin chain from its environment but preserves the internal interactions. (**B**) Measured expectation values $\langle\sigma_j^x\rangle$ after initializing the state $|{+}{+}{+}{+}\rangle$ and applying $[U_F]^N$ with $\theta = \pi$. Here $t = 2\tau N$ is varied by fixing $\tau = 3.5$ ms and varying N. The blue (orange) points show the evolution with (without) spin-spin interactions (*27*). Blue lines: numerical simulations of only the four-spin system (*27*). Measurements in this figure and hereafter are corrected for state preparation and measurement errors.

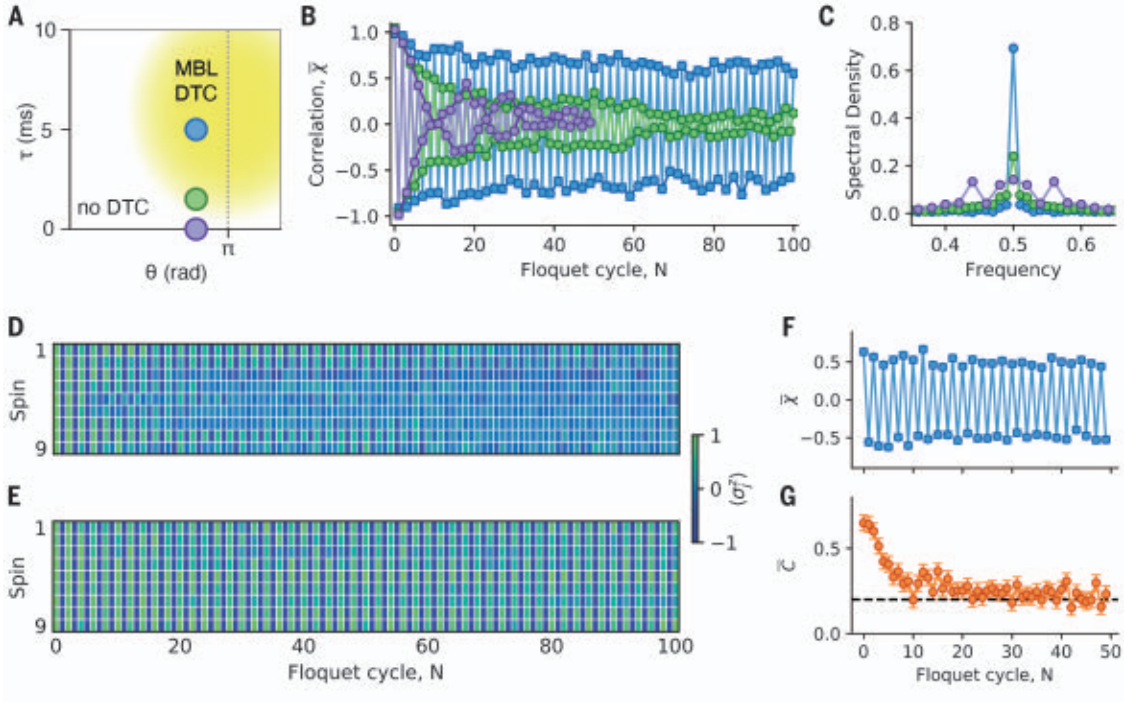


Fig. 3. Discrete time crystal in the nine-spin chain. (**A**) Sketch of the phase diagram as a function of τ and θ when applying the Floquet sequence $[U_F]^N$ (Fig. 1B) (*4*). The yellow region indicates the many-body–localized DTC phase. The colored points mark three combinations of $\{\theta, \tau\}$ that illustrate the DTC phase transition. Additional data for other values are given in the supplementary materials (*27*). (**B**) Averaged two-point correlation $\bar{\chi}$ as a function of the number of Floquet cycles N for $\theta = 0.95\pi$ and initial state $|{\uparrow}{\uparrow}{\uparrow}{\uparrow}{\uparrow}{\uparrow}{\uparrow}{\uparrow}{\uparrow}\rangle$. Without interactions [purple (*27*)], $\bar{\chi}$ decays quickly. With small interactions ($\tau = 1.55$ ms, green), the system is on the edge of the transition to the DTC phase. With strong interactions ($\tau = 5$ ms, blue), the subharmonic response is stable and persists over all 100 Floquet cycles. (**C**) The corresponding Fourier transforms show a sharp peak at $f = 0.5$ emerging as the system enters the DTC phase. (**D** and **E**) Individual spin expectation values $\langle\sigma_j^z\rangle$ for interaction times $\tau = 1.55$ ms (D) and $\tau = 5$ ms (E). (**F** and **G**) Averaged two-point correlation $\bar{\chi}$ (F) and coherence $\bar{C}$ (G) after preparing the superposition state $[\cos(\pi/8)|{\uparrow}\rangle + \sin(\pi/8)|{\downarrow}\rangle]^{\otimes 9}$ and applying $[U_F]^N$ with $\tau = 5$ ms. The subharmonic response in $\bar{\chi}$ is preserved, whereas $\bar{C}$ quickly decays because of interaction-induced local dephasing. The dashed line in (G) indicates a reference value for $\bar{C}$ measured after preparing the state $|{\uparrow}\rangle^{\otimes 9}$ (*27*).

is therefore $W/J_0 \sim 1.5$, comparable with previous theoretical studies of MBL DTCs (*4*, *9*, *26*).

To reveal the signature spatiotemporal order of the DTC phase, one must prepare a variety of initial states and perform site-resolved measurements (*9*, *26*). We use a combination of new and existing methods to realize the required initialization, single-spin control, and individual single-shot measurement for all spins in the chain (Fig. 1B).

First, we initialize the spins through a recently introduced dynamic nuclear polarization sequence called PulsePol (*34*). This sequence polarizes nuclear spins in the vicinity of the NV center and prepares the 1D chain in the state $|\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\rangle$. We analyze and optimize the polarization transfer in the supplementary materials (*27*). Subsequently, each spin can be independently rotated to an arbitrary state by selective rf pulses (*27*).

Second, after Floquet evolution, we read out the spins by sequentially mapping their $\langle\sigma_j^z\rangle$ expectation values to the NV electronic spin (*27*) and measuring the electronic-spin state by resonant optical excitation (*28*). Spins j = 2, 5, 6, 8 can be directly accessed using previously developed electron-nuclear two-qubit gates (*28*). To access the other spins (j = 1, 3, 4, 7, 9), which couple weakly to the NV, we develop a protocol based on nuclear-nuclear two-qubit gates through spin-echo double resonance [details and characterization are given in the supplementary materials (*27*)]. We use these gates to map the spin states to other, directly accessible spins in the chain. Figure 1E shows the measured $\langle\sigma_j^z\rangle$ expectation values after preparing the state $|\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\rangle$.

We verify that we can isolate the dynamics of a subset of spins by studying the first four spins of the nine-spin chain (Fig. 2A). We prepare the superposition state $|++++\rangle$, where $|+\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$, and apply $[U_F]^N$ with $\theta = \pi$. We first verify that the state is preserved when each spin is individually decoupled to remove interactions (Fig. 2B) (*27*). By contrast, with internal interactions, the four spins entangle and undergo complex dynamics. The measured evolution matches a numerical simulation containing only the four spins, indicating that the system is strongly interacting and protected from external decoherence.

With this capability confirmed, we turn to the nine-spin chain and the DTC phase. The expectation for the DTC phase is a long-lived period-doubled response that is stabilized against perturbations of U_F through many-body interactions. To illustrate this, we set $\theta = 0.95\pi$, a perturbation from the ideal value of π, and tune the system through the DTC phase transition by changing τ, which effectively sets the interaction strength (Fig. 3, A to C).

We first investigate the state $|\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\rangle$ and consider the averaged two-point correlation function $\bar{\chi} = \frac{1}{L}\sum_{j=1}^{L}\langle\sigma_j^z(N)\rangle \mathrm{sgn}[\langle\sigma_j^z(0)\rangle]$, where $\langle\sigma_j^z(N)\rangle$ is the expectation value at Floquet cycle N for spin j. Without interactions, the deliberate under-rotations ($\theta < \pi$), in combination with naturally present noise in the applied control fields, lead to a rapid decay (Fig. 3, B and C). By introducing moderate interactions (τ = 1.55 ms), the system is on the edge of the phase transition, and the interactions begin to stabilize the subharmonic response (Fig. 3, B to D). Finally, for strong interactions (τ = 5 ms), the subharmonic response is stabilized despite the perturbations of θ (Fig. 3, B, C, and E). The individual spin measurements confirm that the spins are synchronized, and the signature long-lived spatiotemporal response is observed (Fig. 3E).

To rule out trivial noninteracting explanations, we prepare the superposition state $[\cos(\pi/8)|\uparrow\rangle + \sin(\pi/8)|\downarrow\rangle]^{\otimes 9}$ and perform full single-qubit tomography for each spin for different values of N (*9*). The two-point correlation $\bar{\chi}$ shows a persistent subharmonic response similar to that of the initial state $|\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\rangle$ (Fig. 3F). By contrast, the coherence $\bar{C} = \frac{1}{L}\sum_{j=1}^{L}\sqrt{\langle\sigma_j^x\rangle^2 + \langle\sigma_j^y\rangle^2}$ shows a quick decay on a time scale of ~10 Floquet cycles, indicating rapid local dephasing due to internal many-body interactions that generate entanglement across the system (Fig. 3G).

Although the results shown in Fig. 3 are consistent with a DTC, these measurements

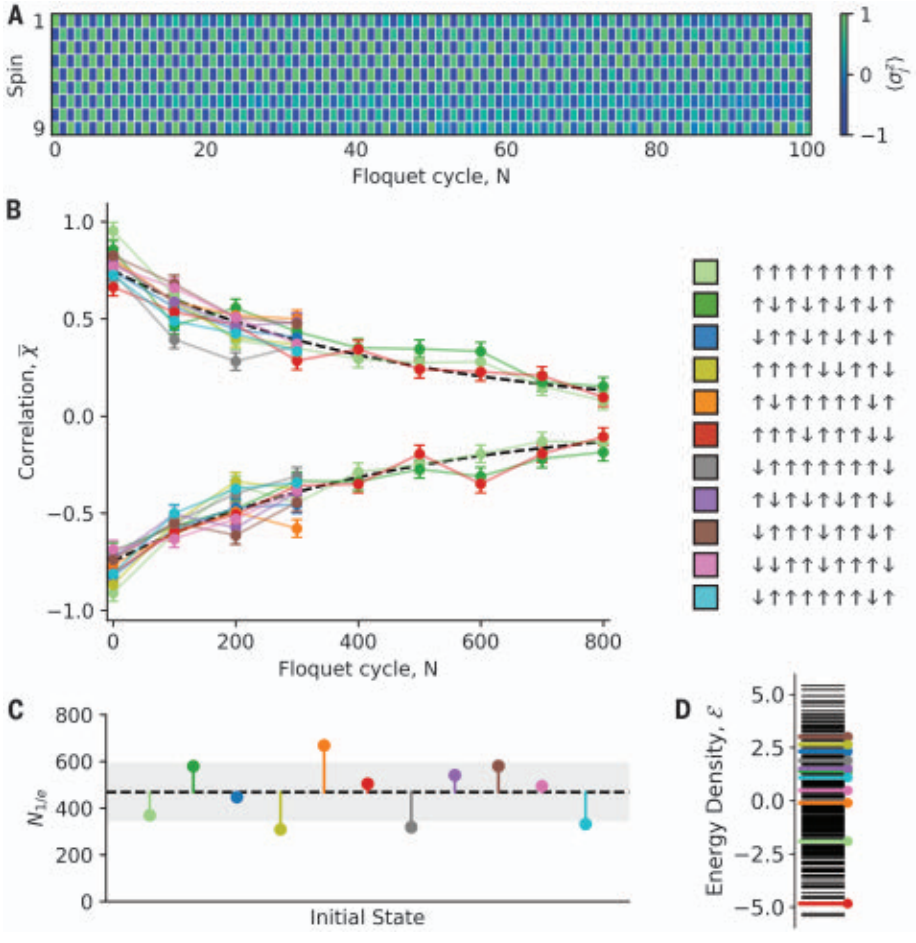


Fig. 4. Observation of the DTC response for generic initial states. (**A**) Individual spin expectation values $\langle\sigma_j^z\rangle$ as a function of N after initializing the spins in the Néel state $|\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\rangle$ and applying $[U_F]^N$ for $\theta = 0.95\pi$ and τ = 5 ms. (**B**) Average correlation for even (upper curve) and odd (lower curve) N for nine randomly chosen initial states, plus the polarized state and the Néel state (indicated in the legend) with $\theta = 0.95\pi$ and τ = 5 ms. Each data point is the average over even/odd integers in the range N to $N + 10$. Three of the states are measured up to N = 800, the others to N = 300. The dashed black line is a fit of $|\bar{\chi}|$, averaged over all states using a phenomenological function $f(N) = Ae^{-N/N_{1/e}}$, giving A = 0.75(2) and $N_{1/e}$ = 463(36). (**C**) $N_{1/e}$ for each initial state, extracted from a fit to $f(N)$ for the data in (B). The gray shaded region indicates the measurement uncertainty ($\pm 2\sigma$ around the mean) obtained through a Monte Carlo sampling of the fitting procedure (*27*). (**D**) Calculated energy density $\mathcal{E}$ for all possible states of the form $\otimes_j^L |m_j\rangle$, $m_j \in \{\uparrow, \downarrow\}$ (black lines). Colored lines: states indicated in the legend shown in (B).

alone do not distinguish the many-body–localized DTC phase from prethermal responses (*4*, *13*, *26*). In particular, the hallmark of the MBL DTC phase is robust time-crystalline order for generic initial states. Conversely, prethermal responses only exhibit long-lived oscillations for a particular range of initial states (*21*, *26*).

We study a range of generic initial states of the form $\otimes_j^L |m_j\rangle$, $m_j \in \{\uparrow, \downarrow\}$, including the Néel state $|\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\rangle$ (Fig. 4A) and nine additional random states (Fig. 4, B and C) . To illustrate that a variety of states are considered, we evaluate their energy density $\mathcal{E} = \langle H_{\text{eff}} \rangle / J_0 L$, where J_0 is the average nearest-neighbor coupling strength (Fig. 1D) and H_{eff} is the leading order term in the Floquet-Magnus expansion of U_{F} (*27*). The selected initial states extend across the energy spectrum (Fig. 4D).

The response up to $N = 800$ shows a stable period-doubled signal for all states, consistent with a DTC stabilized by MBL (Fig. 4, A and B). The $1/e$ decay value averaged over the states is $N_{1/e} = 463(36)$ Floquet cycles—corresponding to a time of ~4.6 s (Fig. 4B), and little dependence on the state is observed (Fig. 4C). By contrast, numerical calculations for a nine-spin chain with the same average couplings, but without disorder so that there is no MBL, show a strongly state-dependent response (*27*). Some initial states show a rapid decay, falling to $1/e$ within ~30 Floquet cycles and crossing through $\bar{\chi} = 0$ within 300 cycles, showing that such a prethermal response would be well distinguished within the experimental lifetime.

Although the DTC phase in an ideal system is predicted to persist to arbitrary times, any experimental implementation inevitably decays because of finite-size effects or environmental decoherence. Numerical calculations for the spin chain without decoherence yield a long-lived response up to ~10^6 Floquet cycles, as well as an exponential growth of the DTC lifetime with system size (*27*). This result shows that the chain is sufficiently large to display the hallmark divergence of MBL (*9*) and that its finite size does not limit the observed DTC lifetime. A characterization of the decoherence [$T_2 > 4800(900)$ periods] and relaxation (T_1, none observed) times for the spins shows that these are negligible over the time scale of the experiments (*27*). Therefore, the observed decay likely originates from residual interactions with the spin environment owing to imperfect decoupling under the Floquet sequence with $\theta \neq \pi$ (*27*). Such decoherence might be mitigated in the future with improved decoupling sequences.

In conclusion, we present an observation of the hallmark signatures of the many-body–localized DTC phase. Compared with previous experiments, the observed time-crystalline response is stable for generic initial states, demonstrating robust DTC order across the many-body spectrum. This result highlights the importance of both many-body interactions and disorder for stabilizing MBL-DTC order. The developed methods also provide new opportunities to investigate other Floquet phases of matter, including topologically protected phases (*4*), and time-crystalline order in a variety of settings complementary to MBL, such as open systems where the interplay between dissipation and interactions leads to distinct DTC phenomena (*6*, *7*, *35*).

From a broader perspective, this work introduces a quantum simulator based on individually controllable solid-state spins that is naturally suited to studying many-body dynamics. By connecting different subsets of spins, larger one-dimensional chains and two- and three-dimensional systems can be realized (*36*, *37*). The combination of complete programmability through universal individual control, excellent coherence, and site-selective measurement enables the realization of a wide variety of many-body Hamiltonians. Future scalability beyond tens of spins might be achieved by exploiting spins external to the diamond (*38*, *39*), by linking multiple electronic-spin defects through dipolar coupling (*40*), by photonic remote entanglement (*41*), or by combinations of these methods.

Note added in proof: While this manuscript was undergoing peer review, a related preprint appeared that reports the observation of an MBL DTC using superconducting qubits (*42*).

REFERENCES AND NOTES

1. F. Wilczek, *Phys. Rev. Lett.* **109**, 160401 (2012).
2. A. Shapere, F. Wilczek, *Phys. Rev. Lett.* **109**, 160402 (2012).
3. H. Watanabe, M. Oshikawa, *Phys. Rev. Lett.* **114**, 251603 (2015).
4. V. Khemani, R. Moessner, S. L. Sondhi, arXiv:1910.10745 [cond-mat.str-el] (2019).
5. D. V. Else, C. Monroe, C. Nayak, N. Y. Yao, *Annu. Rev. Condens. Matter Phys.* **11**, 467–499 (2020).
6. Z. Gong, R. Hamazaki, M. Ueda, *Phys. Rev. Lett.* **120**, 040404 (2018).
7. N. Y. Yao, C. Nayak, L. Balents, M. P. Zaletel, *Nat. Phys.* **16**, 438–447 (2020).
8. H. Keßler *et al.*, *Phys. Rev. Lett.* **127**, 043602 (2021).
9. D. V. Else, B. Bauer, C. Nayak, *Phys. Rev. Lett.* **117**, 090402 (2016).
10. V. Khemani, A. Lazarides, R. Moessner, S. L. Sondhi, *Phys. Rev. Lett.* **116**, 250401 (2016).
11. N. Y. Yao, A. C. Potter, I.-D. Potirniche, A. Vishwanath, *Phys. Rev. Lett.* **118**, 030401 (2017).
12. J. Zhang *et al.*, *Nature* **543**, 217–220 (2017).
13. A. Kyprianidis *et al.*, *Science* **372**, 1192–1196 (2021).
14. S. Choi *et al.*, *Nature* **543**, 221–225 (2017).
15. J. O'Sullivan *et al.*, *New J. Phys.* **22**, 085001 (2020).
16. J. Rovny, R. L. Blum, S. E. Barrett, *Phys. Rev. Lett.* **120**, 180603 (2018).
17. S. Pal, N. Nishad, T. S. Mahesh, G. J. Sreejith, *Phys. Rev. Lett.* **120**, 180602 (2018).
18. J. Smits, L. Liao, H. T. C. Stoof, P. van der Straten, *Phys. Rev. Lett.* **121**, 185301 (2018).
19. S. Autti, V. B. Eltsov, G. E. Volovik, *Phys. Rev. Lett.* **120**, 215301 (2018).
20. P. Frey, S. Rachel, arXiv:2105.06632v1 [quant-ph] (2021).
21. D. V. Else, B. Bauer, C. Nayak, *Phys. Rev. X* **7**, 011026 (2017).
22. F. Machado, D. V. Else, G. D. Kahanamoku-Meyer, C. Nayak, N. Y. Yao, *Phys. Rev. X* **10**, 011043 (2020).
23. P. Peng, C. Yin, X. Huang, C. Ramanathan, P. Cappellaro, *Nat. Phys.* **17**, 444–447 (2021).
24. D. A. Abanin, E. Altman, I. Bloch, M. Serbyn, *Rev. Mod. Phys.* **91**, 021001 (2019).
25. L. D'Alessio, M. Rigol, *Phys. Rev. X* **4**, 041048 (2014).
26. M. Ippoliti, K. Kechedzhi, R. Moessner, S. Sondhi, V. Khemani, *PRX Quantum* **2**, 030346 (2021).
27. See supplementary materials.
28. C. E. Bradley *et al.*, *Phys. Rev. X* **9**, 031045 (2019).
29. M. H. Abobeih *et al.*, *Nature* **576**, 411–415 (2019).
30. We note that avalanche instabilities might destabilize MBL in power-law interacting systems, although such effects are outside of current experimental and numerical capabilities (*33*).
31. N. Y. Yao *et al.*, *Phys. Rev. Lett.* **113**, 243002 (2014).
32. A. L. Burin, *Phys. Rev. B* **91**, 094202 (2015).
33. W. De Roeck, F. Huveneers, *Phys. Rev. B* **95**, 155129 (2017).
34. I. Schwartz *et al.*, *Sci. Adv.* **4**, eaat8978 (2018).
35. A. Lazarides, S. Roy, F. Piazza, R. Moessner, *Phys. Rev. Res.* **2**, 022002 (2020).
36. L. M. K. Vandersypen, I. L. Chuang, *Rev. Mod. Phys.* **76**, 1037–1069 (2005).
37. J. Choi *et al.*, *Phys. Rev. X* **10**, 031002 (2020).
38. J. Cai, A. Retzker, F. Jelezko, M. B. Plenio, *Nat. Phys.* **9**, 168–173 (2013).
39. I. Lovchinsky *et al.*, *Science* **355**, 503–507 (2017).
40. F. Dolde *et al.*, *Nat. Phys.* **9**, 139–143 (2013).
41. M. Pompili *et al.*, *Science* **372**, 259–264 (2021).
42. X. Mi *et al.*, arXiv:2107.13571 [quant-ph] (2021).
43. J. Randall, C. E. Bradley, F. V. van der Gronden, A. Galicia, M. H. Abobeih, M. Markham, D. J. Twitchen, F. Machado, N. Y. Yao, T. H. Taminiau, Data and plotting code for 'Many-body-localized discrete time crystal with a programmable spin-based quantum simulator,' Zenodo (2021). https://doi.org/10.5281/zenodo.5636045.

ACKNOWLEDGMENTS

We thank W. Hahn and V. V. Dobrovitski for valuable discussions and A. Breitweiser for experimental assistance. **Funding:** This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW) through a Vidi grant and as part of the Frontiers of Nanoscience (NanoFront) program and the Quantum Software Consortium program (Project no. 024.003.037/3368). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 852410). This project (QIA) has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 820445. F.M. acknowledges support from the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division and the Division of Chemical Sciences, Geosciences and Biosciences at LBNL under Contract no. DE-AC02-05-CH11231. N.Y.Y. acknowledges support from the DARPA DRINQS program (D18AC00033), the Army Research Office (grant no. W911NF2110262), the David and Lucile Packard foundation, and the W. M. Keck foundation. **Author contributions:** J.R., C.E.B., F.V.vdG., A.G., and T.H.T. devised and performed the experiments, developed theoretical calculations, and performed numerical simulations. J.R., C.E.B., F.V.vdG., A.G., F.M., N.Y.Y., and T.H.T. analyzed the data. J.R., C.E.B., M.H.A., and T.H.T. prepared the experimental apparatus. M.M. and D.J.T. grew the diamond sample. J.R., C.E.B., F.M., N.Y.Y., and T.H.T. wrote the manuscript with input from all authors. T.H.T. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The underlying data and software code for generating the plots presented in the main text and supplementary materials are available at Zenodo (*43*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk0603
Materials and Methods
Figs. S1 to S9
Tables S1 to S6
References (*44*–*58*)

25 June 2021; accepted 21 October 2021
Published online 4 November 2021
10.1126/science.abk0603

QUANTUM SIMULATION

Information scrambling in quantum circuits

Xiao Mi[1]†, Pedram Roushan[1]†, Chris Quintana[1]†, Salvatore Mandrà[2,3], Jeffrey Marshall[2,4], Charles Neill[1], Frank Arute[1], Kunal Arya[1], Juan Atalaya[1], Ryan Babbush[1], Joseph C. Bardin[1,5], Rami Barends[1], Joao Basso[1], Andreas Bengtsson[1], Sergio Boixo[1], Alexandre Bourassa[1,6], Michael Broughton[1], Bob B. Buckley[1], David A. Buell[1], Brian Burkett[1], Nicholas Bushnell[1], Zijun Chen[1], Benjamin Chiaro[1], Roberto Collins[1], William Courtney[1], Sean Demura[1], Alan R. Derk[1], Andrew Dunsworth[1], Daniel Eppens[1], Catherine Erickson[1], Edward Farhi[1], Austin G. Fowler[1], Brooks Foxen[1], Craig Gidney[1], Marissa Giustina[1], Jonathan A. Gross[1], Matthew P. Harrigan[1], Sean D. Harrington[1], Jeremy Hilton[1], Alan Ho[1], Sabrina Hong[1], Trent Huang[1], William J. Huggins[1], L. B. Ioffe[1], Sergei V. Isakov[1], Evan Jeffrey[1], Zhang Jiang[1], Cody Jones[1], Dvir Kafri[1], Julian Kelly[1], Seon Kim[1], Alexei Kitaev[1,7], Paul V. Klimov[1], Alexander N. Korotkov[1,8], Fedor Kostritsa[1], David Landhuis[1], Pavel Laptev[1], Erik Lucero[1], Orion Martin[1], Jarrod R. McClean[1], Trevor McCourt[1], Matt McEwen[1,9], Anthony Megrant[1], Kevin C. Miao[1], Masoud Mohseni[1], Shirin Montazeri[1], Wojciech Mruczkiewicz[1], Josh Mutus[1], Ofer Naaman[1], Matthew Neeley[1], Michael Newman[1], Murphy Yuezhen Niu[1], Thomas E. O'Brien[1], Alex Opremcak[1], Eric Ostby[1], Balint Pato[1], Andre Petukhov[1], Nicholas Redd[1], Nicholas C. Rubin[1], Daniel Sank[1], Kevin J. Satzinger[1], Vladimir Shvarts[1], Doug Strain[1], Marco Szalay[1], Matthew D. Trevithick[1], Benjamin Villalonga[1], Theodore White[1], Z. Jamie Yao[1], Ping Yeh[1], Adam Zalcman[1], Hartmut Neven[1], Igor Aleiner[1], Kostyantyn Kechedzhi[1]*, Vadim Smelyanskiy[1]*, Yu Chen[1]*

Interactions in quantum systems can spread initially localized quantum information into the exponentially many degrees of freedom of the entire system. Understanding this process, known as quantum scrambling, is key to resolving several open questions in physics. Here, by measuring the time-dependent evolution and fluctuation of out-of-time-order correlators, we experimentally investigate the dynamics of quantum scrambling on a 53-qubit quantum processor. We engineer quantum circuits that distinguish operator spreading and operator entanglement and experimentally observe their respective signatures. We show that whereas operator spreading is captured by an efficient classical model, operator entanglement in idealized circuits requires exponentially scaled computational resources to simulate. These results open the path to studying complex and practically relevant physical observables with near-term quantum processors.

The realization of quantum computers was motivated by their ability to simulate dynamical processes that are challenging for classical computation. A physical process that fully leverages the computational power of quantum processors is quantum scrambling, which describes how interaction in a quantum system disperses local information into the system's many degrees of freedom (*1*–*5*). Quantum scrambling is the underlying mechanism for the thermalization of isolated quantum systems (*6*–*8*), and accurately modeling its dynamics is the key to resolving a number of open questions, including the fast-scrambling conjecture for black holes (*2*, *3*), the nature of strange metals (*9*, *10*), and many-body localization (*11*). Understanding scrambling also provides a basis for designing algorithms in quantum benchmarking or machine learning that would benefit from the efficient exploration of Hilbert spaces (*12*–*14*).

A precise formulation of quantum scrambling is found in the Heisenberg picture, where quantum operators evolve and quantum states are stationary. Similar to classical chaos, scrambling manifests itself as a butterfly effect, wherein a local perturbation is rapidly amplified over time (*15*, *16*). More specifically, the perturbation is realized as an initially local unitary operator $\hat{O}$ acting on one of the N qubits Q_b. When the quantum system undergoes a unitary process $\hat{U}$, $\hat{O}$ acquires a time dependence and becomes $\hat{O}(t) = \hat{U}^\dagger \hat{O} \hat{U}$. The resulting $\hat{O}(t)$ can be expanded as $\hat{O}(t) = \sum_{i=1}^{n_B} w_i \hat{B}_i$, where $\hat{B}_i = \hat{b}_1(i) \otimes \hat{b}_2(i) \otimes \ldots$ are basis operators consisting of products of single-qubit operators $\hat{b}_j(i)$, each acting on a different (j^{th}) qubit, and w_i are their coefficients.

Quantum scrambling is enabled by two different mechanisms (*16*–*20*): (i) Operator spreading, wherein basis operators are transformed such that each $\hat{B}_i$ involves more nonidentity single-qubit operators over time and (ii) generation of operator entanglement (*21*), which is reflected in the growth, in time, of the minimum number of terms n_B needed to expand $\hat{O}$ into products of single-qubit operators, with a broad distribution of coefficients w_i. Independent characterizations of these two mechanisms are essential for a complete understanding of the nature of quantum scrambling. Additionally, operator spreading can be effectively mapped to classical dynamics (*16*, *18*–*20*, *22*, *23*), whereas for the circuits studied in this work, operator entanglement is a quantum process. As such, quantum simulation of operator entanglement dynamics has the potential to deliver quantum computational advantages in modeling physical phenomena. However, these two mechanisms are often intertwined, and signatures of operator entanglement have not been identified in past experimental studies of quantum scrambling (*24*–*28*).

In this work, we report independent observations of operator spreading and operator entanglement in a two-dimensional (2D) quantum system of 53 superconducting qubits. Our approach is based on measuring the correlator between $\hat{O}(t)$ and another unitary operator $\hat{M}$, which is a Pauli operator on a different qubit Q_1

$$C(t) = \left\langle \hat{O}^\dagger(t) \hat{M}^\dagger \hat{O}(t) \hat{M} \right\rangle \qquad (1)$$

Here, $\langle \ldots \rangle$ denotes the expectation value over a particular quantum state. $C(t)$ is known as the out-of-time-order correlator (OTOC) (*29*) and is related to the commutator $\left[\hat{O}(t), \hat{M}\right]$ by $\mathrm{Re}[C(t)] = 1 - \frac{1}{2}\left\langle \left[\hat{O}(t), \hat{M}\right]^\dagger \left[\hat{O}(t), \hat{M}\right] \right\rangle$. By measuring C over quantum evolution with microscopic differences, operator spreading may be characterized through the decay of the average OTOC value $\bar{C}$ from 1 as $\hat{O}(t)$ and $\hat{M}$ become noncommuting. In the fully scrambled limit, where $\hat{O}(t)$ and $\hat{M}$ have random commutation, $\bar{C}$ approaches an exponentially small limit $-\frac{1}{2^{2N}-1}$ (*17*). If operator entanglement is also generated, the growing number of contributions to C from basis operators $\hat{B}_i$ with similar magnitudes of w_i results in the decay of OTOC fluctuation δ_C as well. This property will be used as a witness of operator entanglement for a class of quantum circuits below.

The experimental scheme, described in Fig. 1A, utilizes an interferometric protocol that maps C onto the projection $\langle \hat{\sigma}_y \rangle$ of an ancilla qubit, Q_a (*30*, *31*). For this work, $\hat{U}$ is realized using quantum circuits comprising random single-qubit gates and fixed two-qubit gates (Fig. 1B); this choice was made because of the ease of controlling different scrambling mechanisms in such circuits. OTOC measurements are first conducted on a 1D chain of 21 qubits (Fig. 1C). We use the qubit at one end of the chain as Q_a and successively choose qubits Q_2 through

[1]Google Research, Mountain View, CA, USA. [2]QuAIL, NASA Ames Research Center, Moffett Field, CA, USA. [3]KBR, Inc., Houston, TX, USA. [4]USRA Research Institute for Advanced Computer Science, Mountain View, CA, USA. [5]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA. [6]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, USA. [7]Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA. [8]Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA. [9]Department of Physics, University of California, Santa Barbara, CA, USA. *Corresponding author. Email: kostyantyn@google.com (K.K.); smelyan@google.com (V.S.); bryanchen@google.com (Y.C.) †These authors contributed equally to this work.

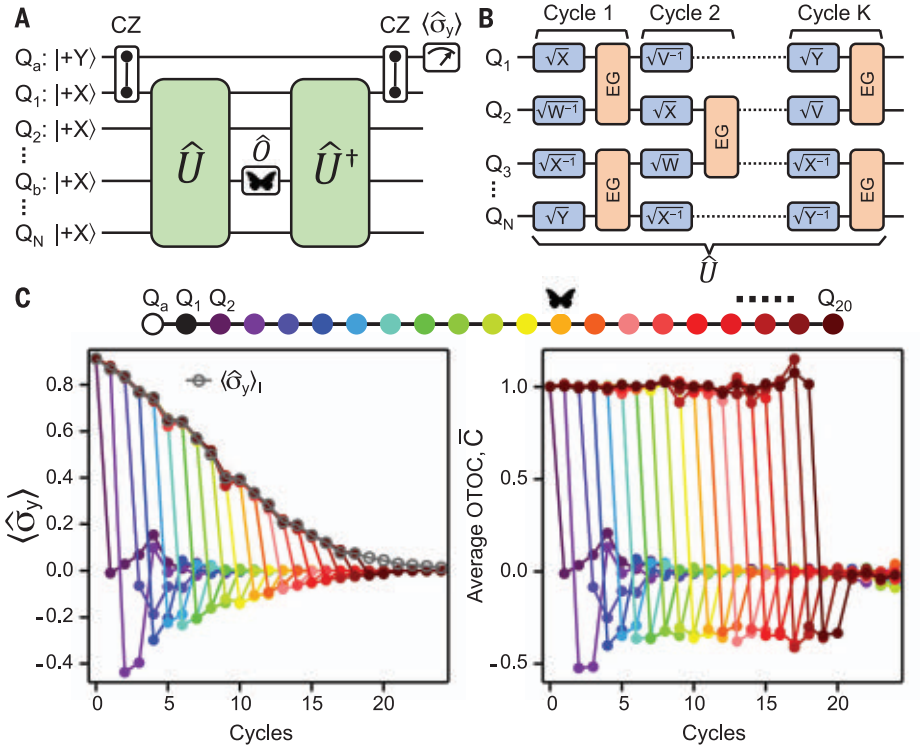


Fig. 1. OTOC measurement protocol. (**A**) Experimental scheme: A quantum circuit $\hat{U}$ and its inverse $\hat{U}^\dagger$ are successively applied to a quantum system (qubits Q_1 through Q_N), with a local operator $\hat{O} = \hat{\sigma}_x^{(b)}$ in between (butterfly symbol). Q_a is entangled with Q_1 through controlled-phase (CZ) gates, and its projection to the *y* axis of the Bloch sphere $\langle\hat{\sigma}_y\rangle$ is measured at the end. The unitary operator $\hat{M}$ is $\hat{\sigma}_z^{(1)}$ in this experiment. +*X* and +*Y* denote single-qubit states along the *x* and *y* axes, respectively, of the Bloch sphere. (**B**) The structure of $\hat{U}$ consists of *K* cycles: Each cycle includes one layer of single-qubit gates $\left(\text{randomly chosen from } \sqrt{X^{\pm1}}, \sqrt{Y^{\pm1}}, \sqrt{W^{\pm1}}, \text{and } \sqrt{V^{\pm1}}\right)$ and one layer of two-qubit entangling gates (EGs). Here, $W = \frac{X+Y}{\sqrt{2}}$ and $V = \frac{X-Y}{\sqrt{2}}$. (**C**) (Left) The filled circles represent cycle-dependent $\langle\hat{\sigma}_y\rangle$ measured with Q_b successively chosen from Q_2 through Q_N. Here, $N = 20$. The open gray circles represent $\langle\hat{\sigma}_y\rangle_I$, which is $\langle\hat{\sigma}_y\rangle$ without applying $\hat{O}$. The data are averaged over 60 random circuit instances. (Right) Average OTOCs $\bar{C}$ for different Q_b, obtained by dividing the corresponding $\langle\hat{\sigma}_y\rangle$ by $\langle\hat{\sigma}_y\rangle_I$.

Q_{20} as Q_b. A two-qubit gate (iSWAP) is applied to each pair (Q_j, Q_{j+1}), where j = 0,2,4... in odd circuit cycles and j = 1,3,5... in even circuit cycles.

In the left panel of Fig. 1C, experimental values of $\langle\hat{\sigma}_y\rangle$ are shown for different numbers of cycles in $\hat{U}$. Here, we have averaged the data over different circuit instances to initially focus on operator spreading. A sharp decrease is seen in each $\langle\hat{\sigma}_y\rangle$ when the number of circuit cycles has first exceeded the number of qubits between Q_b and Q_1—the minimum time required for $\hat{M}$ and the time-evolved operator $\hat{O}(t)$ to overlap and become noncommuting. This propagating behavior, although indicative of operator spreading, is also complicated by errors in the quantum circuits, such as qubit decoherence or any mismatch between $\hat{U}$ and $\hat{U}^\dagger$. These noisy effects are evident from the decay in $\langle\hat{\sigma}_y\rangle$ while $\hat{O}(t)$ and $\hat{M}$ still commute. To clearly distinguish noise from scrambling, we also measure $\langle\hat{\sigma}_y\rangle$ in the absence of $\hat{O}$ (shown as $\langle\hat{\sigma}_y\rangle_I$ in the left panel of Fig. 1C) (*32*). This control experiment realizes an identity operation, and the decay of $\langle\hat{\sigma}_y\rangle_I$ is induced purely by noise (*25*). As such, the difference between $\langle\hat{\sigma}_y\rangle_I$ and each $\langle\hat{\sigma}_y\rangle$ in Fig. 1C after the overlap of $\hat{O}(t)$ and $\hat{M}$ constitutes an unambiguous signature of operator spreading in our experiment.

In addition to verifying operator spreading, $\langle\hat{\sigma}_y\rangle_I$ approximates circuit fidelity (*25*) and therefore enables the reconstruction of ideal circuit dynamics when used to normalize each $\langle\hat{\sigma}_y\rangle$ (*32*). We test this error-mitigation protocol and plot the ratio $\bar{C} = \langle\hat{\sigma}_y\rangle/\langle\hat{\sigma}_y\rangle_I$ in the right panel of Fig. 1C (the data for the rest of this work are shown after similar normalization procedures). $\bar{C}$ exhibits nearly noise-free features of operator spreading: For each location of Q_b, $\bar{C}$ retains values of ~1 when $\hat{O}(t)$ and $\hat{M}$ are nonoverlapping and commute with each other. After $\hat{O}(t)$ and $\hat{M}$ have overlapped and ceased to commute, $\bar{C}$ converges to much smaller values. Additionally, we observe that the time evolution of $\bar{C}$ for each Q_b resembles a ballistically propagating wave. The front of each wave coincides with the edge of the light cone associated with Q_1—i.e., the set of qubits that have been entangled with Q_1. This profile is attributed to the iSWAP gates used in these circuits, which spread single-qubit operators at the same rate as their light cones expand (*33*, *34*). For generic quantum circuits, the spreading velocity (also known as the butterfly velocity) is typically slower. Using the full 2D system, we next demonstrate how the evolution of $\bar{C}$ may be used to diagnose the butterfly velocity of operator spreading.

In Fig. 2A, the spatial distribution of $\bar{C}$ is shown for five different numbers of cycles in $\hat{U}$, with iSWAP still being the two-qubit gate. We see that the number of qubits with $\bar{C} < 1$ rapidly increases with the number of cycles, consistent with the spatial spread of the time-evolved operator $\hat{O}(t)$. Moreover, for each circuit cycle, the values of $\bar{C}$ abruptly change across the edge of the light cone associated with Q_1 (dashed lines in Fig. 2A). By contrast, the spatiotemporal evolution of $\bar{C}$ shown in Fig. 2B is significantly different. Here, the iSWAP gates are replaced with $\sqrt{\text{iSWAP}}$ gates, and the decay of $\bar{C}$ is slower. Qubits far from Q_1 retain $\bar{C}$ close to 1 even after 22 cycles. The sharp, step-like spatial transition seen with iSWAP is also absent for $\sqrt{\text{iSWAP}}$. Instead, $\bar{C}$ changes in a gradual fashion as Q_b moves further away from Q_1.

The different OTOC behaviors can alternatively be seen in the full temporal evolution of four specific qubits (Fig. 2C). For iSWAP, the shape of the OTOC wavefront remains sharp and relatively insensitive to the location of Q_b, similar to the 1D example in Fig. 1C. On the other hand, the wavefront propagates more slowly for $\sqrt{\text{iSWAP}}$ and also broadens as the distance between Q_b and Q_1 increases. As a result, more circuit cycles are required before $\bar{C}$ reaches 0 for $\sqrt{\text{iSWAP}}$. The wavefront behavior seen with $\sqrt{\text{iSWAP}}$ is similar to that observed for generic quantum circuits analyzed in past works (*18*, *19*, *35*).

The observed features of average OTOCs are quantitatively understood by mapping operator spreading to a classical Markov process involving population dynamics (*36*). In this model, the 2D qubit lattice is populated by fictitious particles representing two copies of a single-qubit operator. The initial state of the entire system is a single particle at the site of Q_b. Whenever a two-qubit gate is applied to two neighboring lattice sites, their particle occupation changes between four possible states: ◇◇ (both empty), ◇♦(left empty, right filled), ♦◇ (right empty, left filled), and ♦♦

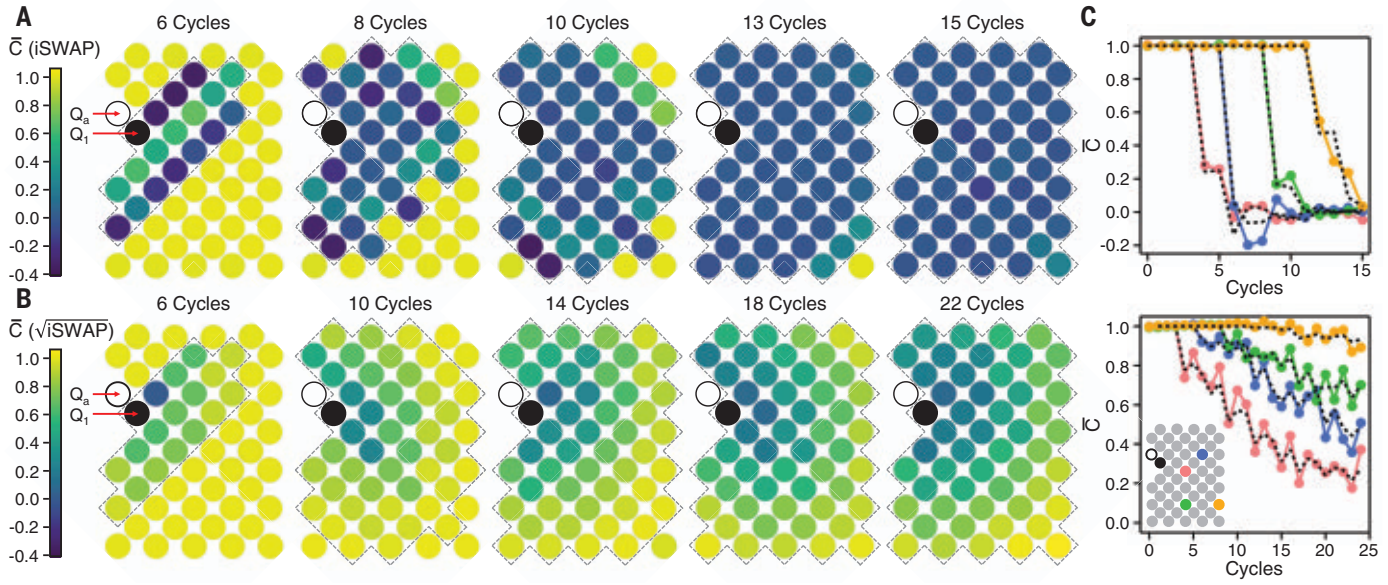


Fig. 2. OTOC propagation and speed of operator spreading. (**A**) Spatial profiles of average OTOCs, $\bar{C}$, measured on the full 53-qubit processor. Q_a and Q_1 are indicated by the red arrows. The colors of the other filled circles represent $\bar{C}$ with different choices of Q_b. The two-qubit gates are iSWAP and applied between all nearest-neighbor qubits, in the same order as done in (*38*). The dashed lines delineate the light cone of Q_1. The data are averaged over 38 circuit instances. (**B**) Similar to (A) but with $\sqrt{\text{iSWAP}}$ as the two-qubit gates. Here, $\bar{C}$ is averaged over 24 circuit instances. (**C**) Cycle-dependent $\bar{C}$ for four different choices of Q_b. The top and bottom panels shows data with iSWAP and $\sqrt{\text{iSWAP}}$, respectively, as the two-qubit gates. The colors of the data points indicate the locations of Q_b (inset in bottom panel). Dashed lines show theoretical predictions based on a classical population dynamics model. A two-qubit error rate $p_2 = 0.012$ is included in the prediction for $\sqrt{\text{iSWAP}}$ circuits, and noiseless simulation can be found in fig. S24 (*36*).

(both filled). The transition probabilities are described by the stochastic matrix

$$\Omega = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-a-b & a & b \\ 0 & a & 1-a-b & b \\ 0 & \frac{b}{3} & \frac{b}{3} & 1-\frac{2}{3}b \end{pmatrix} \quad (2)$$

where $a = \frac{1}{3}\sin^4\theta$ and $b = \frac{1}{3}\left(\frac{1}{2}\sin^2 2\theta + 2\sin^2\theta\right)$, with θ being the swap angle of the two-qubit gate. The average probability of finding a particle at the site of Q_1 is then used to estimate $\bar{C}$. We note that the effect of gate errors can also be included by multiplying all transition probabilities (except $\diamond\diamond \rightarrow \diamond\diamond$) by a constant factor $e^{-\frac{16}{15}p_2}$, where p_2 is an error rate associated with each two-qubit cycle (*36*). In this classical picture, the OTOC wavefront corresponds to the boundary separating the empty region from the region populated by particles.

The difference in OTOC propagation between iSWAP and $\sqrt{\text{iSWAP}}$ is then captured by the dependence of Ω on θ: After each application of an iSWAP gate ($\theta = \pi/2$), the particle occupation changes (with the exception of $\diamond\diamond$). In particular, any previously empty sites will be filled, and the region populated by particles always grows. This leads to the observed maximal butterfly velocity. By contrast, the application of any $\sqrt{\text{iSWAP}}$ gate ($\theta = \pi/4$) can leave the particle occupation unchanged with a probability $\frac{5}{12}$. $\bar{C}$ therefore decays more slowly in this case, and its broadening is explained by the fact that the wavefront spreads at a different velocity for each trial of the Markov process. The predicted values of $\bar{C}$, plotted as dashed lines in Fig. 2C, agree well with the experimental data for iSWAP even when no noise is added ($p_2 = 0$). In the case of $\sqrt{\text{iSWAP}}$, we find that population dynamics agree equally well with experimental data when an empirical error rate $p_2 = 0.012$ is included in the simulation.

Unlike operator spreading, an efficient classical description of operator entanglement is not known to exist for our system. For nonintegrable systems in particular, circuit-to-circuit fluctuation of OTOCs away from asymptotic limits cannot be modeled by population dynamics (*36*). Resolving the growth of operator entanglement is also difficult with a quantum processor because it is often accompanied by increased operator spreading (*16*–*19*). We overcome this challenge by gradually adjusting the composition of $\hat{U}$ and $\hat{U}^\dagger$, realizing a group of circuits with predominantly Clifford gates $\left(\text{iSWAP}, \sqrt{X^{\pm1}} \text{ or } \sqrt{Y^{\pm1}}\right)$ (*37*) and a small number of non-Clifford gates $\left(\sqrt{W^{\pm1}} \text{ and } \sqrt{V^{\pm1}}\right)$. For such circuits, $\hat{O}(t)$ may be conveniently expanded into products of single-qubit Pauli operators (Pauli strings) for analysis of operator entanglement, where numerical simulation studies demonstrate the same qualitative behavior as the minimal product basis $\hat{B}_i$ (*36*). This is because sufficiently long random Clifford circuits map a Pauli string into another weakly correlated Pauli string, and as a result, this basis cannot be improved by single-qubit rotations [see section VII of (*36*)]. The transformation of Pauli strings by representative gates in our circuits is illustrated in Fig. 3A: Clifford gates preserve the total number of Pauli strings but generate operator spreading given that iSWAP can increase the number of nonidentity Pauli operators. By contrast, non-Clifford gates generate operator entanglement by transforming one single Pauli string into a superposition of multiple Pauli strings, maintaining the spatial extent of operator spreading in the process.

These distinctive properties of Clifford and non-Clifford gates therefore provide us with a way to independently tune one scrambling mechanism without affecting the other. We now focus on operator entanglement and measure the circuit-to-circuit fluctuation of OTOCs, as shown in Fig. 3B. Here, the number of circuit cycles is fixed at 12, and the number of non-Clifford gates in $\hat{U}$, N_{wv}, is successively changed from 0 to 32. For each N_{wv}, the individual OTOCs C of 130 random circuit instances are measured using a modified normalized procedure [fig. S13 (*36*)]. At $N_{\text{wv}} = 0$, where the circuits consist of only Clifford gates, we see that C takes discrete values of 1 or −1.

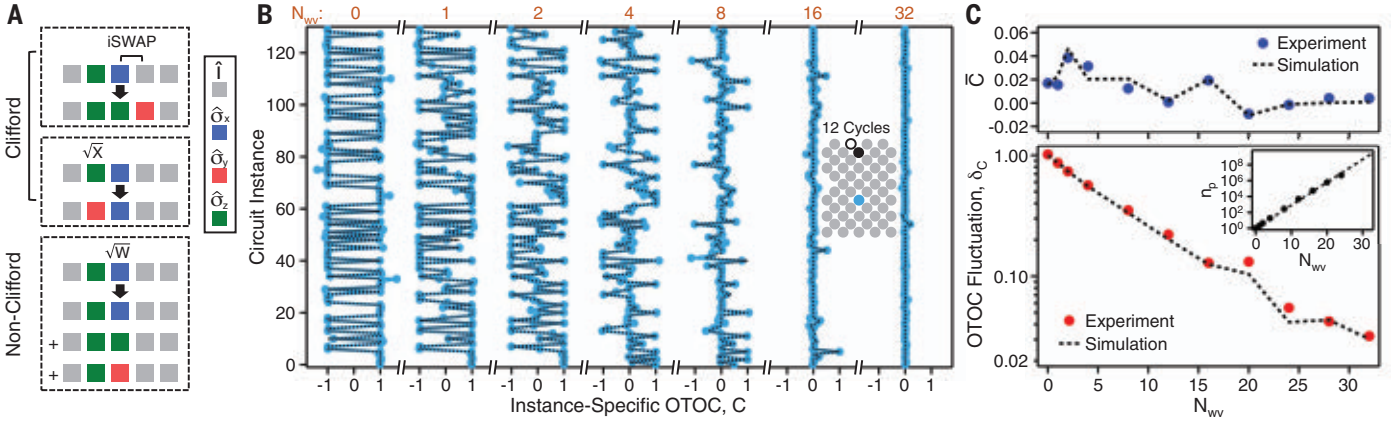


Fig. 3. OTOC fluctuations and signatures of operator entanglement. (**A**) Example transformation of a product of Pauli operators (Pauli string) by different quantum gates. A single Pauli string $\hat{I}^{(1)}\hat{\sigma}_z^{(2)}\hat{\sigma}_x^{(3)}\hat{I}^{(4)}\hat{I}^{(5)}$ is mapped either into a different Pauli string by a Clifford gate or a superposition of multiple Pauli strings (coefficients not shown) by a non-Clifford gate. (**B**) OTOCs of individual random circuit instances, C, measured with the number of non-Clifford gates in $\hat{U}$, N_{wv}, fixed at different values. Dashed lines are numerical simulation results. For each circuit, the non-Clifford gates are injected at random locations within the intersection between the light cones of Q_b and Q_1. The inset shows locations of Q_a (black-outlined unfilled circle), Q_1 (black filled circle), and Q_b (blue filled circle) as well as the number of circuit cycles with which the data are taken. Here, and also in Fig. 4, error bars are omitted because a sufficient number of samples was taken to ensure that the statistical uncertainty is ≤0.01 (*36*). (**C**) The mean $\bar{C}$ (top) and RMS values δ_C (bottom) of C for different N_{wv}. Dashed lines are computed from the numerically simulated values in (B). (Inset) Numerically computed average numbers of Pauli strings in the time-evolved operator $\hat{O}(t)$, n_p, for the experimental circuits. Dashed line is an exponential fit, $n_p \approx 2^{0.96N_{wv}}$.

This is expected because the time-evolved operator $\hat{O}(t)$ is a single Pauli string and therefore either commutes or anticommutes with $\hat{M}$. As more non-Clifford gates are introduced into the circuits, C starts to assume intermediate values between ±1 and converges toward 0.

The mean $\bar{C}$ and fluctuation [i.e., root mean square (RMS) value] δ_C of C are then computed from experimental data and plotted against N_{wv} in Fig. 3C. We observe different behaviors for $\bar{C}$ and δ_C: $\bar{C}$ remains largely constant and close to 0, confirming that operator spreading remains unaffected by the increasing number of non-Clifford gates. Conversely, δ_C decays from an initial value of 1 and is suppressed by almost two orders of magnitude as N_{wv} increases from 0 to 32, which indicates that $\hat{O}(t)$ expands into a large number, n_p, of weakly correlated Pauli strings with comparable coefficients. This is confirmed by numerically computed values of n_p, which increase exponentially with N_{wv} (inset in Fig. 3C). The growth of n_p necessarily implies substantial operator entanglement (increase in n_B) as well as growth of more–commonly used metrics, such as the operator entanglement entropy [see section VII of (*36*) for a detailed analysis]. These results demonstrate that the decay of OTOC fluctuation allows the growth of operator entanglement to be experimentally diagnosed.

To determine the accuracy of our measurements, the OTOCs of experimental circuits are simulated using a Clifford-expansion method (*36*) and overlaid on the data in Fig. 3, B and C. The close agreement between experiment and simulation motivates us to further explore how our experimental accuracy changes with classical simulation complexity. This is done by systematically increasing the number of iSWAPs in the quantum circuits, N_{iswap} (here, N_{iswap} counts only iSWAP gates that lie within the light cones of Q_a and Q_1). At the same time, N_{wv} is kept at a large value such that the Clifford-expansion simulation method used in Fig. 3 is challenging to perform. Instead, we use tensor contraction to approximately simulate C; these simulations replicate idealized circuits and do not take experimental imperfections into account. Tensor contraction was chosen because it is the best-performing algorithm for simulating Sycamore random circuits with low errors compared with the ideal outcomes (*36*, *38*–*40*). Figure 4A shows representative data for three circuit configurations with different values of N_{iswap} along with the corresponding numerical simulation results. As N_{iswap} and the computational complexity for tensor-contraction increase, we observe that the OTOC fluctuation decreases. The agreement between experiment and simulation also degrades—a result of increased experimental errors, such as qubit decoherence and imperfect circuit inversion, that are not taken into account by the simulation.

To quantify these observations, we define an ideal OTOC signal as the fluctuation δ_C computed from the simulated values of C. We also define the experimental error as the RMS deviation between the simulated and the measured values of C. Both quantities are shown as functions of N_{iswap} in the upper panel of Fig. 4B. The ratio of the two is the experimental signal-to-noise ratio (SNR) for C and is plotted in the bottom panel of Fig. 4B. The estimated time to simulate a single value of C for idealized circuits using tensor contraction on a central processing unit (CPU) core (*36*), t_{sim}, is plotted in Fig. 4C; in estimating t_{sim}, we have only counted the time needed to simulate C above the experimental error shown in Fig. 4B. For $N_{iswap} > 280$, we estimate t_{sim} as the expected time needed to contract the tensor network of the quantum circuit a single time. As N_{iswap} increases, t_{sim} increases and the experimental SNR decreases, reaching SNR ≈ 1 for $N_{iswap} = 251$ ($t_{sim} \approx 1$ hour, or 100 hours for all experimental circuits). Although it is still feasible to simulate our current results, we estimate that the classical simulation of idealized circuits likely becomes intractable when $N_{iswap} > 400$, owing to the exponential scaling of t_{sim}. It is possible that classical algorithms may be further improved if the limited experimental accuracy is taken into account. In sections I and VI of (*36*), we present the analysis and simulation results demonstrating that the observed experimental errors are consistent with the level of noise in our device.

We characterize quantum scrambling in a 53-qubit system and find that the accuracy of quantum processors can be significantly improved through effective error mitigation. For example, an SNR of ~1 for OTOCs is achieved

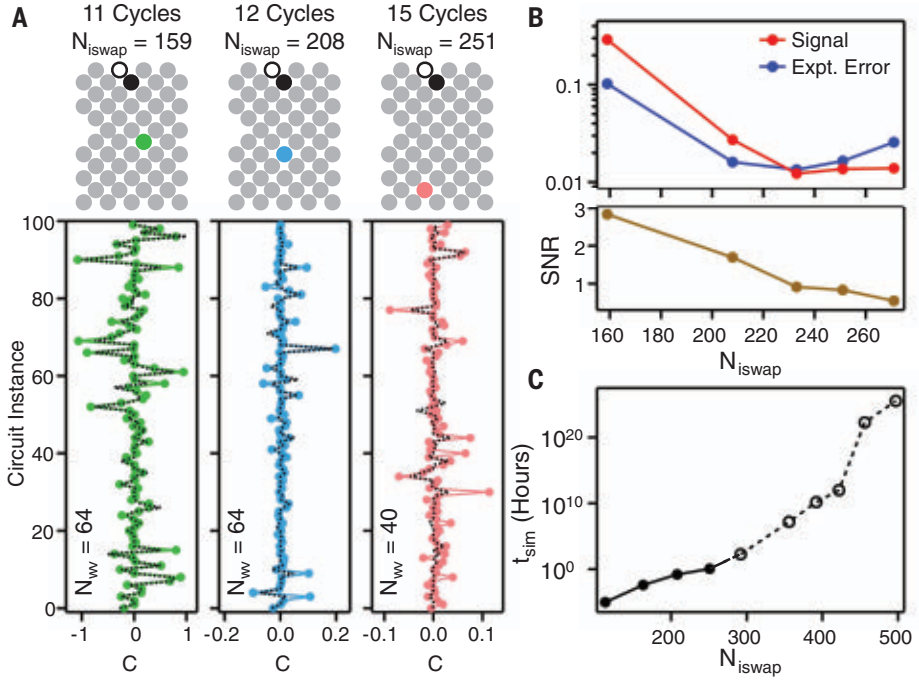


Fig. 4. Classical simulation of quantum scrambling. (**A**) (Top) Three different circuit configurations, each having the same Q_a (black-outlined unfilled circles) and Q_1 (black filled circles) but different Q_b (colored circles) and number of cycles in $\hat{U}$. The number of iSWAPs in $\hat{U}$ and $\hat{U}^\dagger$ that affect classical simulation costs, N_{iswap}, are indicated for each configuration. (Bottom) Instance-dependent OTOCs measured for each configuration. The dashed lines are simulation results of idealized circuits using tensor-contraction methods. N_{wv} is also indicated for each configuration. (**B**) (Top) OTOC signal size and experimental error as functions of N_{iswap}. (Bottom) SNR as a function of N_{iswap}. SNR equals the ratio of OTOC signal size to experimental error. N_{wv} = 64 for the first three values of N_{iswap}, and N_{wv} = 40 for the last two values. The reason for decreasing N_{wv} is to increase the signal size such that it can be resolved with less statistical averaging (*36*). (**C**) Estimated time t_{sim} needed to simulate C of a single 53-qubit idealized circuit with a variable number of iSWAPs, N_{iswap}, on a single CPU core [6 billion floating point operations (Gflop) per second]. Here, the filled and unfilled circles represent circuit sizes at which experimental data are available or unavailable, respectively, for comparison with simulation [see text and (*36*) for details].

for N_{iswap} = 251, where the circuit fidelity is merely 3% (*36*). The average OTOC measurements presented here may be used to efficiently benchmark performances of large quantum processors owing to the efficient classical simulation of operator spreading. On the other hand, the nontrivial computational cost needed to simulate our experimental results, related to operator entanglement, indicates that future quantum processors may be able to shed light on certain classically challenging physical phenomena. Lastly, our experimental protocols may be readily applied to studying other quantum dynamics of interest, such as the 2D XY model or the transverse Ising model (*41*, *42*). We include preliminary experimental measurements of OTOCs in these dynamics in figs. S3 to S5 (*36*). Extending such quantum simulations to larger systems, where they may reveal computationally hard features, is a focus of current research. As the fidelity of quantum processors continues to increase, modeling scrambling in quantum gravity and unconventional quantum phases may become a reality as well (*2*, *3*, *9*, *10*).

REFERENCES AND NOTES

1. P. Hayden, J. Preskill, *J. High Energy Phys.* **2007**, 120 (2007).
2. Y. Sekino, L. Susskind, *J. High Energy Phys.* **2008**, 65 (2008).
3. N. Lashkari, D. Stanford, M. Hastings, T. Osborne, P. Hayden, *J. High Energy Phys.* **2013**, 22 (2013).
4. W. Brown, arXiv:1210.6644 [quant-ph] (2013).
5. S. H. Shenker, D. Stanford, *J. High Energy Phys.* **2014**, 67 (2014).
6. J. M. Deutsch, *Phys. Rev. A* **43**, 2046–2049 (1991).
7. M. Srednicki, *Phys. Rev. E* **50**, 888–901 (1994).
8. M. Rigol, V. Dunjko, M. Olshanii, *Nature* **452**, 854–858 (2008).
9. M. Blake, R. A. Davison, S. Sachdev, *Phys. Rev. D* **96**, 106008 (2017).
10. D. Ben-Zion, J. McGreevy, *Phys. Rev. B* **97**, 155117 (2018).
11. D. M. Basko, I. L. Aleiner, B. L. Altshuler, *Ann. Phys.* **321**, 1126–1205 (2006).
12. J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, H. Neven, *Nat. Commun.* **9**, 4812 (2018).
13. E. Knill *et al.*, *Phys. Rev. A* **77**, 012307 (2008).
14. J. Haferkamp *et al.*, arXiv:2002.09524 [quant-ph] (2020).
15. D. A. Roberts, D. Stanford, L. Susskind, *J. High Energy Phys.* **2015**, 51 (2015).
16. I. L. Aleiner, L. Faoro, L. B. Ioffe, *Ann. Phys.* **375**, 378–406 (2016).
17. D. A. Roberts, B. Yoshida, *J. High Energy Phys.* **2017**, 121 (2017).
18. A. Nahum, S. Vijay, J. Haah, *Phys. Rev. X* **8**, 021014 (2018).
19. C. W. von Keyserlingk, T. Rakovszky, F. Pollmann, S. L. Sondhi, *Phys. Rev. X* **8**, 021013 (2018).
20. T. Rakovszky, F. Pollmann, C. W. von Keyserlingk, *Phys. Rev. X* **8**, 031058 (2018).
21. P. Zanardi, *Phys. Rev. A* **63**, 040304 (2001).
22. Although other classical pictures of operator spreading may arise for k-spin when k is close to infinity (*23*), they do not apply here because qubits are spin-1/2.
23. B. V. Fine, T. A. Elsayed, C. M. Kropf, A. S. de Wijn, *Phys. Rev. E* **89**, 012923 (2014).
24. J. Li *et al.*, *Phys. Rev. X* **7**, 031011 (2017).
25. M. Gärttner *et al.*, *Nat. Phys.* **13**, 781–786 (2017).
26. K. A. Landsman *et al.*, *Nature* **567**, 61–65 (2019).
27. M. S. Blok *et al.*, *Phys. Rev. X* **11**, 021010 (2021).
28. M. K. Joshi *et al.*, *Phys. Rev. Lett.* **124**, 240505 (2020).
29. A. I. Larkin, Y. N. Ovchinnikov, *Sov. Phys. JETP* **28**, 1200 (1969).
30. B. Swingle, G. Bentsen, M. Schleier-Smith, P. Hayden, *Phys. Rev. A* **94**, 040302 (2016).
31. C in some cases can have a nonnegligible imaginary part, which is measured through the $\langle \hat{\sigma}_x \rangle$ of Q_a. For this work, we only measure the real part of C, which is equal to $\langle \hat{\sigma}_y \rangle$ of Q_a.
32. B. Swingle, N. Yunger Halpern, *Phys. Rev. A* **97**, 062113 (2018).
33. B. Bertini, P. Kos, T. Prosen, *Phys. Rev. Lett.* **123**, 210601 (2019).
34. P. W. Claeys, A. Lamacraft, *Phys. Rev. Res.* **2**, 033032 (2020).
35. V. Khemani, D. A. Huse, A. Nahum, *Phys. Rev. B* **98**, 144304 (2018).
36. Supplementary materials are available online.
37. D. Gottesman, arXiv:quant-ph/9807006 [quant-ph] (1998).
38. F. Arute *et al.*, *Nature* **574**, 505–510 (2019).
39. J. Gray, S. Kourtis, *Quantum* **5**, 410 (2021).
40. C. Huang *et al.*, arXiv:2005.06787 [quant-ph] (2020).
41. J. M. Kosterlitz, *J. Phys. C Solid State Phys.* **7**, 1046–1060 (1974).
42. D. S. Fisher, *Phys. Rev. B* **51**, 6411–6461 (1995).
43. Google AI Quantum and collaborators, Data and software code for "Information scrambling in quantum circuits", *Zenodo* (2021); https://doi.org/10.5281/zenodo.5548891.

ACKNOWLEDGMENTS

P.R. and X.M. acknowledge fruitful discussions with P. Zoller, B. Vermersch, A. Elben, and M. Knapp. **Funding:** S.Ma. and J.Ma. acknowledge support from the NASA Ames Research Center and support from the NASA Advanced Supercomputing Division for providing access to the NASA HPC systems, Pleiades and Merope. S.Ma. and J.Ma. also acknowledge support from the AFRL Information Directorate under grant no. F4HBKC4162G001. J.Ma. is partially supported by NAMS contract no. NNA16BD14C. S.Ma. is also supported by the Prime contract no. 80ARC020D0010 with the NASA Ames Research Center. **Author contributions:** V.Sm., K.K., X.M., and P.R. devised the experiment. X.M., C.Q., and P.R. executed the experiment on the Google quantum hardware. X.M., P.R., K.K., and Y.C. wrote the manuscript. X.M., S.Ma., J.Ma., and K.K. wrote the supplementary materials. V.Sm., I.A., X.M., K.K., S.Ma., and J.Ma. provided theoretical support, analysis techniques, and numerical computations. I.A. and K.K. developed the Markov process model. S.Ma. designed and performed the large-scale numerical simulation, including the algorithms and software development. J.Ma. performed the noisy numerical simulations. P.R., Y.C., V.Sm., and H.N. led and coordinated the project. Infrastructure support was provided by the Google Quantum AI hardware team. The NASA Advanced Supercomputing Division at NASA Ames provided the infrastructure to run high-performance computing (HPC) simulations. All authors contributed to revising the manuscript and the supplementary materials. **Competing interests:** The authors declare no competing interest. **Data and materials availability:** All experimental and numerical data in the main text and supplementary materials, along with the software code for generating quantum circuits, measurements, population dynamics simulation, and tensor contraction simulation are available at Zenodo (*43*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg5029
Materials and Methods
Supplementary Text
Figs. S1 to S29
References (*44*–*75*)

9 January 2021; accepted 19 October 2021
Published online 28 October 2021
10.1126/science.abg5029

MAGNETISM

Atomic-scale visualization of topological spin textures in the chiral magnet MnGe

Jacob Repicky[1], Po-Kuan Wu[1], Tao Liu[1,2], Joseph P. Corbett[1], Tiancong Zhu[1], Shuyu Cheng[1], Adam S. Ahmed[1], N. Takeuchi[3], J. Guerrero-Sanchez[3], Mohit Randeria[1], Roland K. Kawakami[1], Jay A. Gupta[1]*

Topological spin textures in chiral magnets such as manganese germanide (MnGe) are of fundamental interest and may enable magnetic storage and computing technologies. Our spin-polarized scanning tunneling microscopy images of MnGe thin films reveal a variety of textures that are correlated to the atomic-scale structure. Our images indicate helical stripe domains, in contrast to bulk, and associated helimagnetic domain walls. In combination with micromagnetic modeling, we can deduce the three-dimensional (3D) orientation of the helical wave vectors, and we find that three helical domains can meet in two distinct ways to produce either a "target-like" or a "π-like" topological spin texture. The target-like texture can be reversibly manipulated through either current/voltage pulsing or applied magnetic field, which represents a promising step toward future applications.

Topological spin textures in chiral magnets are of interest both to fundamental science, through Berry phase–induced Hall effects, and for potential device applications, including magnetic racetrack memories and neuromorphic computing (*1*–*8*). Competing magnetic interactions lead to spin textures such as helices, where spins periodically tumble with a characteristic pitch length, and magnetic skyrmions, which are whirling localized textures with a total π rotation of the spins. In many cases of interest, the competition is between ferromagnetic exchange, which favors aligned spins, and the Dzyaloshinskii-Moriya interaction (DMI), which favors perpendicular spins and arises from spin-orbit coupling in the presence of broken inversion symmetry (*9*). The noncentrosymmetric "B20" crystal structure breaks bulk inversion symmetry, and magnetic skyrmions were first discovered in B20 MnSi (*9*) and FeGe (*10*). In these materials, the magnetic phase diagram and its evolution from bulk crystals to thin films are now well understood. Within the B20 family, MnGe is an intriguing outlier (*11*), with a helical pitch length of 2.8 nm that is more than an order of magnitude smaller than in other B20 crystals (*12*) and whose bulk phase diagram shows unusual "hedgehog-antihedgehog" crystals, for reasons that are not well understood (*13*–*15*). Spin-polarized scanning tunneling microscopy (SP-STM) is uniquely suited to probe the rich physics of such nanoscale spin textures in real space and provides microscopic insights that complement those obtained from ensemble techniques that may average over different chiral domains (*16*).

[1]Department of Physics, The Ohio State University, Columbus, OH 43210, USA. [2]University of Electronic Science and Technology of China, Chengdu 610054, China. [3]Centro de Nanociencias y Nanotecnologia, Universidad Nacional Autónoma de México, Apartado Postal 14, Ensenada Baja California, Código Postal 22800, Mexico.
*Corresponding author. Email: gupta.208@osu.edu

We used SP-STM to probe the magnetism on the surface of 80-nm-thick MnGe(111) films with atomic resolution. Our SP-STM images revealed a variety of topological spin textures depending on the local nanoscale structure, which we interpreted using micromagnetic modeling that builds on recent advances in the theory of helimagnetic domain walls (*17*). Figure 1, A and B, shows atomically resolved topographic images of the MnGe(111) surface. These images are consistent with the B20 structure of MnGe, which features alternating Mn and Ge layers with atoms arranged in triangular lattices of single atoms or trimers (*18*). The structure comprises a quadruple layer of alternating planes of Mn and Ge atoms with sparse or dense packing; the atomic arrangements exactly repeat after a sequence of three such layers. The relative stacking order and orientation of these layers determine the structural and magnetic chiralities (*19*–*22*). The surface lattice constant from these images is 0.67 ± 0.01 nm, within experimental uncertainty of the expected bulk value for the (111) surface (0.678 nm). Point defects on the surface are imaged with bright and dark contrast, but they do not affect the magnetic textures reported here.

In addition to the topographic information, Fig. 1, A and B, shows a subtle (~5 pm) periodic modulation of the atomic corrugation, reflecting the surface magnetic texture picked up by the SP-STM tip. This modulation is evident in the topographic linecut shown in Fig. 1C. To better isolate the stripe pattern from the topography in Fig. 1A, we performed a fast Fourier transform (FFT) (Fig. 1D). In addition to the primary hexagonal spot pattern from the MnGe atomic lattice, there are satellite spots corresponding to scattering vectors of $\pm\mathbf{q}$, rotated by $\phi \sim 14°$ with respect to the lattice. An inverse FFT image of the area, computed with only the atomic lattice and satellite spots, clearly resolves the stripe pattern while removing obscuration from the point defects (Fig. 1E). In bulk MnGe crystals, a 3D hedgehog lattice was observed with Lorentz transmission electron microscopy (LTEM) (*12*), which would yield a 2D lattice projected onto the surface, in contrast to the observed stripe pattern here (*18*). Furthermore, from the FFT analysis we measure a stripe period of 5.96 nm, which is considerably larger than the helical pitch length for bulk MnGe (2.8 nm).

This stripe pattern is consistent with a $1\mathbf{Q}$ helical state in these MnGe thin films, in contrast to the $3\mathbf{Q}$ state observed in bulk crystals. A priori, one could explain the stripe contrast with helices anchored to the surface plane as reported for FeGe (*17*), but this is contradicted by the larger observed pitch length. Prior neutron scattering studies in MnGe thin films indicate that the magnitude of $Q = 2.2\ \mathrm{nm}^{-1}$ is unchanged from the bulk value (*23*), and we expect that any additional surface-specific effects, such as reduced exchange or enhanced surface DMI, would lead to an even smaller pitch length, in contrast to the observation. Instead, we consider the tilting of $\mathbf{Q}$ toward the film normal [111] direction by a polar angle θ, which was invoked in the neutron studies (*23*). Following conventions for topological defects in chiral magnets (*24*, *25*), states with wave vectors $\mathbf{Q}$ and $-\mathbf{Q}$ correspond to the same helical structure, and we chose a positive projection along the surface normal $\hat{z} = (111)$. We defined the surface wave vector $\mathbf{q}$ as the projection of $\mathbf{Q}$ in the plane of the surface (i.e., $q = Q \sin\theta$), so that the polar tilt angle can be related to the observed real-space periodicity by $\theta = \sin^{-1}[2\pi/(Q \cdot 5.96\ \mathrm{nm})] = 28.6°$, compared to the bulk angle of 54.7° with $\mathbf{Q}$s along (100). Our estimated q is roughly consistent with the neutron studies, where a linearly decreasing tilt angle with decreasing film thickness down to 160 nm was attributed to strain-dependent magnetic anisotropy (*23*).

To directly probe the sensitivity of the spin helices to strain in real space, we imaged regions of the film where small curvatures are indicative of inhomogeneous strain. For example, in a different microscopic region of the sample shown in Fig. 2A, three terraces were observed, separated in height by steps of one quadruple layer each in the layered MnGe structure. Focusing on the middle terrace, topography line profiles show small (<0.1%) but significant bowing and curvature of the surface along the horizontal direction (Fig. 2B, blue profile). We note that these images are atomically resolved and the lattice spacing does not show any significant variation, but our experimental uncertainty in this measurement (~0.5%) is larger than the 0.1% height variation shown in the image.

To examine the stripe patterns over larger distances in such areas, we simultaneously mapped the differential conductance signal,

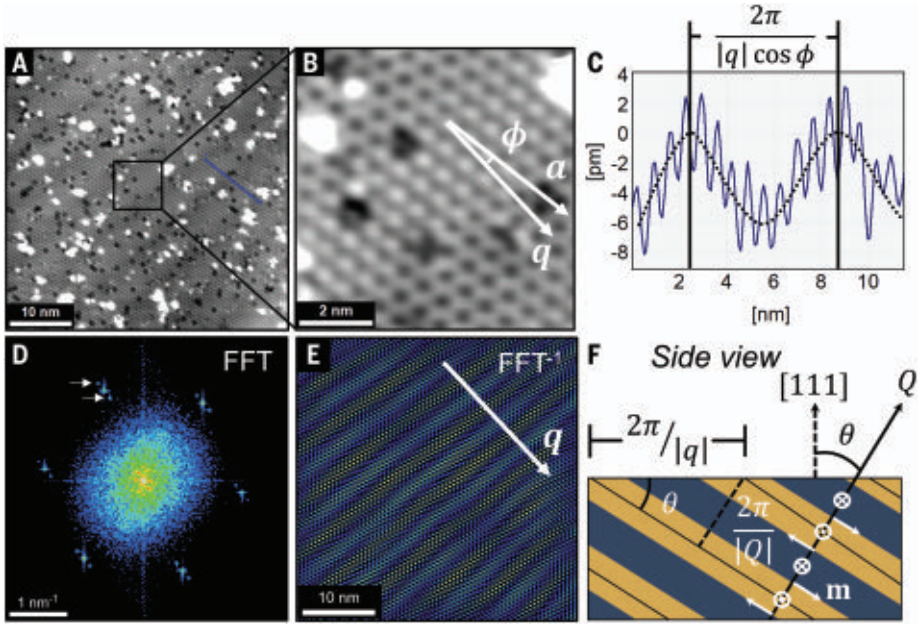


Fig. 1. Atomic-resolution SP-STM imaging of spin helices in MnGe. (**A**) Topographic image of the MnGe(111) surface (−0.17 V, 0.54 nA, T = 5 K). The helical magnetic texture is imaged as modulation in the atomic corrugation. (**B**) Close-up view of the outlined area in (A). Bright and dark lattice spots are observed, depending on the relative alignment between tip and surface spins. In this area, the surface projection of the helical wave vector **q** is rotated from the atomic lattice vector by an angle $\phi \approx 14°$. (**C**) Line profile taken at the blue line in (A) along nearest-neighbor atoms on the surface, showing the helical periodicity of 5.96 nm. The dotted line is a guide to the eye. (**D**) FFT image of the area in (A). Arrows indicate satellite peaks due to the surface-projected **q**. (**E**) Inverse FFT image produced by passing only the atomic lattice and satellite peaks in (D), allowing an unobstructed view of the stripe pattern and its effect on the atomic corrugation. (**F**) Schematic side view of the 3D helical texture showing the relation among **Q**, **q**, and the real-space modulations.

which relates to the spin-dependent density of states. Our SP-STM simulations (fig. S17) explore the expected magnetic contrast for textures with varying **Q** and tip magnetization vector $\mathbf{m}_{tip}$. Because the textures here represent a helical whirling of spins with both in-plane and out-of-plane components, magnetic contrast occurs for any combination of **Q** and $\mathbf{m}_{tip}$, except for the special case $\mathbf{m}_{tip} \parallel \mathbf{Q}$ where there is zero contrast. Figure 2C shows an SP-STM map of the area, where faint magnetic contrast reveals stripes along different directions and more complicated patterns. During repeated imaging of this area, we observed a complete reversal in magnetic contrast associated with an inversion of the tip's spin polarization (Fig. 2D). To confirm these stripes as magnetic in origin, we computed a difference image in Fig. 2E, as topographic or electronic contributions to the STM image would not invert under otherwise identical tunneling conditions. Whereas the magnetic contrast is absent in the sum image (fig. S16), the difference image accentuates intersections, bowing, and terminations of the stripe patterns in this area. We attribute these spin textures to local variations in the orientation of **Q**. For example, the observed stripe periodicity in Fig. 2E varies in the range of 6 to 10 nm depending on position, corresponding to respective variations $28° > \theta > 17°$ in the polar angle of **Q**. The stripe curvature and intersection points furthermore indicate helical domains with distinct azimuthal angles of **Q**.

To understand these features, we used a phenomenological model that builds on recent advances in the theory of topological domain walls in helimagnets (*17*) and uses inputs from neutron data (*23*) to constrain the magnetic anisotropy and hence the orientations of the helical wave vectors (*18*). The structure of a domain wall between two helical regions depends primarily on the angle θ_{12} between their wave vectors $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Three fundamental types of helical domain walls have been reported recently in magnetic force microscopy imaging of B20 FeGe (*17*). Where $\theta_{12} \lesssim 85°$, "type I" walls are observed, which are smooth and free of disclination defects or phase mismatch. We found that for the parameters relevant for our MnGe films, the type I domain walls are energetically preferred.

Our micromagnetic modeling shows a distortion of the helices near the domain wall (Fig. 2, F and G). There are two characteristic surface projections depending on whether the $\mathbf{Q}_{1,2}$ vectors are oriented toward or away from the intersection domain wall plane. For **Q**s oriented toward the wall, the intersection plane is characterized by series of sharp, nested vertices along the domain wall (Fig. 2F). In contrast, for **Q**s oriented away from the wall, the domain wall is characterized by a nesting of more gradual, curved helices (Fig. 2G). In both cases, the surface projections $\mathbf{q}_1$, $\mathbf{q}_2$ make an in-plane angle $\phi_{12} = 120°$ independent of θ_{12}. The difference between these projections reflects rounding of the helical stripes in proximity to the domain wall and the surface.

In good agreement with our modeling, both projections of type I domain walls are observed experimentally (Fig. 2E, dashed boxes). The domain wall in the left box shows the nested, sharp vertex-like structure expected from Fig. 2F. By considering the 3D nature of $\mathbf{Q}_i$, we can extract the angle θ_{12} geometrically using

$$\cos\theta_{12} = \frac{Q_1 \cdot Q_2}{|Q_1||Q_2|} = \sin\theta_1 \sin\theta_2 \cos\phi_{12} + \cos\theta_1 \cos\theta_2 \quad (1)$$

where $\theta_1 = 26°$, $\theta_2 = 19°$ are the polar angles calculated in each domain from the period of the stripes, and $\phi_{12} = 113°$ is estimated as the angle between the stripe patterns on either side of the domain wall. (The simplest model that ignores surface effects predicts a 120° angle, as noted above.) This then gives an angle $\theta_{12} = 37°$ between $\mathbf{Q}_1$ and $\mathbf{Q}_2$ in this region, which is within the established regime for a type I domain wall (*17*). The right box in Fig. 2E shows the other surface projection of a type I domain wall, characterized by a nesting of rounded helical stripes, and can be analyzed in a similar way to give $\theta_{12} = 30°$, also within the type I regime.

More complex magnetic textures can be found at the intersections of domain walls. Our modeling shows that the intersection of two domain walls must necessarily involve at least one wall that is of type II or type III, which are energetically unfavorable and not observed experimentally (fig. S6). We found, however, that three type I domain walls can meet along an axis perpendicular to the surface and can lead to two distinct spin textures depending on the orientations of the $\mathbf{Q}_i$s. The spin texture in Fig. 3A results when all three **Q**s are oriented toward or away from the intersection axis, and exhibits a core region that is wrapped with closed helical loops. These textures closely resemble topological defects known as "target" states or 2π-disclination defects (*26*, *27*). The second "π" texture results from the arrangement of $\mathbf{Q}_i$ as shown in Fig. 3B. Both of these textures have nonzero topological charge density, concentrated in the vicinity of the domain walls, which

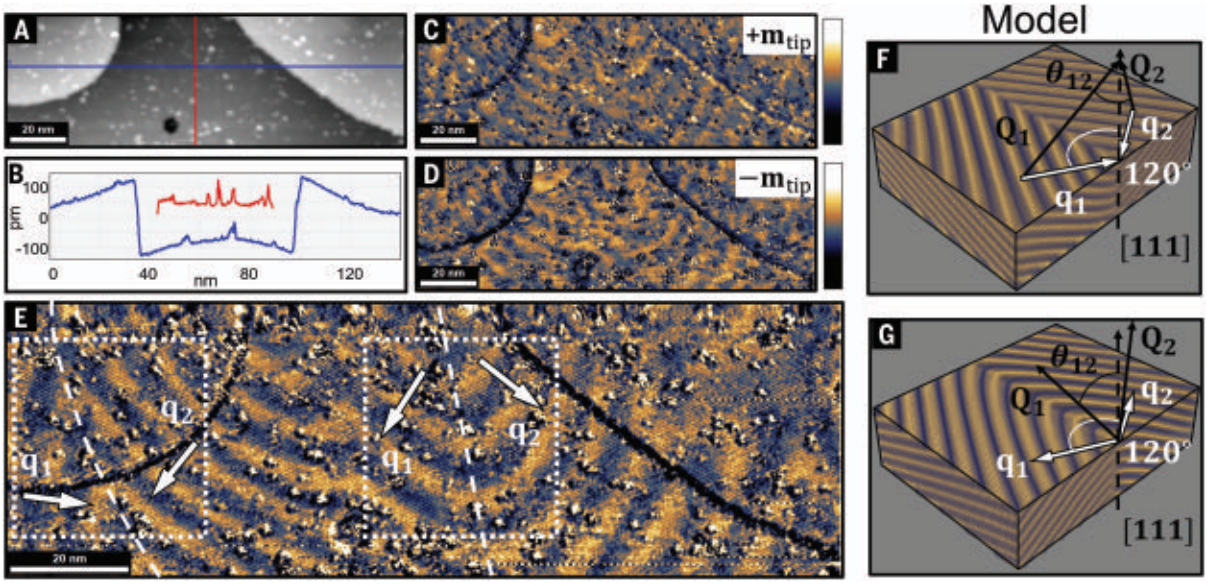
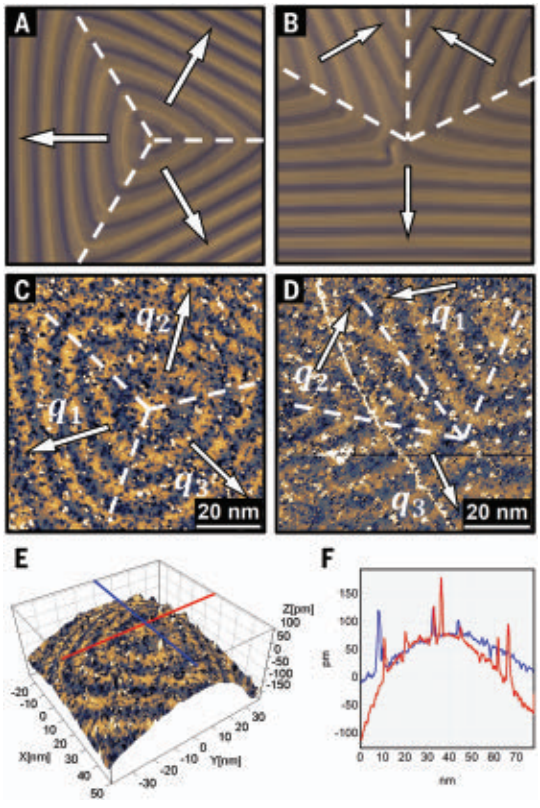


Fig. 2. Spectroscopic imaging of helical domains in a bowed region of the surface. (**A**) Topographic image of a bowed region of the surface containing three atomic terraces (0.17 V, 1.0 nA, T = 5 K). (**B**) Line profiles taken along the blue and red lines in (A). (**C** and **D**) Subsequent *dI/dV* images of the same region as in (A) (0.17 V, 1.0 nA). The contrast of the helical texture is subtly visible and inverts upon reversal of the tip magnetization vector $\mathbf{m}_{tip}$. (**E**) Difference image [(C) − (D)] showing a variety of helical textures. Two type I domain walls are boxed and indicated by the dashed lines. (**F**) Micromagnetic model of a domain wall where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are separated by an angle θ_{12} and point toward the intersection plane. The surface projections ($\mathbf{q}_1$, $\mathbf{q}_2$) intersect at a domain wall showing a nesting of sharp vertices. (**G**) Micromagnetic model of a domain wall where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ point away from the intersection plane, resulting in a nesting of more rounded helices.

Fig. 3. Modeling and observation of target and π textures. (**A**) Micromagnetic model of the target spin texture, with arrows indicating surface **q**s pointing away from the intersection axis. This results in rounded triangular rings wrapping a central core. (**B**) Model of the π texture, where two **q** vectors point inward toward each other and the third points away, resulting in two rounded domain walls meeting a sharp domain wall. (**C**) SP-STM image of the target texture (−0.31 V, 0.22 nA). (**D**) SP-STM image of a π texture (−0.31 V, 0.20 nA). The SP-STM images in (C) and (D) are shown without additional processing. (**E**) Three-dimensional view of the area hosting the target texture, showing curvature of the surface. Topographic information is shown along the *z* axis and the color scale is a *dI/dV* overlay from the image shown in (C). (**F**) Line profiles taken along the *x* (blue) and *y* (red) directions in (E) to show in more detail the curvature of the surface in this area.



oscillates in sign as one moves outward from the core of the texture (fig. S9). In the absence of a well-defined boundary, however, neither the target nor the π texture have a quantized topological charge. Our modeling also shows that the core of the target texture consists of a string of alternating hedgehogs and anti-hedgehogs that is oriented perpendicular to the surface (fig. S10).

Experimentally, we found both the target and π textures in a region of the film where there was also nanoscale curvature. Figure 3C shows the conjunction of three type I walls to form the target texture, with closed helical loops wrapping around a central ~10-nm core. The isometric topographic image (Fig. 3E) and topographic linecuts (Fig. 3F) of this area indicate that the core is localized to the region of convex curvature to within a few nanometers in both the horizontal and vertical directions, although there was some variation in this proximity among the other target textures observed (fig. S14). Figure 3D shows the π texture in an adjacent region spanning two terraces separated by an atomic step across the middle. The local curvature in this region is slightly concave and connects adjoining regions with target textures and convex curvature (fig. S14).

Applications of magnetic skyrmions rely on the ability to manipulate these spin textures, which has been demonstrated with a variety of stimuli (*28*–*31*), including STM pulsing (*30*) and pressure (*31*). We found that the target texture can be similarly manipulated by local current/voltage pulses. Figure 4, A to D, shows a sequence of SP-STM images where current/voltage pulses were applied to the core region using the STM tip. The initial state of the target texture in Fig. 4A displayed a bright core and several surrounding closed loops. After imaging, the STM tip was positioned over the core and a short (0.5 s) current/voltage

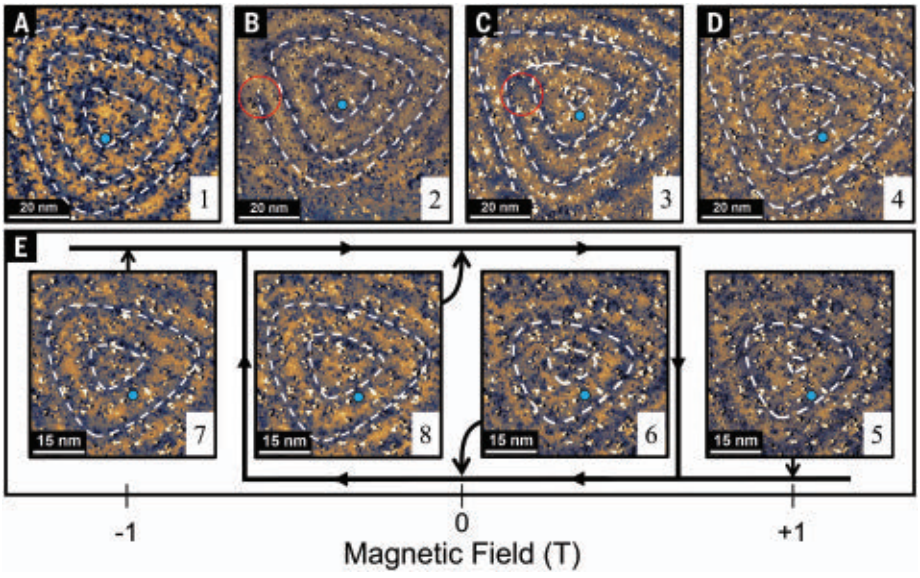


Fig. 4. STM tip and magnetic field manipulation of a target texture. (**A** to **D**) SP-STM images of the target spin texture in different configurations. Between each image, current/voltage pulses were applied with the STM tip (~2.0 V, 0.5 s, $I \lesssim 1\,\mu$A due to 300-pm approach). White dashed lines are guides to the eye showing wrapping around the core region. The red circle in (B) and (C) indicates a disclination defect that is generated, moved, and then annihilated. The blue dot represents the same atomically registered fixed point in the images, showing motion of the core. (**E**) SP-STM images showing the magnetic field dependence of the texture. Image 5, taken after image 4 with an out-of-plane magnetic field of +1 T, shows reversed contrast. Subsequent imaging (images 6 to 8) was performed in the indicated loop, with a clear hysteresis effect evident upon comparison of images 6 and 8. All images were taken under conditions of −0.3 V, 0.2 nA, and T = 5 K and are shown without any additional processing.

pulse was applied. A new disclination defect (branch point) appeared in the subsequent SP-STM image (Fig. 4B, red circle), which represents a discrete change in the topological charge density within this region. The core center also shifted by ~5 nm with respect to the fixed background of atomic defects. Subsequent pulsing shifted the disclination defect closer to the core, which itself changed contrast and size (Fig. 4C). A final cycle of pulsing annihilated the disclination defect and further shifted the core (Fig. 4D). Although this series of images was taken with different atomic tip terminations (and thus possibly different magnetic contrast), the topology of the final state was identical to the starting state (Fig. 4A). This indicates that the target texture had been reversibly manipulated through a landscape of metastable topological states.

We also observed hysteretic behavior of the target texture with applied out-of-plane magnetic field, as shown in Fig. 4E. After acquiring the SP-STM image at zero field (image 4 in Fig. 4D), the magnetic field was ramped up to +1 T. As shown in SP-STM image 5 in Fig. 4E, the core region is reduced in size and shows faintly dark contrast (dashed lines). Ramping the field back down to 0 T (image 6) shows an expansion of the wrapping and increased dark contrast of the core region. The core region flips back to bright contrast and shrinks when the field is ramped to −1 T (image 7), and then expands again when the field is ramped back down to 0 T (image 8). Comparing images 4 and 6 reveals a clear hysteresis effect, and the identical magnetic contrast in images 4 and 8 indicate that this effect is reversible. Consistent with previous SP-STM studies using bulk Cr tips (*30*), this hysteretic behavior cannot be explained as a field-dependent polarization of the tip. In fig. S17 we show SP-STM simulations indicating that if the tip polarization vector were changing with applied field, there would be changes in stripe contrast and/or anisotropy, which are not observed in Fig. 4E. Instead, our micromagnetic modeling explains the contrast reversal as arising from a net magnetic moment associated with the finite-volume texture, which aligns with the applied magnetic field. A π phase shift is induced when the magnetic field switches direction, leading to a reversal of contrast in the SP-STM images (fig. S10).

The topological spin textures observed in our thin films are distinct from those in bulk samples, which raises questions concerning the interplay of bulk and surface magnetism (*32*). How, or whether, the textures we observed extend into the bulk of the films can be further explored by comparing bulk measurements of topological Hall effect with magnetic imaging techniques such as (surface-sensitive) SP-STM and (volumetric) Lorentz transmission electron microscopy. Although additional study is needed to establish quantitative correlation, the association of these textures with local strain may reflect an interplay of magnetostriction effects, which have been studied in B20 materials (*33*–*35*), and further advances strain as an additional tuning parameter in thin-film devices for future memory and logic applications (*12*, *31*, *36*).

REFERENCES AND NOTES

1. A. Fert, N. Reyren, V. Cros, *Nat. Rev. Mater.* **2**, 17031 (2017).
2. A. Soumyanarayanan, N. Reyren, A. Fert, C. Panagopoulos, *Nature* **539**, 509–517 (2016).
3. R. Wiesendanger, *Nat. Rev. Mater.* **1**, 16044 (2016).
4. C. H. Back *et al.*, *J. Phys. D* **53**, 363001 (2020).
5. K. M. Song *et al.*, *Nat. Electron.* **3**, 148–155 (2020).
6. S. S. P. Parkin, M. Hayashi, L. Thomas, *Science* **320**, 190–194 (2008).
7. R. Tomasello *et al.*, *Sci. Rep.* **4**, 6784 (2014).
8. A. Fert, V. Cros, J. Sampaio, *Nat. Nanotechnol.* **8**, 152–156 (2013).
9. S. Mühlbauer *et al.*, *Science* **323**, 915–919 (2009).
10. X. Z. Yu *et al.*, *Nat. Mater.* **10**, 106–109 (2011).
11. N. Kanazawa *et al.*, *Phys. Rev. B* **86**, 134425 (2012).
12. T. Tanigaki *et al.*, *Nano Lett.* **15**, 5438–5442 (2015).
13. M. Bornemann *et al.*, *J. Phys. Condens. Matter* **31**, 485801 (2019).
14. T. T. J. Mutter, A. O. Leonov, K. Inoue, *Phys. Rev. B* **100**, 060407 (2019).
15. Y. Fujishiro, N. Kanazawa, Y. Tokura, *Appl. Phys. Lett.* **116**, 090501 (2020).
16. M. Deutsch *et al.*, *Phys. Rev. B* **90**, 144401 (2014).
17. P. Schoenherr *et al.*, *Nat. Phys.* **14**, 465–468 (2018).
18. See supplementary materials.
19. J. P. Corbett *et al.*, *ACS Appl. Mater. Interfaces* **12**, 9896–9901 (2020).
20. S. V. Grigoriev *et al.*, *Phys. Rev. Lett.* **110**, 207201 (2013).
21. M. Trabel *et al.*, *J. Appl. Phys.* **121**, 245310 (2017).
22. D. Morikawa *et al.*, *Phys. Rev. Mater.* **4**, 014407 (2020).
23. N. Kanazawa *et al.*, *Phys. Rev. B* **96**, 220414 (2017).
24. F. Li, T. Nattermann, V. L. Pokrovsky, *Phys. Rev. Lett.* **108**, 107203 (2012).
25. T. Nattermann, V. L. Pokrovsky, *J. Exp. Theor. Phys.* **127**, 922–932 (2018).
26. F. Zheng *et al.*, *Phys. Rev. Lett.* **119**, 197205 (2017).
27. D. Cortés-Ortuño *et al.*, *Phys. Rev. B* **99**, 214408 (2019).
28. W. Jiang *et al.*, *Science* **349**, 283–286 (2015).
29. N. Ogawa, S. Seki, Y. Tokura, *Sci. Rep.* **5**, 9552 (2015).
30. N. Romming *et al.*, *Science* **341**, 636–639 (2013).
31. Y. Nii *et al.*, *Nat. Commun.* **6**, 8539 (2015).
32. Y. Fujishiro *et al.*, *Nat. Commun.* **12**, 317 (2021).
33. E. Franus-Muir, M. L. Plumer, E. Fawcett, *J. Phys. C* **17**, 1107–1114 (1984).
34. A. E. Petrova, S. M. Stishov, *Phys. Rev. B* **94**, 020410 (2016).
35. Y. Hu, B. Wang, *New J. Phys.* **19**, 123002 (2017).
36. Y. Fujishiro *et al.*, *Nat. Commun.* **10**, 1059 (2019).
37. MnGe Data repository; https://doi.org/10.5281/zenodo.4013735.

ACKNOWLEDGMENTS

We thank S. Mueller for helpful discussions on the data analysis. **Funding:** Primary support was provided by DARPA grant D18AP00008. Partial support was provided by US Department of Energy grant DE-SC0016379 for SP-STM technique development. J.G.S. and N.T. thank DGAPA-UNAM projects IN101019 and IA100920 and CONACyT grant A1-S9070 of the Call of Proposals for Basic Scientific Research 2017–2018 for partial financial support. Calculations were performed in the DGCTIC-UNAM Supercomputing Center, project LANCAD-UNAMDGTIC-368. **Author contributions:** J.R. and J.P.C. performed SP-STM experiments and analyzed data; T.L.,T.Z., S.C., and A.S.A. grew the thin films and characterized the structural and magnetic properties; P.-K.W. and M.R. performed theoretical modeling; N.T. and J.G.-S. performed density functional theory calculations for the surface structural study; and R.K.K., M.R., and J.A.G. helped in writing the manuscript and analyzing the results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** See (*37*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd9225
Materials and Methods
Supplementary Text
Figs. S1 to S19
References (*38*–*43*)
Movie S1

20 July 2020; accepted 19 October 2021
10.1126/science.abd9225

PEROVSKITES

Synthesis of $LaWN_3$ nitride perovskite with polar symmetry

Kevin R. Talley[1,2], Craig L. Perkins[1], David R. Diercks[2], Geoff L. Brennecka[2]*, Andriy Zakutayev[1]*

Oxide materials with the perovskite structure have been used in sensors and actuators for half a century, and halide perovskites transformed photovoltaics research in the past decade. Nitride perovskites have been computationally predicted to be stable, but few have been synthesized, and their properties remain largely unknown. We synthesized and characterized a nitride perovskite lanthanum tungsten nitride ($LaWN_3$) in the form of oxygen-free sputtered thin films, according to spectroscopy, scattering, and microscopy techniques. We report a large piezoelectric response measured with scanning probe microscopy that together with synchrotron diffraction confirm polar symmetry of the perovskite $LaWN_3$. Our $LaWN_3$ synthesis should inspire growth of other predicted nitride perovskites, and measurements of their properties could lead to functional integration with nitride semiconductors for microelectromechanical devices.

Nitride materials are revolutionizing the way humans access information and communicate with others. For example, 4th-generation (4G) wireless networks feature piezoelectric aluminum nitride (AlN) film bulk acoustics resonators (FBARs). Radio-frequency (RF) transistors based on semiconducting gallium nitride (GaN) are becoming an important part of the 5G telecommunication technology. The emerging telecommunication infrastructure would further benefit from improved piezoelectric materials, especially if they are easy to integrate with nitride semiconductors. We synthesized lanthanum tungsten nitride ($LaWN_3$) thin films with a perovskite crystal structure, polar symmetry, and strong piezoelectric response. Synthesis of this nitride member of the broad family of perovskite structured materials with ABX_3 stoichiometry (for example, oxides, halides, and chalcogenides) suggests that other computationally predicted nitride perovskites with useful properties should be also possible to synthesize.

Materials with the perovskite crystal structure (Fig. 1A) are arguably the single most famous class of oxide compounds (*1*). Oxide perovskites, such as $Pb(Zr,Ti)O_3$ (PZT) and $(Ba,Sr)TiO_3$ (BST), with strong piezoelectric response have been extensively used for ceramic capacitors (*2*), microelectromechanical actuators (*3*), electrochemical cells (*4*), and many other applications (*5*, *6*) over the past century (*7*). In the past decade, research activity on halide (X = Cl, Br, or I) perovskites, such as $CH_3NH_3PbI_3$ and $CsPbI_3$, has skyrocketed because of their potential application as inexpensive and efficient optoelectronic devices (*8*). Giant optical anisotropy and nonlinear optics applications have recently attracted attention to chalcogenide (X = S or Se) perovskites, such as $BaTiS_3$ or $SrTiS_3$ (*9*). In contrast to oxides, chalcogenides, and halides, very few experimental reports of nitride perovskites exist in crystallographic databases or the literature (Fig. 1B). The few reported perovskites with high N content include powder $TaThN_3$ synthesized from oxide precursors (*10*) and powder $LaReN_3$ synthesized from azide precursors (*11*). Other known intermetallic materials, such as Mg_3SbN and Mn_3CuN, have an anti-perovskite structure and low N content (*12*). The relatively low number of reported nitride perovskites is surprising because pnictide (X = N or P) ABX_3 materials, including perovskites and others, are statistically more likely than the halide ABX_3 materials (Fig. 1B) because of a larger possible number of cation combinations that satisfy ~9 collective anion valence versus the combinations for satisfying ~3 collective anion valence (table S1). This leads us to the question of how to discover nitride perovskites and evaluate their potential properties.

Computationally driven experimental discovery is an effective approach to predict and synthesize new materials (*13*). In the field of nitrides, we recently synthesized ~10 ternary nitride materials out of ~200 computational predictions (*14*). Among nitride perovskites, $LnMN_3$ (Ln=La, Ce, Eu, Yb, M=W, and Re) materials were predicted by other groups to be thermodynamically stable (*15*, *16*), with lanthanum tungsten nitride ($LaWN_3$) having large 1750 meV/formula unit (f.u.) formation enthalpy. $LaWN_3$ was also predicted to have ferroelectric properties, with a large 61 $\mu C/cm^2$ spontaneous polarization and a small 110-meV barrier to polarization reversal (*17*). However, synthesis of $LaWN_3$ by using traditional bulk solid-state chemistry methods remains challenging (*18*), often leading to oxynitrides, such as $LaWO_{0.6}N_{2.4}$ (*19*). If pure nitride perovskites can be synthesized, a century of

[1]Materials Science Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA. [2]Department of Metallurgical and Materials Engineering, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA.
*Corresponding author. Email: andriy.zakutayev@nrel.gov (A.Z.); geoff.brennecka@mines.edu (G.L.B.)

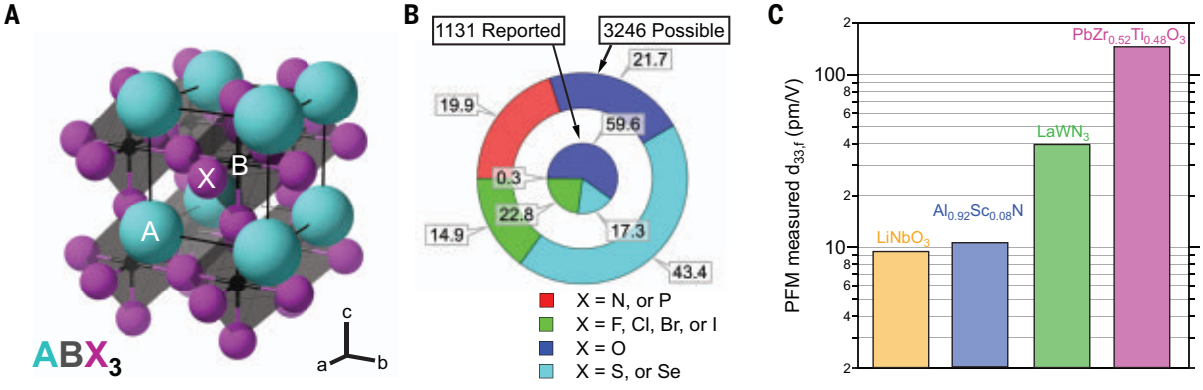


Fig. 1. Nitrides and other materials with the perovskite structure. (**A**) Cubic ABX_3 perovskite unit cell showing the larger A cation sites and smaller B cation sites in BX_6 octahedra. (**B**) Comparison of anion diversity in possible charge-balanced versus experimentally reported ABX_3 compounds, showing that pnictide perovskites are more likely but less reported than halide perovskites. (**C**) Comparison of piezoelectric materials measured in this work, demonstrating strong piezoelectric response of $LaWN_3$.

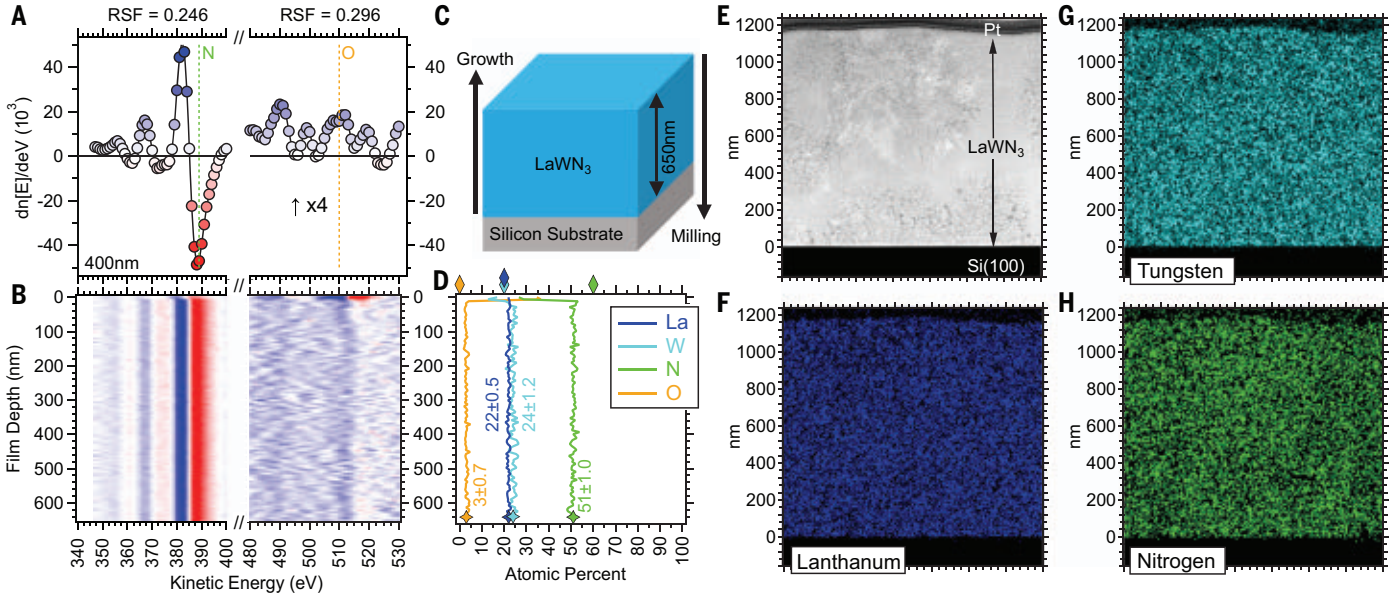


Fig. 2. Chemical composition of $LaWN_3$ thin films. (**A**) Differentiated AES results and (**B**) depth-resolved color intensity map for O and N, showing negligible O signal beyond a thin surface layer, where RSF is a relative sensitivity factor. (**C**) Depth profile and (**D**) resulting element concentration for all elements, with the average and ideal composition indicated with stars and diamonds at the bottom and top, respectively. (**E** to **H**) STEM and EDX images, showing a polycrystalline microstructure and chemical homogeneity of La, W, and N in $LaWN_3$ thin films.

experience in perovskite property engineering (*7*) could be combined with decades of advances in nitride semiconductor integration (*20*). This would likely have an enormous impact on both fundamental and applied research. We synthesized $LaWN_3$ nitride perovskite with polar symmetry and measured a piezoelectric response comparable with that of oxide perovskites and much greater than that of other known nitrides (Fig. 1C).

We used physical vapor deposition (combinatorial cosputtering) to synthesize crystalline $LaWN_3$ thin films on a heated substrate in ultrahigh vacuum to minimize O contamination, and with a N plasma source to maximize N incorporation (*21*). We detected no measurable O (below 3%) throughout the thickness of films beyond a thin nanometer-scale surface oxide layer (Fig. 2, A to D) by using Auger electron spectroscopy (AES), even after 72 hours of atmospheric exposure. These measurements were performed for a sample with the cation-stoichiometric composition (La/W = 1) as determined with x-ray fluorescence (XRF). However, our AES measurements indicate some N loss (51 atomic % instead of 60 atomic %), which can be written as $LaWN_{3-x}$ (x = 0.5) or $LaWN_{2.5}$. This N deficiency may result from either preferential N removal during AES depth profiling measurement (*21*) or the well-known tendency of the perovskite structure to accommodate large anion deficiency (*4*, *11*). Our scanning transmission electron microscopy (STEM) measurements with energy dispersive x-ray (EDX) analysis from the cross section of an identical film (Fig. 2, E to H) confirm a polycrystalline microstructure (150 to 200 nm grain size) and demonstrate chemical homogeneity on the nanometer scale. The corresponding x-ray diffraction (XRD) patterns of the cation-stoichiometric composition (La/W = 1) are consistent with the modeled phase-pure perovskite reference pattern, with W or WN and amorphous minor secondary phases at W- and La-rich compositions, respectively (fig. S1). We performed electrical and optical property measurements as a function of cation composition, which show 10^{-4} to 10^{4} ohm cm resistivity and 1.0 to 2.5 eV optical absorption onset with increasing La content (fig. S2). The upper bound of these measurements is the most representative of $LaWN_3$ because of the optoelectronically inert character of the amorphous lanthanum oxide second phase at La-rich compositions.

To determine the crystal structure of $LaWN_3$, we synthesized randomly oriented polycrystalline thin films by using rapid thermal annealing (RTA) of atomically dispersed amorphous La-W-N precursor films. These samples were sputter deposited on glass substrates and protected from oxidation with an AlN capping layer (*21*). The capped amorphous sample libraries were also free of O (fig. S3) and had a distinct color change close to the La/W = 1 composition (fig. S4), from black on the W-rich side to translucent yellow on the La-rich side. This yellow color is indicative of a <2.5 eV band gap. After the RTA, we observed a randomly oriented polycrystalline microstructure that was evident from uniform Debye-Scherrer rings (Fig. 3A). Our Rietveld refinement of the integrated XRD patterns (Fig. 3B) shows a majority perovskite phase along with a minority metallic W phase (<5% by volume) and possibly WN phase (<1% by volume) (table S2). For the perovskite crystal structure refinement, we chose candidate space groups (SGs) (table S3) calculated to be within ~100 meV/f.u. of the lowest-energy predicted R3c symmetry of $LaWN_3$ (*17*), as well as the higher-energy $I\bar{4}$ space group reported for the $LaWO_{0.6}N_{2.4}$ oxynitride perovskite (*19*). The structure refinement, performed for the unit cell lattice vectors and angles with all other variables held constant, resulted in low and statistically indistinguishable residuals for both the rhombohedral (R3c, SG 161) and tetragonal ($I\bar{4}$, SG 82) symmetries (fig. S5). We refined unit cell parameter (a) and angle (α) of the ground-state R3c perovskite structure to be a = 5.64 A, α = 60.33° in rhombohedral notation (*21*) and a = 5.67 A, c = 13.79 A in hexagonal notation, which is consistent with theoretical prediction for $LaWN_3$ (*17*). Our TEM-based selected area electron diffraction (SAED) results (Fig. 3, C to F) confirmed the perovskite structure that we determined using XRD and were similarly unable to resolve the nonpolar $I\bar{4}$ (SG 82) versus polar R3c (SG 161) structural distortions.

To distinguish between the two possible polar and nonpolar symmetries of $LaWN_3$ (*17*, *19*), we conducted piezoresponse force microscopy

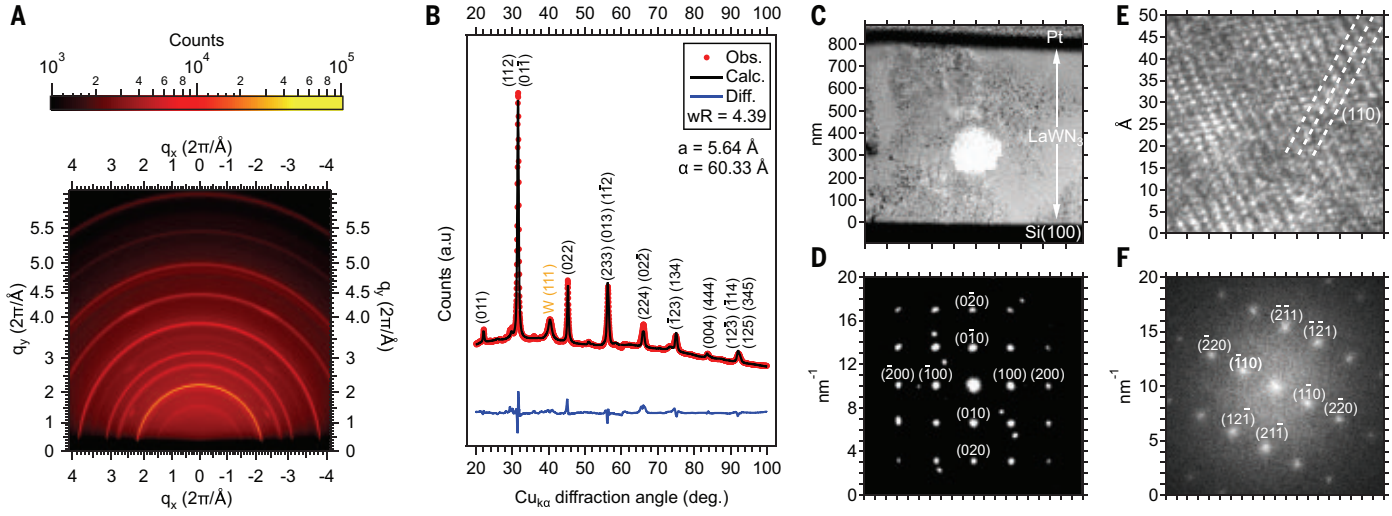


Fig. 3. Crystal structure of $LaWN_3$ thin films. (**A**) Two-dimensional XRD pattern, indicating randomly oriented polycrystalline microstructure. (**B**) Rietveld refinement of XRD data for $LaWN_3$ thin films with a predicted rhombohedral unit cell of R3c symmetry (space group 161) and body-centered cubic (bcc) tungsten (W) minority phase (<5% by volume). (**C**) STEM–HAADF (high angle annular dark field) image of an as-deposited crystalline film highlighting a single grain (white) and (**D**) SAED from this grain showing a pseudo-cubic perovskite [001]–type pattern. (**E**) High-resolution image of a single grain showing the pseudo-cubic (011) lattice spacing and (**F**) the associated fast Fourier transform indexed with a pseudo-cubic [113] type pattern.

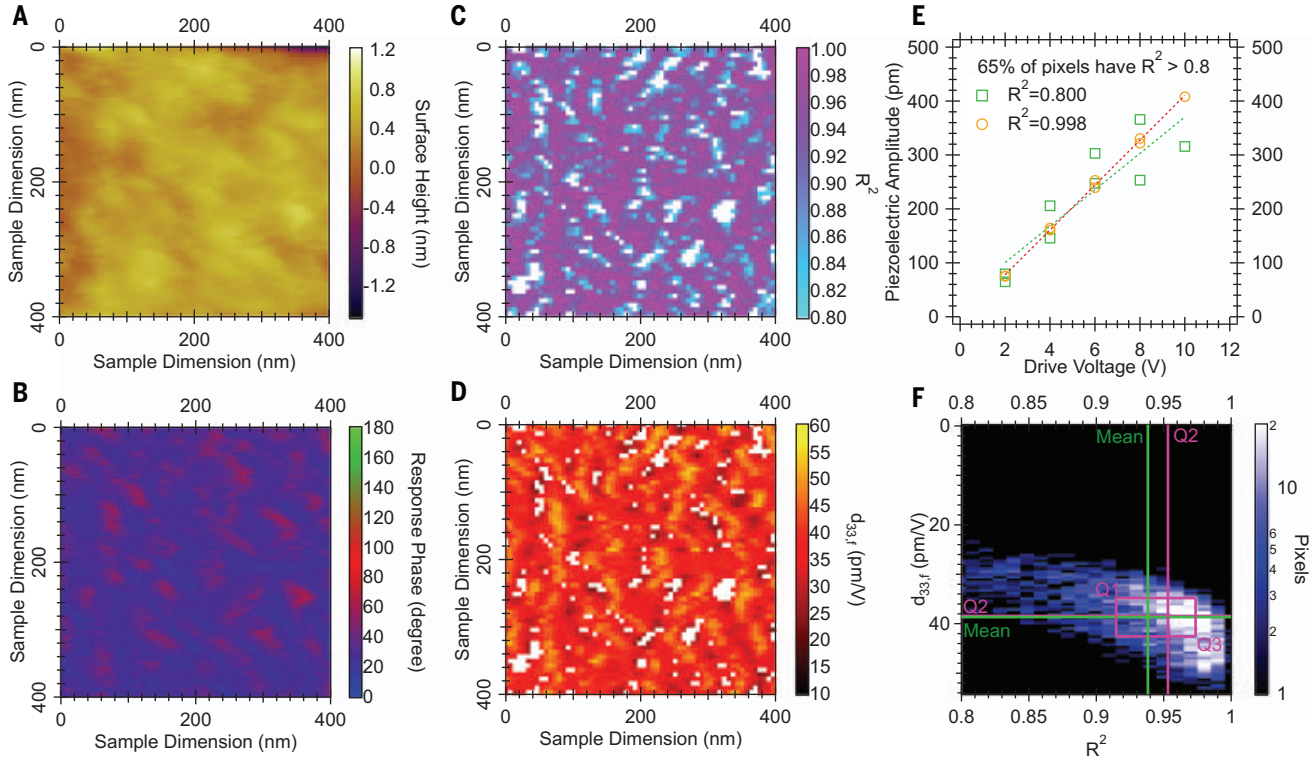


Fig. 4. Piezoelectric properties of $LaWN_3$ thin films. (**A**) Atomically smooth surface of a single $LaWN_3$ grain. (**B**) Phase, (**C**) linearity, and (**D**) slope of each of more than 4000 pixels with R^2 > 0.8 for piezoelectric amplitude versus drive voltage fits. (**E**) The best and worst fits included in this analysis resulting in (**F**) a three-dimensional histogram of these $d_{33,f}$ and R^2 values, indicating a mean (green) and median (magenta) value of the piezoelectric response. Analysis of $LiNbO_3$, PZT, and $(Al_{0.92}Sc_{0.08})N$ reference samples and details of the PFM measurements are provided in the supplementary materials (figs. S7 to S10).

(PFM) measurements (*21*) of the crystalline $LaWN_3$ film synthesized on the heated substrate (Fig. 1 and figs. S1 and S2). We used a <25-nm-radius tip to probe the electromechanical response of uncapped crystalline thin films that were insulating according to conductive atomic force microscopy (c-AFM) measurements (fig. S6). Our PFM results show qualitatively unambiguous piezoelectric response (Fig. 4, A to F), with 65% of more than 4000 pixels in a map (Fig. 4D) having coefficient of determination (R^2) > 0.8 fit of piezoelectric amplitude versus drive voltage (Fig. 4E). Our statistical analysis of these measurement results (Fig. 4F) in terms of mean and median

of all the pixels with $R^2 > 0.8$ indicate $d_{33,f}$ = 40 pm/V magnitude of effective piezoelectric strain coefficient. This $LaWN_3$ value (Fig. 1C) is four times larger than that of the $Al_{0.92}Sc_{0.08}N$ (~10 pm/V) and $LiNbO_3$ (~10 pm/V) reference samples (figs. S7 and S8) yet smaller than the highly engineered $PbZr_{0.52}Ti_{0.48}O_3$ reference sample (~150 pm/V) (fig. S9). Our PFM results clearly indicate a noncentrosymmetric unit cell, supporting the predicted R3c (SG 161) polar symmetry of $LaWN_3$ (*17*) and ruling out $I\bar{4}$ (SG 82) (*19*) as well as other centrosymmetric possibilities within ~100 meV/f.u. from the ground state (table S3). Although we are hesitant to claim quantitative values of piezoelectric coefficient from such PFM measurements (*22*), the results not only confirm polar symmetry of $LaWN_3$ but also indicate its strong piezoelectric response (Fig. 4, A to F).

Considering the computationally predicted ferroelectric character of $LaWN_3$ (17), we attempted to measure $LaWN_3$ polarization reversal (fig. S10, C and D). Our PFM measurements of the crystalline film (*21*) show that the phase of the piezoelectric response of a single 200-nm grain switches in the 0.25 to 0.50 MV/cm range of electric field. These values are similar to those of $PbZr_{0.52}Ti_{0.48}O_3$ that we measured under identical conditions (fig. S10), albeit with 150° instead of 180° phase change, indicating either incomplete switching or substantial charged-defect accumulation in the sample. Because of known challenges with such PFM ferroelectric measurements (*22*), we attempted domain writing across 10- by 10-μm area (fig. S10, E to H), but the results were inconclusive, suggesting the presence of defects at this larger scale. We also attempted macroscale electrical measurements for the samples with a narrow composition gradient and 100-μm-radius contacts (*21*), a process benchmarked by ferroelectric $Al_{1-x}Sc_xN$ thin films synthesized and characterized in our laboratory (*23*). We deposited these (111)-oriented perovskite films (fig. S11) on several conductive substrates (p^+ Si and Pt/Si) under many conditions—such as variations in total power, gas ratio, and film thickness—meant to simultaneously maximize crystallinity and minimize conductivity (fig. S11A). Despite multiple persistent attempts, we observed no definitive polarization-field (*P-E*) ferroelectric loops up to the measurement field of ~1 MV/cm, with the signal dominated by leakage current (fig. S11B). These microscale and macroscale electrical measurements are difficult because of a combination of residual minor impurities, such as metallic W or WN measured with XRD, and point defects, such as N deficiency suggested by AES. Synthesis of higher-quality $LaWN_3$ [theoretical band gap (E_g) = 1.8eV] or wider-gap $LaMoN_3$ (theoretical E_g = 2.7 eV) (*24*) may help decrease the leakage current and help determine whether these materials are indeed ferroelectric.

Nitride perovskites could substantially extend the range of possible applications of existing commercial nitride semiconductor devices. GaN, AlN, and related III-N alloys are well established for electronics in radio-frequency transistors, photonics in light emitting diodes, and telecommunication in film bulk acoustic resonators (*20*). Nitride perovskites may offer additional integration advantages compared with oxide perovskites on wurtzite nitrides. High-quality epitaxial layers of thermodynamically stable nitride perovskites (similar to oxide perovskites) would be easier to synthesize at high temperature than the recently reported metastable (Al,Sc)N wurtzite alloys with high (>30%) Sc content (*23*, *25*). Also, compared with oxide perovskites, nitride perovskites would be easier to integrate with GaN, similar to other nitride wurtzites. This is because there would be no competing N-O anion exchange reaction that results in interfacial layer formation known from growth of oxide perovskites on Si (*26*). Thus, epitaxial integration of nitride perovskites on nitride semiconductors may lead to entirely new types of devices for a broad range of applications (*20*), as highlighted by prospects of quantum computing and single-photon detectors in superconductor-semiconductor nitride heterostructures (*27*).

Our successful synthesis and characterization of $LaWN_3$ perovskite with polar symmetry should lead to more experimental measurements of its properties, as well as growth and characterization of many other theoretically predicted nitride perovskites. In addition to our measured strong piezoelectric response (40 pm/V) and published theoretically predicted ferroelectricity (*17*), there are multiple theoretical predictions of other interesting and useful $LaWN_3$ properties that await experimental confirmation, including spin textures (*28*) and p-type doping (*24*). The relatively narrow band gap of $LaWN_3$ (theoretical E_g = 1.8 eV) (*24*) compared with that of oxide perovskites (for example, E_g = 3.4 eV in $BaTiO_3$) could also offer an advantage in studying a controversial topic of solar energy conversion in perovskite materials with polar symmetry (*29*). Other nitride perovskites computationally predicted to have interesting properties are metallic $TbReN_3$ with very high anisotropy and large saturation magnetization (*16*) and $TaThN_3$ with topological insulating behavior (*30*). Nitride perovskites may also harbor other emergent properties or hidden states because of the mixed covalent and ionic character of the metal-nitrogen bonds that results from smaller electronegativity of N compared with O (*18*). Thus, our work on $LaWN_3$ opens the door to synthesis of other predicted nitride perovskites with exceptional electromechanical, magnetic, optoelectronic, thermoelectric, topological, and quantum properties.

REFERENCES AND NOTES

1. A. S. Bhalla, R. Guo, R. Roy, *Mater. Res. Innov.* **4**, 3–26 (2016).
2. D. Dimos, C. H. Mueller, *Annu. Rev. Mater. Sci.* **28**, 397–419 (1998).
3. P. Muralt, R. G. Polcawich, S. Trolier-McKinstry, *MRS Bull.* **34**, 658–664 (2009).
4. M. Papac, V. Stevanović, A. Zakutayev, R. O'Hayre, *Nat. Mater.* **20**, 301–313 (2021).
5. A. V. Boris *et al.*, *Science* **332**, 937–940 (2011).
6. X. Li *et al.*, *Science* **364**, 1079–1082 (2019).
7. V. M. Goldschmidt, *Naturwissenschaften* **14**, 477–485(1926).
8. M. M. Lee, J. Teuscher, T. Miyasaka, T. N. Murakami, H. J. Snaith, *Science* **338**, 643–647 (2012).
9. S. Niu *et al.*, *Nat. Photonics* **12**, 392–396 (2018).
10. N. E. Brese, F. J. DiSalvo, *J. Solid State Chem.* **120**, 378–380 (1995).
11. S. D. Kloß, M. L. Weidemann, J. P. Attfield, *Angew. Chem. Int. Ed.* **60**, 22260–22264 (2021).
12. R. Niewa, *Eur. J. Inorg. Chem.* **2019**, 3647–3660 (2019).
13. S. Curtarolo *et al.*, *Nat. Mater.* **12**, 191–201 (2013).
14. W. Sun *et al.*, *Nat. Mater.* **18**, 732–739 (2019).
15. R. Sarmiento-Pérez, T. F. Cerqueira, S. Körbel, S. Botti, M. A. Marques, *Chem. Mater.* **27**, 5957–5963 (2015).
16. J. A. Flores-Livas, R. Sarmiento-Pérez, S. Botti, S. Goedecker, M. A. Marques, *J. Phys. Mater.* **2**, 025003 (2019).
17. Y. W. Fang *et al.*, *Phys. Rev. B* **95**, 014111 (2017).
18. F. J. Disalvo, *Science* **247**, 649–655 (1990).
19. P. Bacher *et al.*, *J. Solid State Chem.* **77**, 67–71 (1988).
20. D. Jena *et al.*, *Jpn. J. Appl. Phys.* **58** (SC), SC0801 (2019).
21. Materials and methods are available as supplementary materials.
22. R. K. Vasudevan, N. Balke, P. Maksymovych, S. Jesse, S. V. Kalinin, *Appl. Phys. Rev.* **4**, 021302 (2017).
23. K. Yazawa, D. Drury, A. Zakutayev, G. L. Brennecka, *Appl. Phys. Lett.* **118**, 162903 (2021).
24. S. Singh, M. N. Tripathi, *J. Appl. Phys.* **124**, 065109 (2018).
25. S. Fichtner, N. Wolff, F. Lofink, L. Kienle, B. Wagner, *J. Appl. Phys.* **125**, 114103 (2019).
26. R. A. McKee, F. J. Walker, M. F. Chisholm, *Phys. Rev. Lett.* **81**, 3014–3017 (1998).
27. R. Yan *et al.*, *Nature* **555**, 183–189 (2018).
28. H. J. Zhao *et al.*, *Phys. Rev. B* **102**, 041203 (2020).
29. M.-M. Yang, D. J. Kim, M. Alexe, *Science* **360**, 904–907 (2018).
30. M. C. Jung, K.-W. Lee, W. E. Pickett, *Phys. Rev. B* **97**, 121104 (2018).

ACKNOWLEDGMENTS

Funding: This work was authored at the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, for the US Department of Energy (DOE) under contract DE-AC36-08GO28308. Funding was provided by the Office of Science (SC), Office of Basic Energy Sciences (BES), Materials Chemistry Program, as a part of the Early Career Award "Kinetic Synthesis of Metastable Nitrides" (synthesis and characterization); as well as by the US National Science Foundation (NSF) Designing Materials to Revolutionize and Engineer our Future (DMREF) program (DMREF-1534503), and DARPA Tunable Ferroelectric Nitrides (TUFEN) program (DARPA-PA-19-04-03) (piezoelectric measurements). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US DOE, SC, BES under contract DE-AC02-76SF00515. The views expressed in the article do not necessarily represent the views of the DOE or the US government. **Author contributions:** K.R.T. synthesized the films, performed x-ray fluorescence and scattering measurements and piezoresponse force microscopy, and drafted the original version of the manuscript. C.L.P. performed AES measurements and commented on the manuscript. D.R.D. performed TEM measurements and commented on the manuscript. G.L.B. provided intellectual guidance and physical resources, assisted in data analysis, and edited the manuscript. A.Z. conceived the overall study; assisted in data analysis; provided intellectual guidance and physical resources; and edited, revised, and finalized the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm3466
Materials and Methods
Supplemental Text
Figs. S1 to S11
Tables S1 to S3
References (*31*–*38*)
Data File S1

10 September 2021; accepted 30 October 2021
10.1126/science.abm3466

NEUROSCIENCE

Supramammillary regulation of locomotion and hippocampal activity

Jordan S. Farrell[1]*, Matthew Lovett-Barron[2,3], Peter M. Klein[1], Fraser T. Sparks[4,5,6], Tilo Gschwind[1], Anna L. Ortiz[1], Biafra Ahanonu[7,8], Susanna Bradbury[2], Satoshi Terada[4,5,6], Mikko Oijala[1], Ernie Hwaun[1], Barna Dudok[1], Gergely Szabo[1], Mark J. Schnitzer[7,9], Karl Deisseroth[2,9,10], Attila Losonczy[4,5,6], Ivan Soltesz[1]

Locomotor speed is a basic input used to calculate one's position, but where this signal comes from is unclear. We identified neurons in the supramammillary nucleus (SuM) of the rodent hypothalamus that were highly correlated with future locomotor speed and reliably drove locomotion when activated. Robust locomotion control was specifically identified in *Tac1* (substance P)–expressing (SuM^{Tac1+}) neurons, the activation of which selectively controlled the activity of speed-modulated hippocampal neurons. By contrast, *Tac1*-deficient (SuM^{Tac1-}) cells weakly regulated locomotion but potently controlled the spike timing of hippocampal neurons and were sufficient to entrain local network oscillations. These findings emphasize that the SuM not only regulates basic locomotor activity but also selectively shapes hippocampal neural activity in a manner that may support spatial navigation.

The ability to construct and access a mental map of one's environment during locomotion is an important adaptation to facilitate survival and is supported by tracking self-motion (*1*). Mammalian locomotion is intimately tied to the occurrence of 6- to 12-Hz hippocampal theta oscillations, such that the theta rhythm begins before the onset of self-generated motion and increases in amplitude with respect to speed (*2*–*6*). By temporally organizing the activity of place-coding neuronal assemblies into trajectories across past, present, and future locations, hippocampal theta oscillations are thought to subserve cognitive operations during spatial navigation (*5*–*10*). Tight coupling of theta waves to speed could be the result of shared neural circuitry between self-generated locomotion and theta control, providing a potential speed signal (*1*). Alternatively, speed could be derived from optic flow, vestibular input, or an efference copy from locomotor areas (*11*–*14*). Since the identification of speed-encoding neurons in brain areas that are thought to use a speed signal to calculate position (such as the hippocampus and entorhinal cortex) (*3*, *15*–*18*), research interest in understanding potential sources of speed input has grown.

The medial septum is critical for hippocampal theta activity and is functionally coupled

[1]Department of Neurosurgery, Stanford University, Stanford, CA, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA, USA. [3]Neurobiology Section, Division of Biological Sciences, University of California, San Diego, CA, USA. [4]Department of Neuroscience, Columbia University, New York, NY, USA. [5]Kavli Institute for Brain Sciences, Columbia University, New York, NY, USA. [6]Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. [7]Departments of Biology and Applied Physics, Stanford University, Stanford, CA, USA. [8]Department of Anatomy, University of California, San Francisco, CA, USA. [9]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [10]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA.
*Corresponding author. Email: jsfarrel@stanford.edu

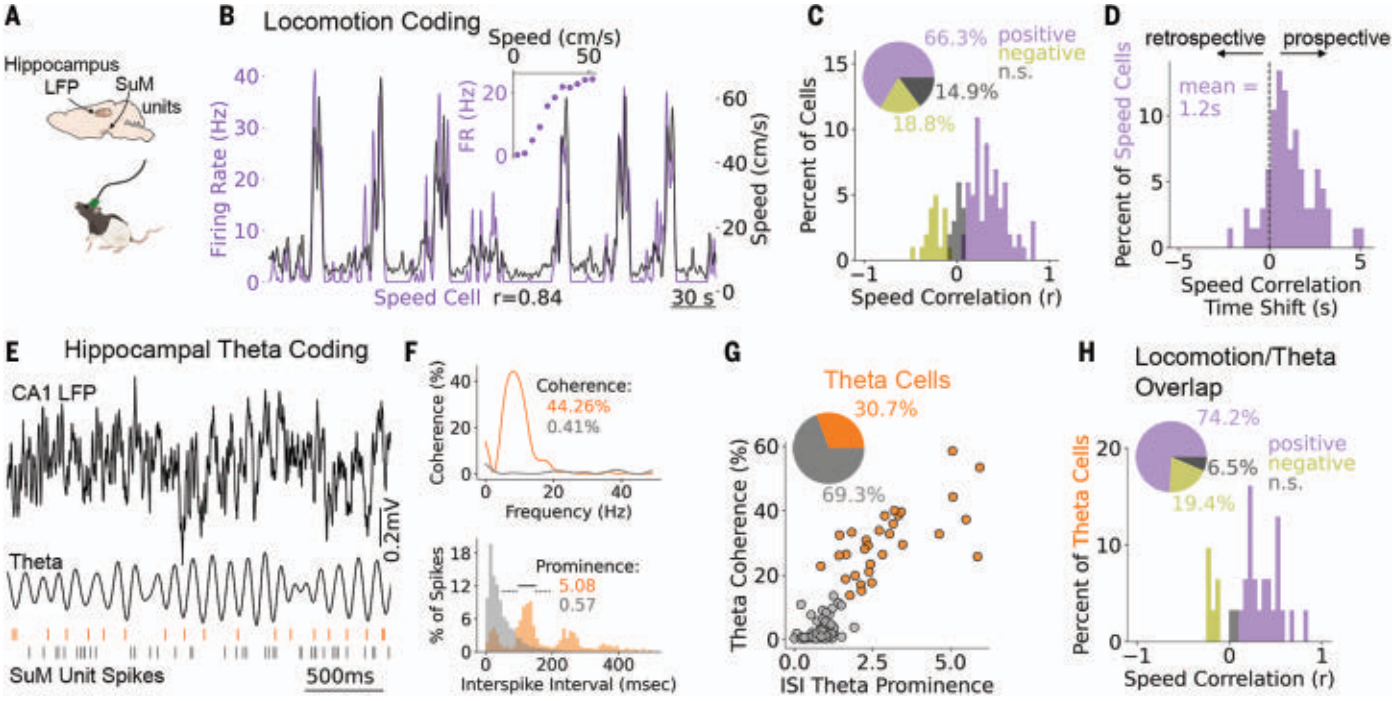


Fig. 1. SuM representation of speed and hippocampal theta oscillations. (**A**) Recording paradigm; data from unrestrained rats (*26*), reanalyzed here for speed- and theta-related investigations. (**B**) Example SuM unit (recorded by tetrodes) that is positively correlated with speed (i.e., a speed cell). The inset represents these data as a scatter plot. (**C**) Distribution of speed versus firing rate Pearson correlation coefficient (*r*) values. The pie chart shows the percentage of units for positive (r = 0.36 ± 0.023, mean ± SEM), negative (r = −0.24 ± 0.024), and nonsignificant (n.s.) cells (r = 0.018 ± 0.012). (**D**) Distribution of temporal offsets for positive speed cells. SEM = 0.19. (**E**) Two example SuM unit spiking activities. The cell represented in orange shows phase-locked firing with respect to hippocampal theta oscillations, whereas the cell represented in gray does not. (**F**) Quantification of theta-related firing for two example units from (E). The top panel shows spike-field coherence; the bottom shows theta-rhythmic spiking. (**G**) Units were clustered into theta cells according to quantification from (F). ISI, interspike interval. (**H**) Distribution of speed scores among clustered theta cells from (G).

to locomotion (*3*, *19*–*21*), but other brain areas have been proposed to contribute to this oscillation. One such area is the supramammillary nucleus (SuM) of the posterior hypothalamus (*22*–*24*), which has recently been shown to have roles in arousal (*25*), spike-timing coordination (*26*), and identification of novelty (*27*). As a proposed theta controller, neural activity in the SuM is likely also related to locomotion, but this has not been systematically investigated. In addition to innervating the medial septum, the SuM has highly divergent outputs that target the midbrain, where locomotor commands are integrated, as well as regions involved in spatial navigation, including the hippocampus, entorhinal cortex, medial prefrontal cortex, nucleus reuniens, and claustrum (*28*). Thus, the SuM projects to theta, locomotor, and spatial navigation circuitry, but the functional relevance of this positioning remains poorly understood.

Using electrophysiological data from rats navigating a continuous alternation task for reward (*26*), we first investigated how the spiking activity of SuM neurons relates to locomotor speed and hippocampal theta oscillations (Fig. 1A). Similar to previous observations in the midbrain locomotor region (MLR) (*29*), we found a large proportion of SuM units with firing rates that were significantly coupled to locomotor speed, with the majority displaying a positive correlation (Fig. 1, B and C, and fig. S1, A, C, and D). Notably, SuM "speed cells" were more closely correlated with future speed (with an average offset of 1.2 s) than real-time speed (Fig. 1D and fig. S1, A and D). After adjusting for temporal offsets, the firing rates of 99% of SuM units were modulated by speed (fig. S1B). Most speed cells retained their correlation with immobility data withheld (fig. S1, E and F) and were more weakly coupled to acceleration than speed (fig. S2). SuM unit activity was also correlated to hippocampal theta amplitude in real time, but this was less accurately modeled than speed (fig. S3). Thus, SuM activity and locomotion are strongly coupled.

We then addressed whether speed-related neuronal activity is propagated to projection targets, because not all locally recorded SuM units are necessarily projection neurons. In vivo two-photon calcium imaging of SuM axon terminals that innervate the dentate gyrus (DG) and the CA2 region of the hippocampus was performed on head-fixed mice (fig. S4A). Indeed, speed-correlated activity in SuM axons was observed in both regions (fig. S4, B and C). Extensive collateralization of SuM axons was also determined (fig. S5) and may contribute to the similar proportion of positively and negatively correlated axons in both regions (fig. S4B).

Next, we examined the coupling between SuM action potentials and hippocampal theta waves (Fig. 1E). High coherence with hippocampal theta oscillations and theta-rhythmic spiking was observed for 30.7% of units (Fig. 1, F and G; see materials and methods). SuM theta cells typically fired near the trough of CA1 theta waves, with a slight prospective bias (fig. S6). Much like the overall SuM population, most SuM theta cells were positively correlated with locomotor speed (Fig. 1H).

To test the functional involvement of the SuM in locomotion and theta activity, we injected the SuM with recombinant adeno-associated virus (rAAV) to express the optogenetic proteins channelrhodopsin-2 (ChR2;

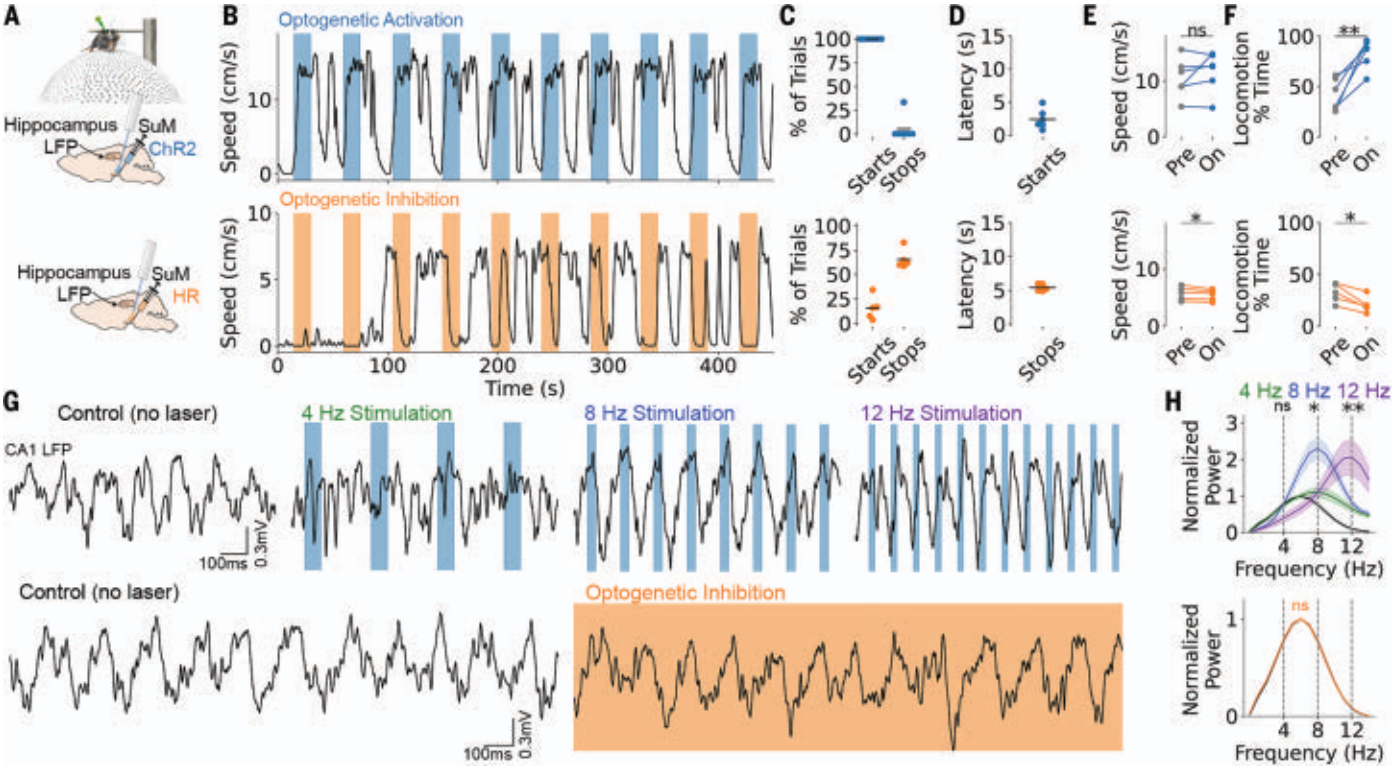


Fig. 2. Optogenetic SuM modulation controls locomotion and hippocampal LFP. (**A**) Pan-neuronal SuM activation with ChR2 (blue) or inhibition with HR (orange) in head-fixed mice on a floating ball. (**B**) Bidirectional locomotor effect with SuM activation (top) and inhibition (bottom). Colored bars denote "laser on" segments. (**C**) Percent of trials where locomotion was initiated or halted for ChR2 (top) and HR (bottom). (**D**) Latency of the start versus stop response. (**E**) Speed during locomotor epochs before (Pre) and during (On) light delivery. ChR2, $t_5 = -1.14$, $P = 0.31$; HR, $t_4 = 3.07$, $P = 0.037$. (**F**) Percent of time spent locomoting before (Pre) and during (On) light delivery. ChR2, $t_5 = -4.84$, $P = 0.0047$; HR, $t_4 = 4.60$, $P = 0.010$. (**G**) (Top) Optogenetic activation at 4, 8, or 12 Hz compared with no laser control. Blue bars denote laser on. (Bottom) Optogenetic inhibition (orange shading) versus no laser control. The scale bar applies across rows. (**H**) Quantification of power spectrum changes normalized to no laser control. (Top) ChR2: 4-Hz power at 4-Hz stimulation, $t_5 = 0.86$, $P = 0.43$; 8-Hz power at 8-Hz stimulation, $t_5 = 5.18$, $P = 0.0035$; 12-Hz power at 12-Hz stimulation, $t_5 = 3.99$, $P = 0.010$. (Bottom) HR: $t_4 = 0.043$, $P = 0.97$. Data are mean ± SEM. ns, not significant. *$P < 0.05$, **$P < 0.01$.

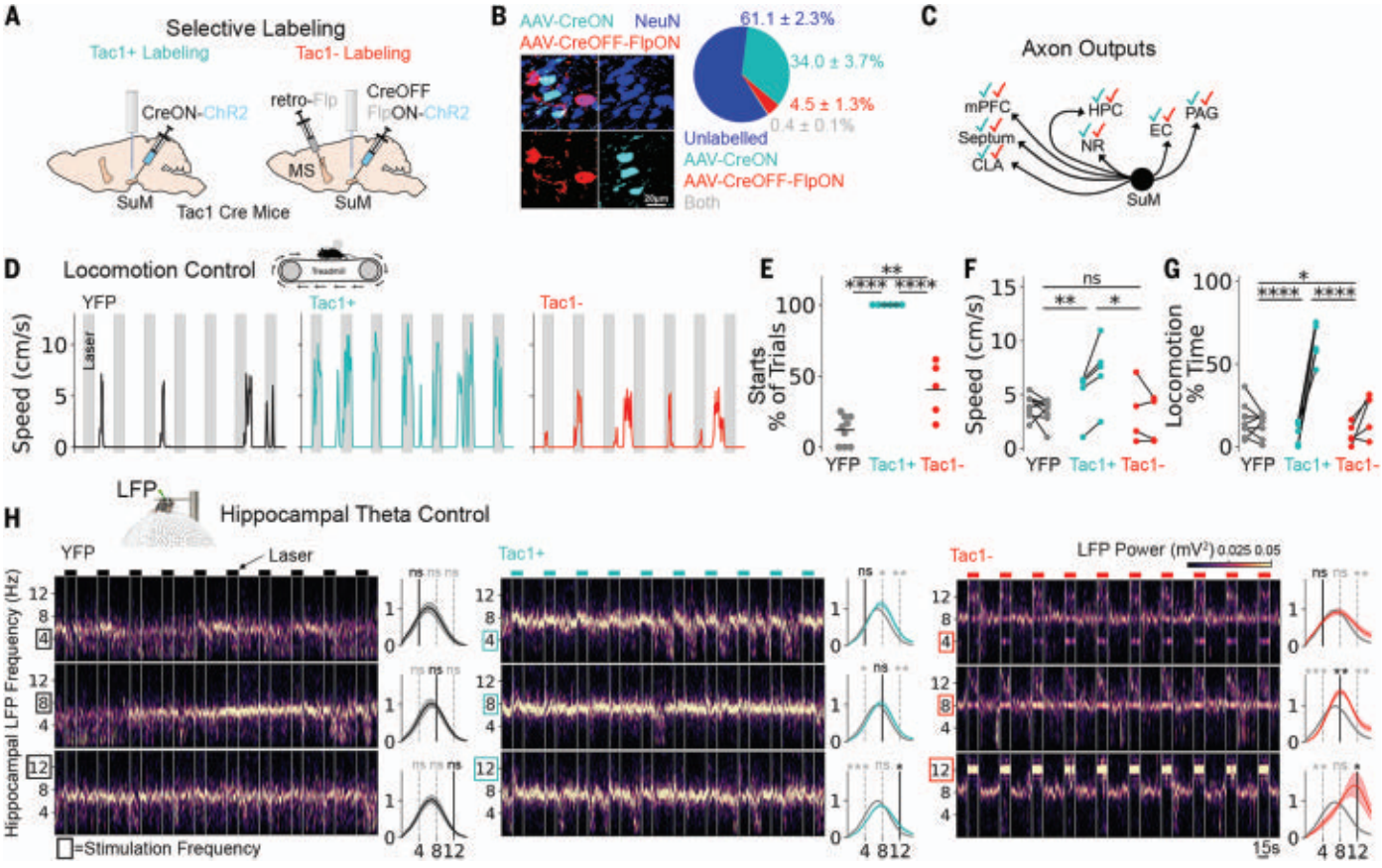


Fig. 3. Cell type dependence of locomotion initiation and LFP entrainment. (**A**) Labeling strategy to target mutually exclusive populations according to the presence or absence of *Tac1*. MS, medial septum. (**B**) Investigation of labeling specificity. (Left) Image showing AAV-labeled *Tac1*+ (CreON) and *Tac1*− cells (CreOFF-FlpON) among other NeuN+ cells in the SuM. (Right) Quantification. (**C**) Schematic showing checkmarks of each color (cyan, SuM^{Tac1+}; red, SuM^{Tac1-}) if axons were found in SuM target regions (see fig. S9). mPFC, medial prefrontal cortex; CLA, claustrum; HPC, hippocampus; NR, nucleus reuniens; EC, entorhinal cortex; PAG, periaqueductal gray. (**D**) Representative locomotor activity during optogenetic activation at 8 Hz. YFP, yellow fluorescent protein. (**E**) Percent of trials with locomotion initiation. Analysis of variance (ANOVA) $F_{2,17} = 103.6$, $P < 0.0001$, Tukey post hoc test. (**F**) Locomotor speed before (left dot) versus during (right dot) light delivery, connected by a line for each mouse. ANOVA was performed to determine group differences on changes in speed (before versus during light delivery). $F_{2,15} = 7.2$, $P = 0.0064$, Tukey post hoc test. (**G**) Percent of time locomoting before (left dot) versus during (right dot) light delivery. Differences in response change were assessed by ANOVA. $F_{2,17} = 57.8$, $P < 0.0001$, Tukey post hoc test. (**H**) Optogenetic stimulation in head-fixed mice on a floating ball at 4, 8, or 12 Hz. Spectrogram (left) and power spectral density changes [right; off (light gray) versus on (black, YFP; cyan, Tac1+; red, Tac1−)] for each condition (columns) at each frequency (rows). Paired *t* tests performed in light-off versus light-on conditions. ns, not significant. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

excitatory) or halorhodopsin (HR; inhibitory) under control of a pan-neuronal promoter (Fig. 2, A and G). Light activation of ChR2 drove locomotion in 100% of trials, with an average latency of 2.4 ± 0.6 s in head-fixed mice (Fig. 2, B to F). Locomotion lagged behind an increase in theta amplitude (fig. S7), which is consistent with prospective encoding of speed but real-time encoding of theta amplitude (Fig. 1D versus fig. S3A). The frequency of laser pulses reliably entrained hippocampal local field potential (LFP) at 8 and 12 Hz (Fig. 2, G and H), similar to previous optogenetic manipulations in the medial septum (*3*, *4*, *20*, *21*). Most CA1 neurons (94%) were entrained by laser pulses; however, the preferred firing phase shifted considerably from spontaneous theta activity (fig. S8). Optogenetic inhibition with HR halted locomotion on 65.6 ± 0.6% of trials, with a latency of 5.4 ± 0.2 s (Fig. 2, B to F). Consistent with previous studies (*26*, *30*), SuM inhibition did not alter hippocampal theta rhythms during locomotion (Fig. 2, G and H). Thus, optogenetic manipulation of the SuM robustly and bidirectionally controls locomotion but does not inhibit spontaneous hippocampal theta waves, despite robust spike and LFP entrainment with ChR2.

Considering the heterogeneity of SuM cell types (*31*), we then determined whether locomotion control and spike and LFP entrainment were dependent on cell type. A subset of SuM neurons express substance P, encoded by the *Tac1* gene, and project to the hippocampus and other regions (*32*). Using the Tac1-Cre mouse line, we targeted two mutually exclusive populations. *Tac1*-expressing (SuM^{Tac1+}) neurons were labeled with a CreON rAAV, whereas a smaller population of *Tac1*-deficient (SuM^{Tac1-}) projection neurons were labeled with an intersectional approach using a CreOFF-FlpON rAAV (*33*), facilitated by a retrograde rAAV carrying Flp in the medial septum (Fig. 3, A and B). The axon outputs of both cell populations were similar, innervating the known and expected SuM target regions (Fig. 3C and fig. S9). However, SuM^{Tac1+} and SuM^{Tac1-} cells had different intrinsic properties (fig. S10). Cell types were further differentiated by the non-uniform axon innervation pattern of the DG granule cell layer (fig. S11C) and increased vesicular γ-aminobutyric acid (GABA) transporter content of SuM^{Tac1-} cells (fig. S11, A and B), consistent with previously identified GABA and glutamate co-releasing SuM cells (*34*). Functionally, SuM^{Tac1+} cell stimulation robustly drove locomotion in 100% of trials (Fig. 3, D to G) but did not entrain hippocampal LFP (Fig. 3H). This locomotion was relatively slow and steady with an alternating

stepping pattern (movie S1), consistent with exploratory locomotion and distinct from fight or flight responses seen with stimulation of other hypothalamic areas (*35*). By contrast, activation of SuM^{Tac1-} cells weakly controlled locomotion (Fig. 3, D to G) but precisely controlled the frequency of hippocampal LFP at 8 and 12 Hz (Fig. 3H).

Given the robust control of movement initiation by SuM^{Tac1+} neuron stimulation, we further examined the role of this cell type in locomotion. Using two-photon calcium imaging in head-fixed mice running on a treadmill, we observed a high proportion of SuM^{Tac1+} cells for which activity was positively correlated with speed and that were active before locomotion onset (fig. S12). Similar results were obtained from *Tac1*+ cells in the ventral hypothalamus of zebrafish during swimming behavior, supporting evolutionary conservation of function (fig. S13). As with broad SuM inhibition, selective optogenetic inhibition of SuM^{Tac1+} neurons also suppressed locomotion in head-fixed mice (fig. S14). Finally, we examined potential SuM^{Tac1+} output pathways that reach the MLR, where the coordination of locomotor input and gait selection takes place (*36*). Using anterograde trans-synaptic tracing to label postsynaptic neurons (*37*), we found that midbrain periaqueductal gray neurons that specifically receive SuM^{Tac1+} input, in turn, project to the MLR (fig. S15).

Finally, we determined how SuM^{Tac1+} and SuM^{Tac1-} neurons alter firing rate and control spike timing of hippocampal neurons. Given the tight coupling of SuM^{Tac1+} cells to locomotion, we hypothesized that spontaneously speed-correlated hippocampal neurons would be particularly sensitive to SuM^{Tac1+} activation. On average, activation of both SuM cell types increased the firing rate of the hippocampal units during locomotion (Fig. 4B) and resulted in similar proportions of units with significant firing rate alterations (Fig. 4A). The effect of SuM^{Tac1+} stimulation, but not SuM^{Tac1-} stimulation, was indeed correlated to the magnitude of spontaneous speed modulation, such that the firing rates of positively correlated speed units increased, and negatively correlated cells decreased with optogenetic stimulation (Fig. 4B). We observed the opposite relationship with optogenetic inhibition of SuM^{Tac1+} cells (fig. S14E). Hippocampal speed cell firing rates were then modeled from speed and laser timing as inputs (Fig. 4C). Modeling firing rates during SuM^{Tac1+} stimulation produced considerably less error than modeling during SuM^{Tac1-} stimulation when laser timing was withheld (Fig. 4C), supporting the notion that the effect of SuM^{Tac1+} stimulation on locomotion and hippocampal speed cell firing rates are coupled.

At a subsecond time scale, we also determined whether hippocampal spike timing was altered relative to each laser pulse. Because SuM^{Tac1-} activation overrode spontaneous hippocampal theta waves and entrained LFP, we hypothesized that spike timing would be more affected by activation of this cell type. Indeed, SuM^{Tac1-} activation entrained spike timing in more units with a greater average effect (Fig. 4, D and E). Moreover, SuM^{Tac1-} activation increased the firing rate of most units in close proximity to the termination of each light pulse, whereas SuM^{Tac1+} activation had mixed responses (Fig. 4F and fig. S16). Unlike SuM^{Tac1-} stimulation, the effect of SuM^{Tac1+} stimulation depended on that unit's speed correlation, such that positively correlated hippocampal speed cells were more likely to fire shortly after laser pulse onset and negatively correlated cells were suppressed (Fig. 4G). A significant correlation between the magnitude of optogenetic entrainment and the magnitude of spontaneous theta entrainment was observed for activation of both cell types; however, the linear fit slope was twice the value for SuM^{Tac1-} activation (Fig. 4H). Lastly, the preferred firing phase relative to hippocampal theta activity was more perturbed by SuM^{Tac1-} cell activation (fig. S17).

The finding that SuM^{Tac1-} cells potently regulate hippocampal spike timing and are sufficient to entrain LFP is notable, given the lack of effect of SuM inhibition on spontaneous hippocampal theta oscillations. However, others have reported no change in hippocampal theta activity with SuM inhibition and lesions

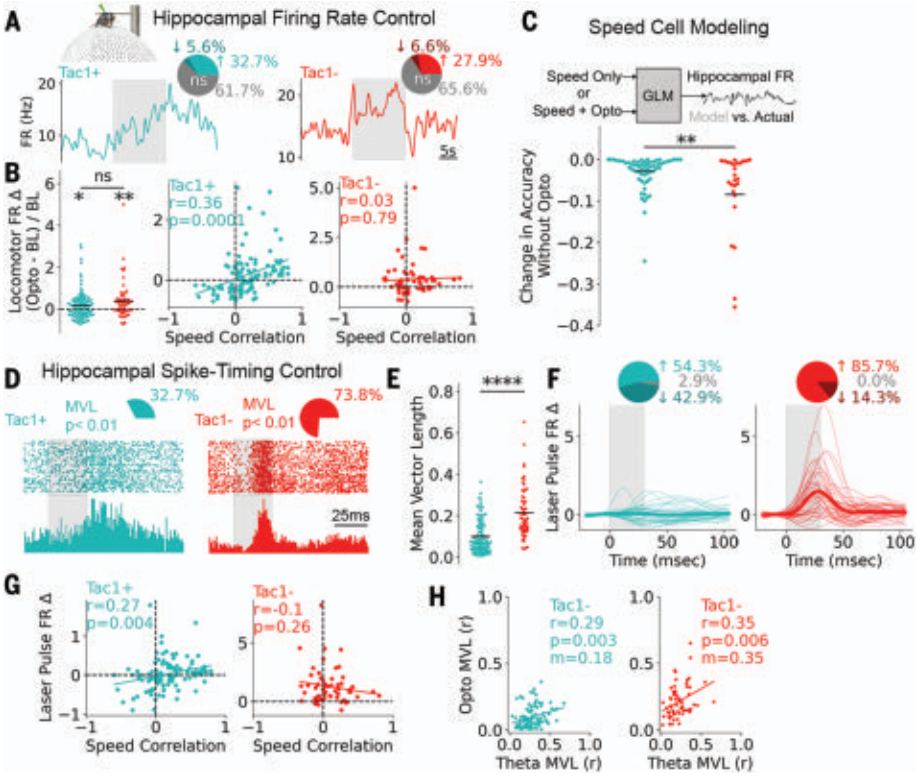


Fig. 4. Hippocampal populations are differentially regulated by SuM cell types. (**A**) Data were obtained from head-fixed mice on a floating ball. Mean firing rate (FR) changes from two example hippocampal cells during SuM^{Tac1+} or SuM^{Tac1-} optogenetic activation (gray shaded area). Pie charts display the proportion of units with significantly altered locomotor firing rates. (**B**) Locomotor firing rate change for light-on versus light-off conditions (one-sample *t* test: SuM^{Tac1+}, t_{105} = 2.56, P = 0.012; SuM^{Tac1-}, t_{57} = 3.28, P = 0.0017; between sample *t* test: t_{162} = 1.69, P = 0.09) and as a function of speed correlation (calculated while laser is off). The *y*-axis label applies to all panels. BL, baseline. (**C**) Generalized linear model of hippocampal speed cell firing rate, with two sets of input [speed only versus speed + optogenetic input (opto)]. The plot shows the change in modelling accuracy when optogenetic information was withheld (t_{89} = 3.28, P = 0.0015). (**D**) Hippocampal spike raster plots with histograms aligned to laser pulses (gray bar) during 8-Hz stimulation. Pie charts show the proportion of units with significantly laser-modulated spike distributions. MVL, mean vector length. (**E**) Quantification of nonuniform spike distributions from (D). t_{166} = 7.39, P = 6.8 × 10^{12}. (**F**) Smoothed spike histograms of significantly laser-modulated cells. Thin lines represent individual cells; the thick line represents the mean. Pie charts show the directionality of modulation. (**G**) Laser pulse–triggered firing rate change plotted against cells' speed correlations (calculated while laser is off). (**H**) Firing rate modulation by spontaneous theta versus optogenetic laser pulses. m, slope. Lines in (B), (G), and (H) represent linear fits.

(*26*, *30*), despite substantial effects with SuM activation (*31*). Recent work demonstrated that a subpopulation of SuM units increase their activity under novel conditions (*27*). SuM cells from this previous study have several characteristics that overlap with SuM^{Tac1-} cells, including the nonuniform axonal innervation pattern and mixed neurotransmitter phenotype in the DG (*34*). In another study, optogenetic inhibition of the SuM affected theta-range spike timing in SuM-connected structures, but only at the decision point of the maze (*26*). Thus, SuM^{Tac1-} cells may be most influential to hippocampal network patterns in particularly salient situations, with little contribution to the mechanisms that underlie spontaneous theta waves. Consistent with this model, SuM^{Tac1-} activation caused a considerable shift in firing phase preferences from spontaneously generated theta rhythm, highlighting potentially different underlying mechanisms for SuM^{Tac1-}-evoked versus spontaneous theta oscillations.

These data also advance our understanding of the SuM's in role in locomotion and identify a cell type with functional properties relevant to spatial navigation. In addition to the tight coupling of SuM^{Tac1+} activity to speed, we found that SuM^{Tac1+} activation robustly drove locomotion while selectively regulating the activity of speed-sensitive hippocampal neurons. These data raise the possibility that SuM^{Tac1+} neurons have a role in distributing a speed signal throughout the SuM's many axon-termination sites and complement recent work outlining a speed-relaying pathway from the MLR to the entorhinal cortex via the septum (*38*). Given that SuM^{Tac1+} cells encode future speed, the SuM may provide its synaptic partners with intended speed, whereas executed speed is propagated from the MLR. Thus, the SuM may support a role in planning and error correction during locomotion by broadcasting a future speed signal.

REFERENCES AND NOTES

1. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Clarendon Press, 1978).
2. C. H. Vanderwolf, *Electroencephalogr. Clin. Neurophysiol.* **26**, 407–418 (1969).
3. T. C. Foster, C. A. Castro, B. L. McNaughton, *Science* **244**, 1580–1582 (1989).
4. F. Fuhrmann *et al.*, *Neuron* **86**, 1253–1264 (2015).
5. L. L. Colgin, *Annu. Rev. Neurosci.* **36**, 295–312 (2013).
6. G. Buzsáki, *Rhythms of the Brain* (Oxford Univ. Press, 2006).
7. J. O'Keefe, M. L. Recce, *Hippocampus* **3**, 317–330 (1993).
8. W. E. Skaggs, B. L. McNaughton, M. A. Wilson, C. A. Barnes, *Hippocampus* **6**, 149–172 (1996).
9. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsáki, *Science* **321**, 1322–1327 (2008).
10. M. R. Mehta, A. K. Lee, M. A. Wilson, *Nature* **417**, 741–746 (2002).
11. W. M. Sheeran, O. J. Ahmed, *Neurosci. Biobehav. Rev.* **108**, 821–833 (2020).
12. R. G. K. Munn, C. S. Mallory, K. Hardcastle, D. M. Chetkovich, L. M. Giocomo, *Nat. Neurosci.* **23**, 239–251 (2020).
13. J. A. Pérez-Escobar, O. Kornienko, P. Latuske, L. Kohler, K. Allen, *eLife* **5**, e16937 (2016).
14. M. G. Campbell *et al.*, *Nat. Neurosci.* **21**, 1096–1106 (2018).
15. M. Iwase, T. Kitanishi, K. Mizuseki, *Sci. Rep.* **10**, 1407 (2020).
16. E. Kropff, J. E. Carmichael, M. B. Moser, E. I. Moser, *Nature* **523**, 419–424 (2015).
17. D. Justus *et al.*, *Nat. Neurosci.* **20**, 16–19 (2017).
18. J. R. Hinman, M. P. Brandon, J. R. Climer, G. W. Chapman, M. E. Hasselmo, *Neuron* **91**, 666–679 (2016).
19. J. D. Green, A. A. Arduini, *J. Neurophysiol.* **17**, 533–557 (1954).
20. H. Dannenberg *et al.*, *J. Neurosci.* **35**, 8394–8410 (2015).
21. I. Zutshi *et al.*, *Curr. Biol.* **28**, 1179–1188.e3 (2018).
22. I. J. Kirk, N. McNaughton, *Neuroreport* **2**, 723–725 (1991).
23. I. J. Kirk, S. D. Oddie, J. Konopacki, B. H. Bland, *J. Neurosci.* **16**, 5547–5554 (1996).
24. B. Kocsis, R. P. Vertes, *J. Neurosci.* **14**, 7040–7052 (1994).
25. N. P. Pedersen *et al.*, *Nat. Commun.* **8**, 1405 (2017).
26. H. T. Ito, E. I. Moser, M. B. Moser, *Neuron* **99**, 576–587.e5 (2018).
27. S. Chen *et al.*, *Nature* **586**, 270–274 (2020).
28. R. P. Vertes, *J. Comp. Neurol.* **326**, 595–622 (1992).
29. A. M. Lee *et al.*, *Neuron* **83**, 455–466 (2014).
30. J. S. Thinschmidt, G. G. Kinney, B. Kocsis, *Neuroscience* **67**, 301–312 (1995).
31. W. X. Pan, N. McNaughton, *Prog. Neurobiol.* **74**, 127–166 (2004).
32. T. Ino *et al.*, *Neurosci. Lett.* **90**, 259–264 (1988).
33. L. E. Fenno *et al.*, *Nat. Methods* **11**, 763–772 (2014).
34. F. Billwiller *et al.*, *Brain Struct. Funct.* **225**, 2643–2668 (2020).
35. Y. Li *et al.*, *Neuron* **97**, 911–924.e5 (2018).
36. V. Caggiano *et al.*, *Nature* **553**, 455–460 (2018).
37. B. Zingg *et al.*, *Neuron* **93**, 33–47 (2017).
38. M. M. Carvalho *et al.*, *Cell Rep.* **32**, 108123 (2020).

ACKNOWLEDGMENTS

The authors thank H. Ito, E. Moser, and M.-B. Moser for generously sharing a previously published dataset for reanalysis for the specific purposes of this study and for providing comments on the manuscript. The authors also thank R. Kumar and S. Felong for technical assistance. **Funding:** This work was supported by a Canadian Institutes of Health Research postdoctoral fellowship (J.S.F.); National Institute of Mental Health (NIMH) K99MH11284002 (M.L.-B.); an American Epilepsy Society (AES) Junior Investigator Award (F.T.S.); Stanford Epilepsy Training Grant 5T32NS007280 funded by the National Institute of Neurological Disorders and Stroke (NINDS) (P.M.K. and E.H.); a Swiss National Science Foundation Postdoctoral Fellowship (T.G.); National Science Foundation Fellowship DGE-114747 (B.A.); a HHMI Gilliam Fellowship for Advanced Study (B.A.); Gates Millennium Scholarship (B.A.); a Japan Society for the Promotion of Science (JSPS) Overseas Fellowship (S.T.); an AES Postdoctoral Fellowship (B.D.); NINDS K99NS117795 (B.D.); National Institute of Health (NIH) 1U19NS104590 (I.S., A.L., and M.J.S.); NIMH 1R01MH124047 and 1R01MH124867 (A.L.); the Kavli Foundation (A.L.); and the NIH, NSF, Gatsby, Fresenius, and NOMIS Foundations (K.D.). **Author contributions:** Conceptualization: J.S.F. and I.S. Methodology: B.A., M.O., and G.S. Investigation: J.S.F., M.L.-B., P.M.K., F.T.S., T.G., A.L.O., and S.B. Formal analysis: J.S.F. M.L.-B., P.M.K., F.T.S., T.G., S.T., M.O., E.H., and B.D. Funding acquisition: M.J.S., K.D., A.L., and I.S. Supervision: M.J.S., K.D., A.L., I.S. Writing – original draft: J.S.F. and I.S. Writing – review & editing: all authors. **Competing interests:** M.J.S. is a scientific cofounder of Inscopix. GRIN lenses used in this studied were purchased from Inscopix. **Data and materials availability:** Code used in this study came from publicly available resources referenced in the materials and methods. Parts of the raw datasets and additional custom code are openly available at solteszlab.com/datasets and will continue to be formatted and uploaded.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh4272
Materials and Methods
Figs. S1 to S17
References (*39*–*65*)
Movie S1

11 March 2021; accepted 29 October 2021
10.1126/science.abh4272

PHYSIOLOGICAL STRESS

Physiological costs of undocumented human migration across the southern United States border

Shane C. Campbell-Staton[1,2,3]*†, Reena H. Walker[4]†, Savannah A. Rogers[5]†‡, Jason De León[6]*, Hannah Landecker[2,7], Warren Porter[8], Paul D. Mathewson[8], Ryan A. Long[4]*

Political, economic, and climatic upheaval can result in mass human migration across extreme terrain in search of more humane living conditions, exposing migrants to environments that challenge human tolerance. An empirical understanding of the biological stresses associated with these migrations will play a key role in the development of social, political, and medical strategies for alleviating adverse effects and risk of death. We model physiological stress associated with undocumented migration across a commonly traversed section of the southern border of the United States and find that locations of migrant death are disproportionately clustered within regions of greatest predicted physiological stress (evaporative water loss). Minimum values of estimated evaporative water loss were sufficient to cause severe dehydration and associated proximate causes of mortality. Integration of future climate predictions into models increased predicted physiological costs of migration by up to 34.1% over the next 30 years.

> "You need to put yourself into the most difficult places that you can where [Border Patrol] can't get to. You understand? Where there are lots of trees, mountains, rocks…off the trail. That's where you need to go. If you walk in the easiest places, they will catch you quick."
> —Lucho, 47-year-old migrant from Jalisco, Mexico, interviewed June 2009 [(*1*), p. 189]

In 1994, the US Border Patrol adopted the border enforcement strategy referred to as Prevention Through Deterrence. This policy sought to dissuade undocumented entry across the southwest border of the United States by fortifying official ports of entry and their surrounding areas for the purpose of redirecting migrants toward more remote regions, including areas of the Chihuahuan and Sonoran Deserts (*1*, *2*). Desert environments are

among the most extreme biomes on the planet and are characterized by high maximum temperatures, large daily thermal fluctuations, and minimal water availability, which challenge biological function. In the decades since the implementation of Prevention Through Deterrence, thousands of migrants have perished in the desert while attempting to circumvent border protection efforts (354.8 ± 71.07 deaths/year; 7805 total reported deaths from 1998 to 2019; data S1), with many more deaths likely unreported (*3*). Interviews with surviving migrants depict experiences of extreme thermohydric stress—dehydration, disorientation, and organ failure—as common elements of the migrant journey:

> "We were dying of thirst. I was hallucinating at that point. We were surrounded by dirt but I kept seeing water everywhere in the desert."
> —Lucho, 47-year-old migrant from Jalisco, Mexico, interviewed June 2009 [(*1*), p. 193]

> "They abandoned me on the mountains. I didn't have food; I didn't have any water… [Our smuggler] abandoned us… He left us without food or water. I felt like I was almost dead. I was drinking my urine and trying to eat cactus. I am doing this for my children."
> —Monica, 27-year-old migrant from Mexico, interviewed May 2013 by J.D.L.

Although the southwest border accounts for 97.86% of total undocumented migrant apprehensions in the United States each year (2000–2019; data S2), the physiological challenges faced by humans attempting to traverse the desert terrain along southern migration routes remain largely unstudied [however, see (*4*, *5*)]. In contemporary society, causes of preventable human death are typically subject to extensive research. Yet unverified assumptions about heat stress and water loss as contributors to undocumented migrant mortality have greatly outpaced empirical work. Ethnographic insights can be synergistically coupled with physiological modeling to better understand how the extremes of the desert environment affect migrant physiology and risk of mortality. In this study, we model the physiological costs of undocumented migration across a portion of the Tucson Sector, a Border Patrol jurisdiction that runs from Yuma, Arizona, to the New Mexico border. Arizona contains 46.9% of the total primary barrier mileage (pedestrian and vehicle) across the southwest border (data S3), and the Tucson Sector, which has been a primary crossing point for migrants for nearly two decades, is characterized by the highest number of known border-crossing fatalities (*1*).

To gain a clearer understanding of human physiological stress in the face of desert migration, we integrated migrant interview data with data on human physiology, morphology, and fine-scale climatic variation to parameterize a spatiotemporally explicit biophysical model, Niche Mapper (*6*). Niche Mapper is based on fundamental principles of heat and mass exchange between a model organism and its environment and solves the energy-balance equation using two submodels that integrate detailed data about the organism and the microclimate of its environment (*6*–*9*). Model outputs include estimated rates of metabolism and/or evaporative water loss necessary for maintaining homeothermy in the modeled environment. We used this model to simulate the physiological costs of migration by estimating rates of evaporative water loss necessary for humans to maintain heat balance while traveling by foot across the desert between Nogales, Mexico, and Three Points, Arizona (*1*, *2*, *4*).

Age, body size, sex, and reproductive status can have substantial impacts on the costs of thermoregulation (*10*–*12*), and several studies have reported a sex bias among migrants, with men being more likely to migrate than women (*13*, *14*). More recent data, however, have shown marked increases in the number of family units and unaccompanied minors (≤17 years old) apprehended along the southwest border (net increases for 2013–2019 for family units of +3189%, 14,855 to 473,682; and for unaccompanied minors of +196%, 38,759 to 76,020; data S4 and S5). Therefore, we used literature-derived, region-specific values of average body mass and physiological parameters to model four demographics representative

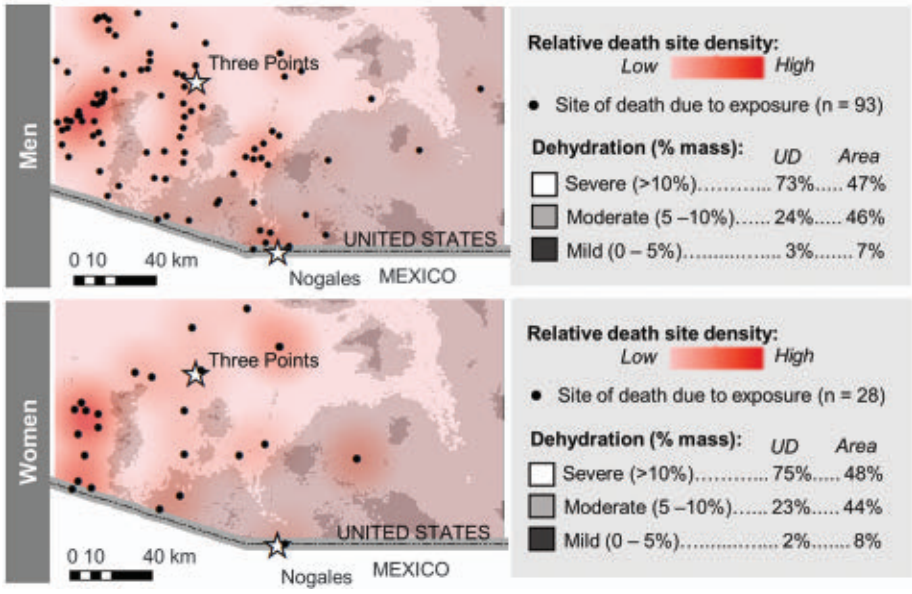


Fig. 1. Spatial relationship between density of migrant death sites and severity of thermohydric costs. Utilization distributions (red scale bar) of adult male (top) and female (bottom) migrant death sites due to exposure overlaid on a spatiotemporally explicit map of the predicted costs (clinical dehydration level as percentage of body mass lost per day through evaporative water loss; gray scale bar) of traveling on foot during June in southern Arizona. Nogales and Three Points are denoted by white stars. Georeferenced migrant death sites (black dots; men, n = 93; women, n = 28) were extracted from Arizona's OpenGIS Initiative for Deceased Migrants database (*21*). Only sites with known sex, age (>18 years old), cause of death (exposure), and month of death (May to September) are included. Red shading depicts the volume of the utilization distribution (95% fixed kernel); darker red indicates areas of greater kernel density volume (i.e., where more deaths occurred). Locations predicted to induce severe dehydration (>10% body mass deficit through evaporative water loss) in June are mapped in white, moderate dehydration (5 to 10% body mass deficit) in light gray, and mild dehydration (0 to 5% body mass deficit) in dark gray (*22*–*24*). The geographic distributions of male nd female deaths were disproportionately concentrated in regions predicted to induce severe dehydration. See figs. S1 and S2 for results from all months.

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. [3]Institute for Society and Genetics, University of California, Los Angeles, CA, USA. [4]Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID, USA. [5]Bioinformatics and Computational Biology, University of Idaho, Moscow, ID, USA. [6]Department of Anthropology and Chicana, Chicano, and Central American Studies, University of California, Los Angeles, CA, USA. [7]Sociology Department, University of California, Los Angeles, CA, USA. [8]Department of Integrative Biology, University of Wisconsin, Madison, WI, USA.
***Corresponding author. Email: scampbellstaton@princeton.edu (S.C.C.-S.); ralong@uidaho.edu (R.A.L.); jdeleon@ucla.edu (J.D.L.)**
†These authors contributed equally to this work.
‡Present address: Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St. Andrews, UK.

of humans attempting to cross the southern US border: (i) a Latin American adult man (60.3 kg), (ii) a nonpregnant Latin American adult woman (57.7 kg), (iii) a 5-year-old Latin American child (19 kg), and (iv) a pregnant Latin American adult woman. [We estimated the costs of migrating during pregnancy by adding the average weight gain reported at 30 weeks of pregnancy (11.3 kg) to the weight of the modeled woman and adjusting physiological parameters to reflect changes due to pregnancy (*15*–*19*).] We validated our model predictions using a combination of published literature on human physiology and metabolic chamber simulations (*19*).

We used Niche Mapper to generate spatiotemporally explicit maps of predicted evaporative water loss in our study area (*9*, *20*). We then evaluated how geographic variation in evaporative water loss during summer (May to September), when most migrant deaths occur (*21*), influenced the distribution of migrant deaths due to exposure. We compiled georeferenced records of adult (>18 years old) male (n = 292) and female (n = 101) migrant deaths from Arizona's OpenGIS Initiative for Deceased Migrants database [1981–2019, n = 3251 (*21*)]. We then partitioned those records by the summer month in which each death occurred. Next, we estimated 95% fixed-kernel utilization distributions (UDs) of migrant deaths due to exposure in each month and calculated the proportion of the volume of each UD that fell within each of three categories of dehydration: mild (0 to 5% body mass lost per day through evaporative water loss), medium (5 to 10% body mass lost per day), or severe (>10% body mass lost per day) (*22*–*24*). Relative differences in costs across space and time were our primary interest in this analysis, and we assumed no water replacement.

The geographic distribution of adult male deaths from exposure was disproportionately concentrated in regions predicted to induce severe or moderate dehydration (i.e., regions with high predicted evaporative water loss) in all summer months ($P < 0.05$ in all monthly comparisons, equality of proportions test; Fig. 1 and fig. S1). By contrast, adult male deaths were either underrepresented ($P < 0.05$) or proportionally represented ($P > 0.05$) within regions that were predicted to induce only mild dehydration (i.e., low predicted evaporative water loss) during each month. Geographic clustering of adult female migrant deaths followed the same pattern; death sites were overrepresented in regions predicted to induce severe or moderate dehydration in all summer months ($P < 0.05$ in all monthly comparisons, equality of proportions test) but were underrepresented or proportionally represented in areas of mild predicted dehydration (Fig. 1 and fig. S2).

Owing to the vast expanse and remoteness of the Arizona desert, many migrant deaths in this region likely go unreported (*3*). Among those bodies that are discovered, many are located by accident, and, as of now, there is no concerted effort at the state or federal levels to systematically retrieve bodies across the region (*25*). Animal scavenging and harsh environmental conditions lead to rapid decomposition and scattering of remains, decreasing the likelihood of discovery while obscuring identification and cause of death (*3*). Although it is often assumed that such deaths are caused by exposure, little empirical data exist regarding how physiological stress may relate to broader geographic patterns of undocumented migrant death throughout the study area.

Thus, we repeated the analysis described above using sites of migrant death for which the cause and timing of death was uncertain (adult males: n = 156; adult females: n = 37; fig S3). We averaged estimates of evaporative water loss across all summer months to account for uncertainty in the estimated time of death. Similar to our previous results, we found that deaths of uncertain cause among both adult men and women were disproportionately clustered within regions predicted to induce severe dehydration ($P < 0.05$ in both comparisons, equality of proportions test; fig. S3) but were significantly underrepresented within areas of moderate and mild dehydration ($P < 0.05$ in both comparisons, equality of proportions test; fig. S3). Though it is difficult to determine the exact proportion of these deaths that were directly due to dehydration and thermohydric stress, the significant correlation between high levels of predicted evaporative water loss and the density of deaths (indicated by the volume of the UD; figs. S4 to S6) strongly implicate temperature and water availability as major contributors to broader patterns of migrant mortality during summer.

To estimate the minimum water requirements associated with travel from Nogales to Three Points, we calculated least-cost paths for each modeled individual (man, nonpregnant woman, pregnant woman, child) during each summer month using evaporative water loss

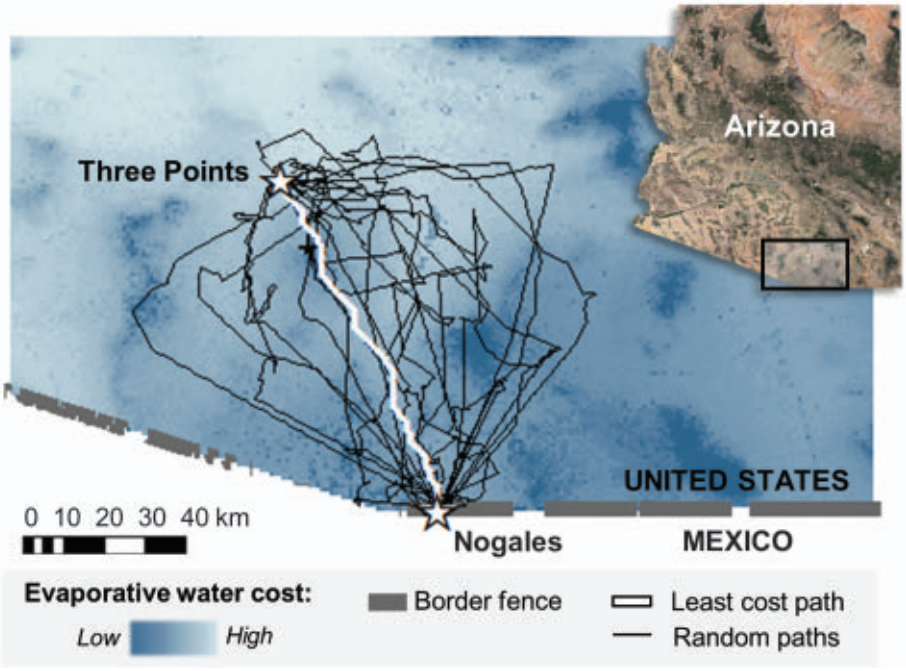


Fig. 2. Potential migration paths between Nogales and Three Points relative to spatial variation in thermohydric costs. Spatial variation in predicted physiological costs (evaporative water loss, liters/day) incurred by undocumented migrants at the southern US border, with least-cost and random paths used to calculate the distribution of estimated water loss incurred by migrants walking between Nogales and Three Points overlaid. The two cities lie on opposite sides of the border between the United States and Mexico, much of which is blocked by a border fence (thick gray lines) that physically separates the two countries (*50*). Random paths (black lines) and the least-cost path (white line) between Nogales and Three Points are overlaid on a raster (800-m resolution) that shows the predicted rate of evaporative water loss across the study area for an adult man traveling on foot during daytime hours in June. Paths were used to quantify the distribution of potential water costs for each combination of modeled individual (man, nonpregnant woman, pregnant woman, child), activity pattern (diurnal, nocturnal), departure time, and summer month (May to September). Satellite imagery for Arizona was adapted from Google Earth Pro (*51*). Methods for extracting spatiotemporally explicit water costs along paths are outlined in fig. S7. Results for all demographics are presented in figs. S8 to S11.

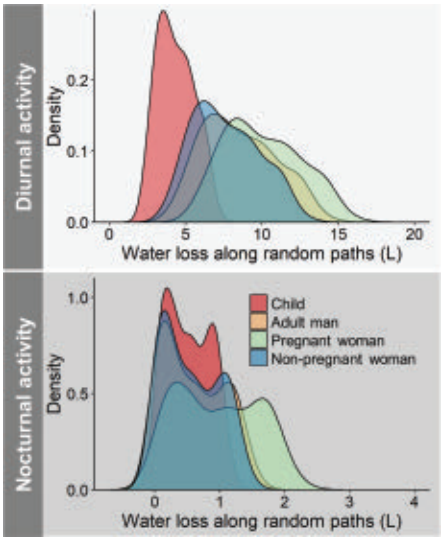


Fig. 3. Water-loss distributions for migrants traveling along random paths between Nogales and Three Points. Distribution of predicted water costs for diurnal (top) versus nocturnal (bottom) travel by four migrant demographics traversing random paths (n = 20) between Nogales and Three Points during summer (May to September). We used a custom algorithm [associated R package: *atob* (*28*)] to generate random trajectories (i.e., paths) at discrete time steps (15-min-step time interval) between an origin (Nogales) and destination (Three Points) using maximum travel speed, total travel time, minimum spatial resolution, and the desired number of routes as user-supplied parameters (*19*). We calculated total water loss along each random path by overlaying paths on spatio-temporally explicit rasters of predicted evaporative water loss for each combination of demographic, activity pattern, departure time, and month, summing the cell-specific evaporative water loss estimates along the path and multiplying the hour-specific rate of water loss in the cell (liters/s) by the time spent in the cell (s) given the modeled migrants' time of departure and movement speed (see materials and methods for a detailed description of this analysis). Diurnal travel produced substantially higher and more variable rates of water loss than nocturnal travel for all migrant demographics.

and terrain slope as equally weighted cost factors (*19*). In addition, because time of day affects costs of thermoregulation (*26*), we modeled least-cost paths under both diurnal and nocturnal activity scenarios. To quantify the cost of traveling along least-cost paths in each modeled scenario (i.e., each combination of migrant demographic, activity pattern, departure time, and month), we (i) extracted the hour-specific evaporative water-loss estimate for each raster cell intersecting the path, (ii) calculated the cell-specific evaporative water loss (liter) by multiplying the hour-specific rate of water loss in the cell (liters/s) by the time spent in the cell (s) given the migrants' time of departure and movement speed, and (iii) summed the cell-specific costs accrued along the path (fig. S7). To capture the full distribution of evaporative water loss along each path, we calculated the cost of traversing the path using all possible hourly start times for both diurnal and nocturnal travel (*19*). Personal accounts of individual journeys across this landscape suggest that in many instances, migrants do not have the knowledge to identify a least-cost path of travel and may instead wander through the desert with little information about the terrain:

> "[We were] in the mountains going up and down, up and down…everybody couldn't take it so the guide abandoned them. I also couldn't take it anymore. I tried again and my toenails fell off. All of these nails came off before. All of them fell off. All of them."
> —Maria, 30-year-old migrant from Oaxaca, Mexico, interviewed May 2013 by J.D.L.

> "When I was walking the last day, I felt like I was already there. I [hallucinated] a lot of things. The last hours, to get to where they were going to pick us up, I didn't see anything. Everything was spinning. I was dehydrated - I don't know what! It hurt here, in my chest."
> —Yoanna, 27-year-old migrant from Puebla, Mexico, interviewed May 2013 by J.D.L.

Therefore, we also generated random paths of migration between Nogales and Three Points using a custom random trajectory algorithm [n = 20 paths per modeled individual per month; Fig. 2; *atob* package (v0.0.0.9000) in program R (v3.5.3) (*27*–*29*)]. By combining elements of random walks and Brownian bridge movement models, the algorithm allowed us to calculate random trajectories between an origin and a destination given a user-supplied time budget for 2 days of travel (a "best-case" scenario based on ethnographic interview data), a maximum movement speed of 1.2 m/s, and a step length (i.e., distance between successive locations along the path) of 800 m (*19*). We quantified the cost of traversing random paths in each modeled scenario following the procedure described for least-cost paths [fig. S7; (*19*)].

Predicted evaporative water loss differed significantly among months and times of day for all migrant demographics. The cost (cumulative evaporative water loss) of traveling along least-cost paths ranged from 0.01 liters (the minimum for all demographics) when traveling only at night in May to 11.9 liters for a pregnant woman traveling during the day in June (figs. S8 to S11). Costs of traveling along more realistic, random paths were considerably higher (mean ± SD: 2.27 ± 0.003 liters higher for random paths across all scenarios). Predicted evaporative water loss peaked in June for all demographics regardless of the time of activity (linear mixed-effects model: $P < 0.001$, Wald test; figs. S8 to S11). Within each month, it was significantly more costly to travel during the day than at night (linear mixed-effects model: $P < 0.001$, Wald test; figs. S8 to S11). The distribution of thermohydric costs incurred during ~48 hours of travel from Nogales to Three Points along random paths was strongly dependent on body size, sex, and reproductive status (Fig. 3). Estimated costs of traveling at night ranged from 0.69 ± 0.49 liters (mean ± SD) for a child to 1.29 ± 0.92 liters for a pregnant woman (Fig. 3). The costs of daytime travel were much higher and ranged from 4.66 ± 1.29 liters for a child to 10.76 ± 2.72 liters for a pregnant woman (Fig. 3). In the absence of water replacement by drinking, these rates of evaporative water loss are sufficient to cause severe dehydration and a diversity of associated conditions such as renal failure, electrolyte abnormalities, uremia, and coronary and cerebral thrombosis that are likely the proximate causes of many migrant deaths (*30*–*38*).

These models reflect contemporary climate conditions, but both changes in climate and human migration are likely to drive future patterns of mortality across the desert (*39*). Population surveys, econometrics, and climate analyses have documented increased human migration globally in association with changing climatic conditions (*40*–*44*). Yet the physiological costs of human migration under such conditions have received relatively little attention. As desert environments become more extreme in the coming decades (*45*, *46*), thermohydric stress and associated risk of mortality endured by migrant populations are also likely to increase. We next modeled the physiological costs of migration (evaporative water loss) based on anticipated climate change over the next 30 years under representative concentration pathway (RCP) 4.5 (i.e., the intermediate mitigation scenario (*47*). We reparameterized the microclimate submodel of Niche Mapper with monthly minimum and maximum temperatures for 2050 by averaging predictions from six locally downscaled, publicly available climate projection models across our study area: ACCESS-1, CanESM2, CCSM4, CNRM-CM5, CSIRO-Mk3, and INM-CM4 (*48*). We chose a 30-year time horizon to facilitate direct comparison with our contemporary climate models, which were parameterized using average temperatures from the previous 30 years (*49*).

Our analysis revealed a significant increase in thermohydric costs of migration for each human demographic under the future climate change projection (linear mixed-effects model, $P < 0.001$, Wald test; Fig. 4). Projected increases in temperature across our study area

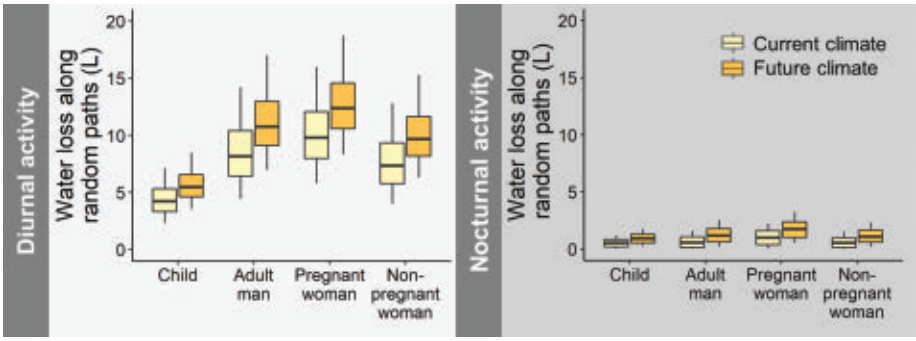


Fig. 4. Predicted costs of migration under current versus future climate scenarios. Distribution of predicted water costs for diurnal (left) versus nocturnal (right) travel by four migrant demographics traversing random paths (n = 20) between Nogales and Three Points during summer (May to September) based on previous 30-year monthly average temperatures ("current climate") and locally downscaled climate projections of temperature change over the next 30 years ("future climate"). Pixel-specific values for future climate conditions were averaged across six publicly available, locally downscaled climate models (ACCESS-1, CanESM2, CCSM4, CNRM-CM5, CSIRO-Mk3, INM-CM4) that projected monthly maximum and minimum temperatures in 2050 under RCP 4.5 (i.e., the intermediate mitigation scenario) (*47*, *48*). Predicted water costs for migration on foot across the southern border increase significantly under this climate-change scenario relative to current conditions for all modeled migrant demographics (linear mixed-effects model, P <0.001, Wald test). Boxes and whiskers represent median, interquartile range, and range of predicted water loss.

under RCP 4.5 over the next 30 years were predicted to increase the average cost of migration along random paths between Nogales and Three Points by 34.1% for an adult man (1.81 ± 0.003 liters), 34.1% for a nonpregnant adult woman (1.63 ± 0.003 liters), 33.1% for a child (0.94 ± 0.001 liters), and 29.6% for a pregnant adult woman (1.90 ± 0.003 liters). Taken together, these results indicate that undocumented migration across the southwest border of the United States will become increasingly dangerous over the next 30 years, which will likely result in increased mortality of migrants.

Biophysical models such as those presented here have been previously used to understand the dynamics of thermohydric stress in nonhuman animals in ecologically relevant contexts. We present a new application of biophysical modeling to human systems, providing a methodological framework for future studies that seek to understand the impacts of past, present, and future climate on human physiology, stress, and evolution [see (*19*) for caveats and considerations]. Our findings demonstrate the utility of social scientific inputs to adapt these modeling tools to humans moving through extreme landscapes. We have used this approach to show that thermohydric stress is a major contributor to patterns of undocumented migrant mortality at the southern US border. Such quantitative frameworks for studying intersecting impacts of social policy and climate change on human stress and physiology will be of increasing importance as the climate warms (*39*), because these methods could be extended to other situations involving outdoor exertion in conditions that challenge physiological homeostasis.

REFERENCES AND NOTES

1. J. De León, M. Wells, *Land of Open Graves: Living and Dying on the Migrant Trail* (Univ. California Press, ed. 1, 2015).
2. R. Rubio-Goldsmith, M. McCormick, D. Martinez, I. Duarte, "The 'Funnel Effect' and recovered bodies of unauthorized migrants processed by the Pima County Office of the Medical Examiner, 1990-2005" (Technical Report SSRN-id3040107, Binational Migration Institute, 2006).
3. J. Beck, I. Ostericher, G. Sollish, J. De León, *J. Forensic Sci.* **60**, S11–S20 (2015).
4. G. A. Boyce, S. N. Chambers, S. Launius, *J. Migr. Hum. Secur.* **7**, 23–35 (2019).
5. J. De León, *Am. Anthropol.* **114**, 477–495 (2012).
6. W. P. Porter, J. W. Mitchell, US Patent US7155377B2 (2006).
7. P. D. Mathewson *et al.*, *J. Ther. Biol.* **94**, (2020).
8. M. Kearney, W. Porter, *Ecol. Lett.* **12**, 334–350 (2009).
9. Y. Natori, W. P. Porter, *Ecol. Appl.* **17**, 1441–1459 (2007).
10. M. Kleiber, *Physiol. Rev.* **27**, 511–541 (1947).
11. D. Gagnon, G. P. Kenny, *J. Physiol.* **590**, 5963–5973 (2012).
12. N. Charkoudian, N. S. Stachenfeld, *Compr. Physiol.* **4**, 793–804 (2014).
13. M. Cerrutti, D. S. Massey, in *Crossing the Border: Research from the Mexican Migration Project*, J. Durand, D.S. Massey, Eds. (Russel Sage Foundation, 2004), pp. 24–26.
14. D. S. Massey, C. Capoferro, *Int. Migr. Rev.* **38**, 1075–1102 (2004).
15. S. C. Walpole *et al.*, *BMC Public Health* **12**, 439 (2012).
16. R. Martorell, L. K. Khan, M. L. Hughes, L. M. Grummer-Strawn, *J. Nutr.* **128**, 1464–1473 (1998).
17. P. Durán *et al.*, *Acta Paediatr.* **105**, e116–e125 (2016).
18. K. M. Rasmussen, A. L. Yaktine, Eds., *Weight Gain During Pregnancy: Reexamining the Guidelines* (National Academies Press, 2009).
19. Materials and methods are available as supplementary materials.
20. M. R. Kearney, W. P. Porter, *Ecography* **40**, 664–674 (2017).
21. Arizona OpenGIS Initiative for Deceased Migrants, Custom map of migrant mortality (2019); https://humaneborders.info/app/map.asp.
22. K. E. D'Anci, F. Constant, I. H. Rosenberg, *Nutr. Rev.* **64**, 457–464 (2006).
23. World Health Organization (WHO), "The treatment of diarrhea: A manual for physicians and other senior health workers" (WHO/CDD/SER/80.2 rev. 4, WHO, 2005).
24. American Academy of Pediatrics Provisional Committee on Quality Improvement, *Pediatrics* **97**, 424–436 (1996).
25. M. Jimenez, "Humanitarian crisis: Migrant deaths at the U.S.-Mexico border" (American Civil Liberties Union, 2009).
26. W. P. Porter, *Ecol. Monogr.* **37**, 273–296 (1967).
27. G. Technitis, W. Othman, K. Safi, R. Weibel, *Int. J. Geogr. Inf. Sci.* **29**, 912–934 (2015).
28. S. Rogers, *atob*: Creates random routes between two points, GitHub (2019); https://github.com/SavannahAlicia/atob/.
29. R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, 2018); www.R-project.org/.
30. J. González-Alonso, R. Mora-Rodríguez, P. R. Below, E. F. Coyle, *J. Appl. Physiol.* **82**, 1229–1236 (1997).
31. H. R. Lieberman, *J. Am. Coll. Nutr.* **26**, 555S–561S (2007).
32. L. Nybo, *J. Appl. Physiol.* **104**, 871–878 (2008).
33. P. M. Clarkson, *Sports Med.* **37**, 361–363 (2007).
34. G. C. Donaldson, W. R. Keatinge, R. D. Saunders, *Int. J. Hyperthermia* **19**, 225–235 (2003).
35. C. L. Chapman *et al.*, *J. Appl. Physiol.* **128**, 715–728 (2020).
36. J. González-Alonso, C. G. Crandall, J. M. Johnson, *J. Physiol.* **586**, 45–53 (2008).
37. J. Tikkanen, O. P. Heinonen, *Eur. J. Epidemiol.* **7**, 628–635 (1991).
38. R. Boni, *Mol. Reprod. Dev.* **86**, 1307–1323 (2019).
39. H. Benveniste, M. Oppenheimer, M. Fleurbaey, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26692–26702 (2020).
40. V. Mueller, C. Gray, K. Kosec, *Nat. Clim. Chang.* **4**, 182–185 (2014).
41. A. Dillon, V. Mueller, S. Salau, *Am. J. Agric. Econ.* **93**, 1048–1061 (2011).
42. L. Marchiori, J. F. Maystadt, I. Schumacher, *J. Environ. Econ. Manage.* **63**, 355–374 (2012).
43. C. L. Gray, V. Mueller, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6000–6005 (2012).
44. T. Halliday, *Econ. Dev. Cult. Change* **54**, 893–925 (2006).
45. J. L. Weiss, J. T. Overpeck, *Glob. Change Biol.* **11**, 2065–2077 (2005).
46. S. R. Loarie *et al.*, *Nature* **462**, 1052–1055 (2009).
47. M. Collins *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker *et al.*, Eds. (Cambridge Univ. Press, 2013).
48. T. Wang, A. Hamann, D. Spittlehouse, C. Carroll, *PLOS ONE* **11**, e0156720 (2016).
49. C. Daly *et al.*, *Int. J. Climatol.* **28**, 2031–2064 (2008).
50. M. J. Cory, Border-fence map. Reveal from The Center for Investigative Reporting and OpenStreetMap contributors (2017); www.arcgis.com/home/item.html?id=2aac3acce15a4b56ab483c48243880d4.
51. Google Earth Pro V.7.3.3.7786, Arizona, United States of America. 111.093731, 34.048928. Eye alt 727.17 mi. SIO, NOAA, US Navy, NGA, GEBCO, 19 January 2021.

ACKNOWLEDGMENTS

We acknowledge the invaluable contributions of the men, women, and children whose personal experiences and stories informed the parameters of this study. We also acknowledge the innumerable lives, both known and unknown, that have been affected by the past and ongoing circumstances that contextualize this work. This study would have been impossible without the bravery and sacrifice of these individuals. **Funding:** Infrastructure support for this study was provided by the University of California, Los Angeles (UCLA); University of Idaho; and University of Wisconsin. **Author contributions:** S.C.C.-S. and R.A.L. conceived of and conceptualized this study. S.A.R., P.D.M., and W.P. developed the methodology described herein. R.H.W., S.A.R., P.D.M., and W.P. performed modeling and validation of the experiments performed. J.D.L. provided ethnographic data used to parameterize the models. All results were visualized by R.H.W. Supervision of experiments and their contextualization were provided by R.A.L., H.L., and J.D.L. The original draft of the manuscript was written by S.C.C.-S., H.L., and R.H.W., then reviewed and edited by S.C.C.-S., R.H.W., S.A.R., J.D.L., H.L., W.P., P.D.M., and R.A.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text and supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh1924
Materials and Methods
Figs. S1 to S18
Tables S1 to S8
References (*52*–*115*)
MDAR Reproducibility Checklist
Data S1 to S6

25 February 2021; resubmitted 4 June 2021
Accepted 21 October 2021
10.1126/science.abh1924

RADIATIVE COOLING

Scalable thermochromic smart windows with passive radiative cooling regulation

Shancheng Wang[1,2]†, Tengyao Jiang[3,4]†, Yun Meng[1,2], Ronggui Yang[5]*, Gang Tan[3]*, Yi Long[1,2,6]*

Radiative cooling materials spontaneously radiate long-wave infrared (LWIR) to the cold outer space, providing cooling power that is preferred in hot seasons. Radiative cooling has been widely explored for walls and roofs but rarely for windows, which are one of the least energy-efficient parts of buildings. We fabricated scalable smart windows using a solution process giving different emissivity (ε) at high ($\varepsilon_{LWIR\text{-}H}$ of 0.61) and low ($\varepsilon_{LWIR\text{-}L}$ of 0.21) temperatures to regulate radiative cooling automatically while maintaining luminous transparency and near-infrared (NIR) modulation. These passive and independent visible–NIR–LWIR regulated smart windows are capable of dynamic radiative cooling for self-adapting applications across different climate zones.

Buildings consume ~40% of global energy, and windows, one of the least energy-efficient parts, account for as much as 60% of their energy loss (*1*–*3*). In the United States, the window-associated heating and cooling energy consumption in buildings has been estimated at 4% of nation's total primary energy usage (*4*). Thermochromic windows are considered a cost-effective, stimulirational, energy-efficient smart window because of their simple structure, passive light modulation, and zero-energy input characteristics (*5*–*7*). The performance indexes of smart windows are luminous transmission (T_{lum}) and solar transmission modulation (ΔT_{sol}), where ΔT_{sol} is defined as the difference in T_{sol} (0.38 to 2.5 μm) between low and high temperatures (Fig. 1A) (*8*). However, the modulation of emissivity (ε) of long-wave infrared (ε_{LWIR}, 2.5 to 25 μm) in windows has rarely been studied.

Radiative cooling (RC) through LWIR spontaneously cools a surface by radiating thermal heat to the cold outer space. In recent years, there have been several applications using this strategy, including a subambient radiative cooler (*9*–*14*), energy-saving film (*15*), RC-based air-conditioning system (*16*–*18*), fabrics (*19*), roof (*20*), and energy-saving paint (*21*) (table S1). However, the fixed high ε_{LWIR} can only provide cooling when the weather is warm. Modulated RC catering to different seasonal weather conditions is needed for regions with hot and cold seasonal temperature fluctuations.

We propose a set of characteristics for an ideal smart window (Fig. 1A) that should have transparency in the visible range (380 to 780 nm) for both low and high temperatures. The T_{sol} has contributions from both the visible and near IR (NIR, 0.78 to 2.5 μm) ranges, which is responsible for heating a room. Changing the NIR state from opaque in the summer to transparent in winter is therefore desirable. Moreover, an ideal smart window should have a high ε_{LWIR} at high temperature ($\varepsilon_{LWIR\text{-}H}$) to promote RC during warm weather and a low ε_{LWIR} at low temperature ($\varepsilon_{LWIR\text{-}L}$) to suppress RC when it is cold outside.

Tuning ε_{LWIR} has been discussed theoretically (*22*–*24*) and demonstrated experimentally (*25*–*27*). However, these experiments tuned the ε_{LWIR} in an opposite way from what is helpful for windows (i.e., low $\varepsilon_{LWIR\text{-}H}$ and high $\varepsilon_{LWIR\text{-}L}$). We developed a transparent, passive RC regulating thermochromic (RCRT) smart window using near room temperature phase change vanadium dioxide (VO_2) and doped VO_2. We aimed to regulate both NIR transmission and RC spontaneously. Such an RCRT window fabricated by a solution process achieves luminous transparency with ΔT_{sol} (9.3%) and $\Delta\varepsilon_{LWIR}$ (defined as the difference between $\varepsilon_{LWIR\text{-}L}$ and $\varepsilon_{LWIR\text{-}H}$) of 0.4. Our energy consumption simulations suggest that the RCRT window gives a higher energy saving than commercial low-E glass across different climate zones.

For our RCRT window, the NIR is blocked in summer to prevent solar heating while the visible transparency remains unchanged (Fig. 1B). This is different from hydrogel-based smart windows, which turn opaque in the summer (*28*, *29*). Because RCRT windows have a high front side LWIR emissivity ($\varepsilon_{LWIR\text{-}H,Front}$), the RC helps to lower the cooling load further. In winter, both visible and NIR are transmitted while $\varepsilon_{LWIR\text{-}L,Front}$ is switched to lower values, reducing RC heat loss. Here, we kept $\varepsilon_{LWIR\text{-}Back}$ low in both conditions to minimize the heat exchange between the indoors and outdoors.

We designed and fabricated our RCRT window using spin coating, resulting in VO_2/spacer/low-E stacking (Fig. 1C) to form a Fabry-Perot resonator (*30*). Such a resonator has low resonance to the LWIR at low temperature but shows a strong Fabry-Perot resonation effect to give much-enhanced LWIR absorption due to the insulator-to-metallic transformation of VO_2 (*31*). The stacking consists of a VO_2 nanocomposite overcoat (Fig. 1D), a poly(methyl methacrylate) (PMMA) lossless spacer layer due to high solar and LWIR transparency (fig. S1), and two layers of indium tin oxide (ITO) low-E coating due to its high visible transmittance (~80%; fig. S2) and low ε_{LWIR} (~0.1; fig. S3). The VO_2 nanoparticles (figs. S4 to S6) are dispersed in PMMA solution and spin coated on PMMA spacer, serving the core functionality of T_{sol} and RC modulation. We compared photos of two transparent smart windows with and without RC regulation with the size up to 5 × 5 cm at both low and high temperatures (Fig. 1E). This solution-based fabrication process provides the potential for scaling up to building-sized windows.

We determined the optical spectra of our RCRT window (Fig. 2A) with T_{lum} of 27.8 and 26.1% at 20° and 90°C, respectively, and a ΔT_{sol} of 9.3%. The low transmittance of wavelength longer than 1500 nm was caused by the strong NIR blocking of the ITO coating (fig. S7). The transition temperature (τ_c) of the VO_2 layer was ~60°C, which can be tuned through doping (fig. S8). The RCRT window showed a promising ε_{LWIR} switching performance. At 20°C, the multilayer structure had a ε_{LWIR} of 0.21 and increased to 0.61 above τ_c, with a $\Delta\varepsilon_{LWIR}$ of 0.40. As a comparison, the spectra of the VO_2 sample without RC regulation (Fig. 2B) gives a typical T_{lum} of 32.2% and ΔT_{sol} of 17% but a negligible $\Delta\varepsilon_{LWIR}$ ($\varepsilon_{LWIR\text{-}L}$: 0.82, $\varepsilon_{LWIR\text{-}H}$: 0.85). The stacking design (Fig. 1D) leads to the favorable $\Delta\varepsilon_{LWIR}$ (Fig. 2A).

We collected images with an IR camera of both samples (Fig. 2C) with the background ε_{LWIR} of 0.5. We observed that from 30° to 60°C, the RCRT window was darker than the background, indicating a lower ε_{LWIR} compared with background (0.21 versus 0.5) because high ε_{LWIR} gives high thermal radiation intensity. With temperature above 70°C, the RCRT window turned brighter than the background because of a higher ε_{LWIR} compared with background (0.61 versus 0.5). By contrast, the sample with no RC regulation showed a constant brighter color than the background, which further demonstrates that it fails to modulate its ε_{LWIR} without structure manipulation.

[1]School of Materials Science and Engineering, Nanyang Technological University, Singapore, 639798, Singapore. [2]Singapore-HUJ Alliance for Research and Enterprise (SHARE), Campus for Research Excellence and Technological Enterprise (CREATE), Singapore, 138602, Singapore. [3]Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, Laramie, WY 82071, USA. [4]School of Environmental Science and Engineering, Nanjing Tech University, Nanjing, Jiangsu 211816, China. [5]State Key Laboratory of Coal Combustion, School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China. [6]Sino-Singapore International Joint Research Institute (SSIJRI), Guangzhou 510000, China.
*Corresponding author. Email: longyi@ntu.edu.sg (Y.L.); gtan@uwyo.edu (G.T.); ronggui@hust.edu.cn (R.Y.)
†These authors contributed equally to this work.

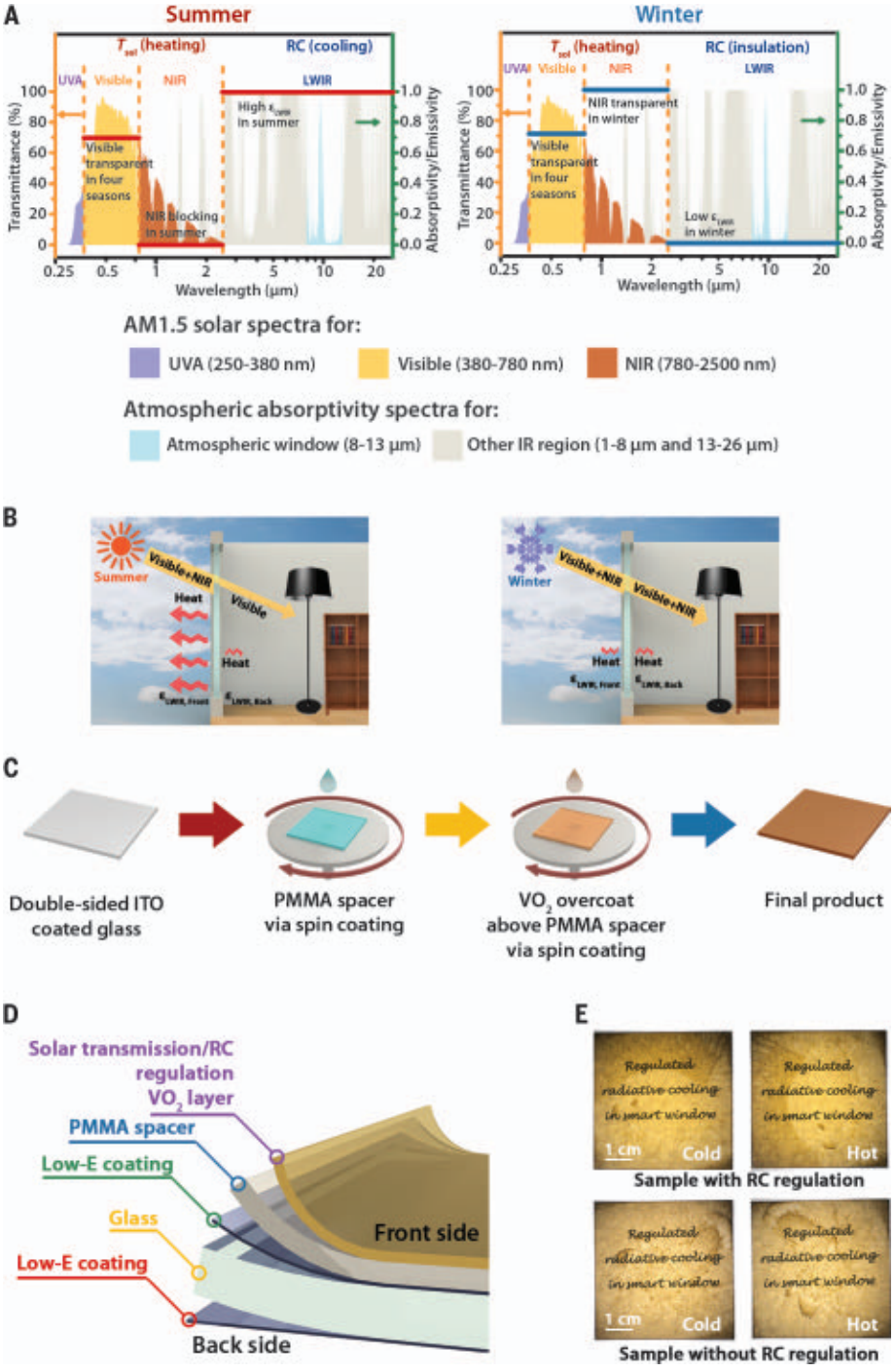


Fig. 1. Structure and concept of an RCRT window. (**A**) Concept of the ideal energy-saving smart window. The red and blue lines represent the spectra for an ideal energy-saving smart window in summer and winter. T_{sol} and thermal radiation in LWIR contribute to heating and cooling, respectively. Visible transparency remains constant. Windows should be NIR blocking and have high $\varepsilon_{LWIR\text{-}H}$ to promote radiative cooling in summer, and the response must be switched to the opposite in the winter. (**B**) Working principle of RCRT window in summer (left) and winter (right). The yellow arrows indicate the T_{sol} and the red arrows represent heat radiation. In summer, the NIR is blocked and visible light transmits through the window with thermal radiation enhanced. In winter, both NIR and visible light enter the room. Heat radiation is suppressed to minimize heat loss. (**C**) Schematics of the fabrication process for the RCRT window. (**D**) Schematic structure of the RCRT window. (**E**) Photos of the RCRT window with RC regulation (top) and the VO_2 window without RC regulation (bottom) at low and high temperature. The size of samples is 5 × 5 cm.

We systematically studied the tunability of the ε_{LWIR} of RCRT window by adjusting the thickness of the PMMA spacer and the VO_2 nanoparticle weight ratio in the overcoat. We established a relationship among PMMA thickness, VO_2 weight ratio, and ε_{LWIR} (fig. S9). By fixing a VO_2 weight ratio of 20:1, $\Delta\varepsilon_{LWIR}$ gradually increased from 0.28 to the maximum of 0.4 and dropped to 0.08 with increasing spacer thickness. We observed a similar trend of $\Delta\varepsilon_{LWIR}$ relating to spacer thickness in our optical simulation using the finite difference time domain method with the VO_2 layer, PMMA spacer, and double-sided ITO glass (fig. S10). Three representative samples are presented, along with their spectra, using different ε_{LWIR} modulation abilities, max $\Delta\varepsilon$ ($\varepsilon_{LWIR\text{-}L}$=0.21 and $\varepsilon_{LWIR\text{-}H}$ = 0.61), max ε ($\varepsilon_{LWIR\text{-}L}$ = 0.47 and $\varepsilon_{LWIR\text{-}H}$ = 0.67), and min ε ($\varepsilon_{LWIR\text{-}L}$ = 0.13 and $\varepsilon_{LWIR\text{-}H}$ = 0.36) (Fig. 3A).

To evaluate the impact of T_{sol} and ε_{LWIR} in different climate zones (*32*), we mapped the heating and cooling energy consumption per window area versus T_{sol} (10 to 90%) and ε_{LWIR} (0.1 to 0.9) across seven climate zones (Fig. 3B for zone 4 and figs. S12 to S17 for the rest of the zones). In our simulation, we used a single-story small office building as a prototype (fig. S11 and table S2A). We observed that in all seven zones, the minimum energy consumption (E_{min}) for cooling happened at T_{sol} = 10% and ε_{LWIR} = 0.9, and the E_{min} for heating occurred at ε_{LWIR} = 0.1 and T_{sol} = 90%. This agrees well with our proposed ideal window (Fig. 1A).

To further investigate the importance of RC regulation, we fixed T_{sol} to 30% for building energy simulation, considering that the building energy standard of American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) recommends the assembly maximum solar heat gain coefficient of 0.21 to 0.40 for climate zones 1 to 7. (*32*) We summarized the total E_{min} by adding heating and cooling from mappings (table S4), and the best-performing static ε_{LWIR} was 0.9 for zones 1 to 4 and 0.1 for the colder zones 6 to 7 (Fig. 3C and fig. S18). Comparatively, the dynamic ε_{LWIR} sample (ε_{LWIR} of 0.9 for cooling and ε_{LWIR} of 0.1 for heating) was able to save up to 124.0 MJ m^{-2} energy annually compared with the best-performing static ε_{LWIR} samples. This comparison demonstrates the capability of a dynamic ε_{LWIR} modulation window to cater the energy-saving demand in different climates.

We conducted energy consumption simulation of a full-scale building with the model of a typical 12-story large office building with a floor dimension of 73 m (240 ft) length × 49 m (160 ft) width × 4 m (13 ft) height and a total window glass area of 4636 m^2 (fig. S19 and table S2B). Whereas the RCRT window with pure VO_2 had a high τ_c of ~60°C, a tungsten-doped VO_2 (W-doped VO_2) max $\Delta\varepsilon$ window, with its low τ_c of ~27.5°C (fig. S20 and table

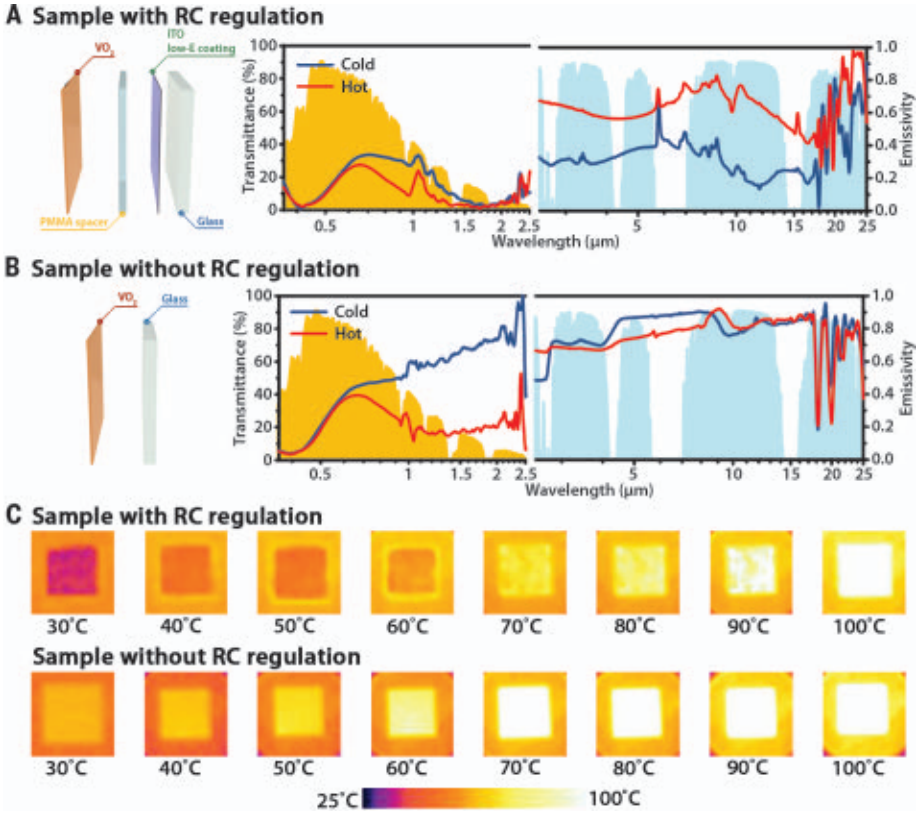


Fig. 2. Optical and thermal performance of the RCRT window. (**A** and **B**) Spectra of the sample with (A) and without (B) RC regulation in the visible-NIR and LWIR range at 20°C (blue line) and 90°C (red line) against a normalized AM1.5 global solar spectrum (yellow shadow) and LWIR atmospheric transmittance window (blue shadow), respectively. The schematics on the left illustrate the structure of the samples with and without RC regulation, respectively. Both samples give solar modulation, but the RCRT window shows a low ε_{LWIR} at 20°C and a high ε_{LWIR} at 90°C, whereas the ε_{LWIR} change for the sample without RC regulation is small. (**C**) IR camera image of the RCRT window (top) and the VO_2 sample without RC regulation (bottom) with temperature increasing from 30° to 100°C. The ε_{LWIR} for background is 0.5. The darker color of the sample with RC regulation compared with background from 30 to 60°C indicates that its ε_{LWIR} is <0.5, whereas a brighter color from 70° to 100°C indicates that ε_{LWIR} switches >0.5. By contrast, from 30° to 100°C, the sample without RC regulation is brighter than the background.

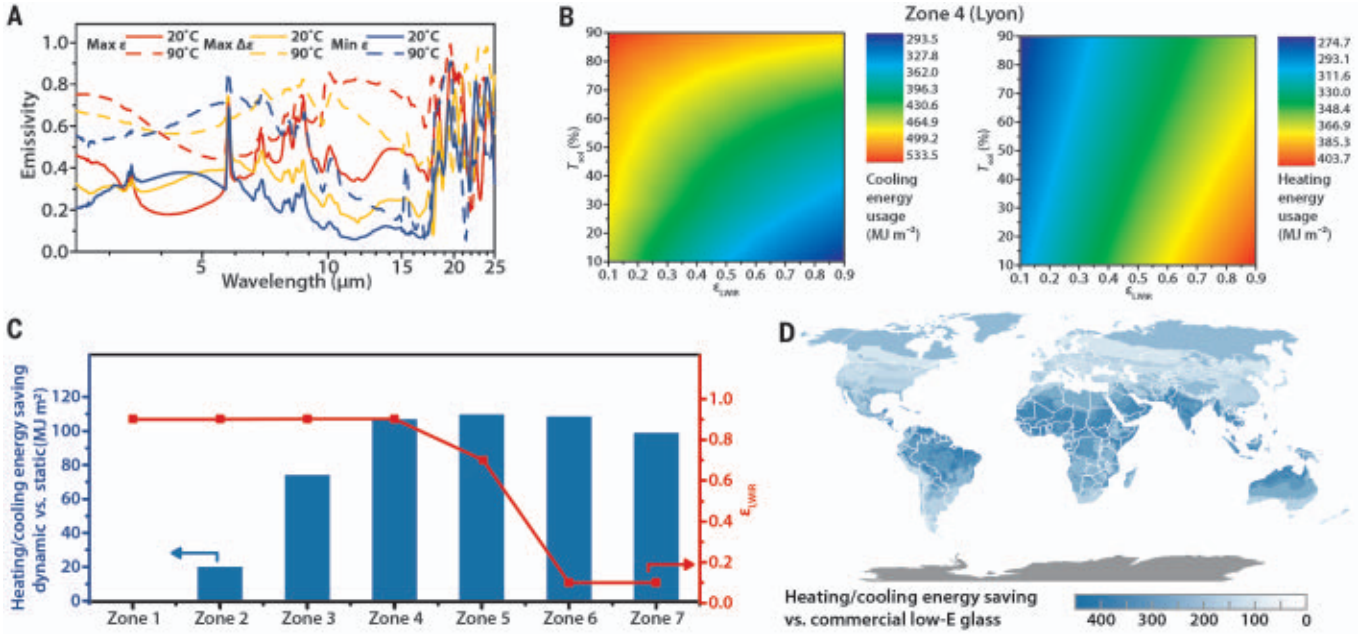


Fig. 3. ε_{LWIR} emissivity regulation and energy consumption simulation for the RCRT window. (**A**) $\varepsilon_{LWIR\text{-}L}$ (solid line) and $\varepsilon_{LWIR\text{-}H}$ (dotted line) spectra for max ε (red line), max Δε (yellow line), and min ε (blue line) VO_2 RCRT samples, respectively. The samples with different ε_{LWIR} regulation abilities were fabricated by varying the thickness of PMMA spacer while fixing the thickness of VO_2-PMMA layer and the VO_2 weight ratio. (**B**) Mapping result on the effects of T_{sol} and ε_{LWIR} to the cooling (left) and heating (right) energy usage in climate zone 4. The minimum energy consumption for cooling happens at ε_{LWIR} = 0.9 and T_{sol} = 10%, and the minimum energy consumption for heating happens at the point with ε_{LWIR} = 0.1 and T_{sol} = 90%. This observation agrees well with the proposed ideal window in Fig. 1A. (**C**) Heating and cooling energy saving (blue columns) of simulated optimized dynamic samples against optimized static samples (red square) from zones 1 to 7. In this simulation, T_{sol} was fixed to 30%. The optimized static ε sample is derived from the point of minimum total energy consumption by adding heating and cooling in seven climate zones by fixing ε. The optimized dynamic ε window has a ε_{LWIR} of 0.9 for cooling and 0.1 for heating. The optimized dynamic ε window has lower energy consumption than optimized static ε window from zones 2 to 7. (**D**) Estimated world heating and cooling energy-saving of a W-doped max Δε sample against a commercial low-E glass as the baseline. RCRT windows outperform the commercial low-E across different climate zones. Unit for the energy-saving is MJ m^{-2}.

S3), has been instead applied in the full-scale building energy simulation. The optical response from 20° to 60°C with the interval of 10°C was used in the simulations (fig. S21). It should be noted that broadband cooler with high emssivity over the infrared spectra at 3-30 μm had a larger net cooling power than the 8- to 13-μm selective cooler when the surface temperature was above or close to the surrounding ambient temperature. (*33*) In the daytime, the exterior surface temperature of window glass is generally higher than the ambient temperature (fig. S22); therefore, we used the broadband emission in our calculations. We compared the energy-saving performance of the W-doped max Δε RCRT sample in different climate zones with a commercial low-E window and a conventional thermochromic window without RC regulation (fig. S23 and table S5). Our RCRT sample yielded a higher energy-saving performance benchmarked by a commercial low-E glass across all different climate zones (Fig. 3D), with energy saving up to 324.6 MJ m^{-2}, which further shows the potential importance of RC regulation and solar modulation in windows.

In summary, we fabricated a thermochromic smart window with tunable ε_{LWIR} based on a W-doped VO_2-PMMA/spacer/low-E stack using a solution process to cater to the dynamic energy demand in different climate zones. The ε_{LWIR} modulation ability of the windows can be tuned by simply adjusting the spacer thickness and VO_2 weight ratio and doping. In the whole-building energy simulations, the RCRT window showed a higher heating and cooling energy savings than a commercial low-E glass across the different climate zones. This type of technology has the potential to cut down the carbon emissions related to heating and cooling, improving the sustainability of buildings. Passive ε_{LWIR} and solar modulation has the potential to be useful for a wide range of heat-regulating applications including windows, walls, roofs, fabrics, and paintings.

REFERENCES AND NOTES

1. L. Zhao, X. Lee, R. B. Smith, K. Oleson, *Nature* **511**, 216–219 (2014).
2. E. Cuce, S. B. Riffat, *Renew. Sustain. Energy Rev.* **41**, 695–714 (2015).
3. Y. Gao *et al*., *Energy Environ. Sci.* **5**, 8234–8237 (2012).
4. E. S. Lee, X. Pang, S. Hoffmann, H. Goudey, A. Thanachareonkit, *Sol. Energy Mater. Sol. Cells* **116**, 14–26 (2013).
5. Y. Y. Cui *et al*., *Joule* **2**, 1707–1746 (2018).
6. Y. Ke *et al*., *Adv. Energy Mater.* **9**, 1902066 (2019).
7. S. Wang *et al*., *Appl. Energy* **211**, 200–217 (2018).
8. C. G. Granqvist, *Phys. Scr.* **32**, 401–407 (1985).
9. A. P. Raman, M. A. Anoma, L. Zhu, E. Rephaeli, S. Fan, *Nature* **515**, 540–544 (2014).
10. L. Zhu, A. P. Raman, S. Fan, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12282–12287 (2015).
11. Z. Chen, L. Zhu, A. Raman, S. Fan, *Nat. Commun.* **7**, 13729 (2016).
12. B. Bhatia *et al*., *Nat. Commun.* **9**, 5001 (2018).
13. J. Mandal *et al*., *Joule* **3**, 3088–3099 (2019).
14. L. Zhou *et al*., *Nat. Sustain.* **2**, 718–724 (2019).
15. Y. Zhai *et al*., *Science* **355**, 1062–1066 (2017).
16. D. Zhao *et al*., *Joule* **3**, 111–123 (2019).
17. E. A. Goldstein, A. P. Raman, S. Fan, *Nat. Energy* **2**, 17143 (2017).
18. Z. Chen, L. Zhu, W. Li, S. Fan, *Joule* **3**, 101–110 (2019).
19. Y. Peng *et al*., *Nat. Sustain.* **1**, 105–112 (2018).
20. T. Li *et al*., *Science* **364**, 760–763 (2019).
21. J. Mandal *et al*., *Science* **362**, 315–319 (2018).
22. H. Ye, X. Meng, B. Xu, *Energy Build.* **49**, 164–172 (2012).
23. H. Ye, X. Meng, L. Long, B. Xu, *Renew. Energy* **55**, 448–455 (2013).
24. M. Ono, K. Chen, W. Li, S. Fan, *Opt. Express* **26**, A777–A787 (2018).
25. C. Xu, M. Colorado Escobar, A. A. Gorodetsky, *Adv. Mater.* **32**, e1905717 (2020).
26. L. Xiao *et al*., *Nano Lett.* **15**, 8365–8370 (2015).
27. M. Li, D. Liu, H. Cheng, L. Peng, M. Zu, *Sci. Adv.* **6**, eaba3494 (2020).
28. Y. Ke *et al*., *Adv. Funct. Mater.* **28**, 1800113 (2018).
29. Y. Zhou *et al*., *Joule* **4**, 2458–2474 (2020).
30. M. A. Kats *et al*., *Appl. Phys. Lett.* **101**, 221101 (2012).
31. S. Taylor, Y. Yang, L. Wang, *J. Quant. Spectrosc. Radiat. Transf.* **197**, 76–83 (2017).
32. ANSI/ASHRAE/IES, "Energy standard for buildings except low-rise residential buildings" (ANSI/ASHRAE/IES, 2019); https://www.ashrae.org/technical-resources/bookstore/standard-90-1.
33. X. Yin, R. Yang, G. Tan, S. Fan, *Science* **370**, 786–791 (2020).

ACKNOWLEDGMENTS

Funding: The research was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) program and by the Singapore-HUJ Alliance for Research and Enterprise (SHARE), the Sino-Singapore International Joint Research Institute (SSIJRI), and Singapore Ministry of Education (MOE) Academic Research Fund Tier One RG103/19. **Author contributions:** Y.L. proposed, designed, and guided the project as principal investigator. S.W. fabricated the samples, performed the optical and thermal analyses, and drafted the manuscript with others. T.J. performed energy-saving performance simulation under the guidance of G.T. Y.M. performed theoretical simulation for the optical properties. Y.L., R.Y., and G.T. discussed and revised the manuscript. All authors checked the manuscript. **Competing interests:** A Singapore provisional patent (10202112639V) related to this work has been filed by Y.L. and S.W. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg0291
Materials and Methods
Figs. S1 to S27
Tables S1 to S6
References (*34*–*67*)

6 December 2020; accepted 26 October 2021
10.1126/science.abg0291

RADIATIVE COOLING

Temperature-adaptive radiative coating for all-season household thermal regulation

Kechao Tang[1,2,3]†, Kaichen Dong[1,2]†, Jiachen Li[2,4]†, Madeleine P. Gordon[4,5], Finnegan G. Reichertz[6], Hyungjin Kim[2,7], Yoonsoo Rho[8], Qingjun Wang[1,2], Chang-Yu Lin[1], Costas P. Grigoropoulos[8], Ali Javey[2,7], Jeffrey J. Urban[5], Jie Yao[1,2], Ronnen Levinson[9], Junqiao Wu[1,2,4]*

The sky is a natural heat sink that has been extensively used for passive radiative cooling of households. A lot of focus has been on maximizing the radiative cooling power of roof coating in the hot daytime using static, cooling-optimized material properties. However, the resultant overcooling in cold night or winter times exacerbates the heating cost, especially in climates where heating dominates energy consumption. We approached thermal regulation from an all-season perspective by developing a mechanically flexible coating that adapts its thermal emittance to different ambient temperatures. The fabricated temperature-adaptive radiative coating (TARC) optimally absorbs the solar energy and automatically switches thermal emittance from 0.20 for ambient temperatures lower than 15°C to 0.90 for temperatures above 30°C, driven by a photonically amplified metal-insulator transition. Simulations show that this system outperforms existing roof coatings for energy saving in most climates, especially those with substantial seasonal variations.

In countries such as the United States, ~39% of the total energy consumption is in buildings (*1*). For the residential housing energy portion, ~51% is consumed for heating and cooling to maintain a desirable indoor temperature (~22°C) (*2*). In contrast to most temperature regulation systems, which require external power input, the mid-infrared (IR) atmospheric transparency window ("sky window") allows thermal radiation exchange between terrestrial surfaces and the 3 K outer space, thus opening a passive avenue for thermal radiative cooling of buildings. This method to cool an outdoor surface such as a roof has been extensively studied in the past (*3*–*6*). It is now advanced by the development of daytime radiative cooling (*7*–*13*) using materials with low solar absorptance and high thermal emittance in the form of thin films (*8*), organic paints (*10*, *14*), or structural materials (*11*).

[1]Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Key Laboratory of Microelectronic Devices and Circuits (MOE), School of Integrated Circuits, Peking University, Beijing 100871, P. R. China. [4]Applied Science and Technology Graduate Group, University of California, Berkeley, CA, 94720, USA. [5]The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [6]East Bay Innovation Academy, 3800 Mountain Blvd., Oakland, CA 94619, USA. [7]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA. [8]Department of Mechanical Engineering, University of California, Berkeley, CA 94720, USA. [9]Heat Island Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
***Corresponding author. Email: wuj@berkeley.edu**
†These authors contributed equally to this work.

Past research on daytime radiative cooling, while successful in reducing cooling energy consumption, typically used materials with fixed, cooling-optimized properties, which efficiently emit thermal radiation even when the temperature of the surface is lower than desired, such as during the night or in the winter. This unwanted thermal radiative cooling will increase the energy consumption for heating and may offset the cooling energy saved in hot hours or seasons. This issue is well acknowledged by the research community, and mitigation of the overcooling has become a timely demand (*15*). To cut the heating penalty from overcooling, a few techniques were recently attempted for switching off thermal radiative cooling at low temperatures (below 22°C). Although effective in switching, these techniques typically require either additional energy input (*16*, *17*) or external activation (*18*), and in some cases, switching is achieved by mechanical moving parts (*19*, *20*). Developing dynamic structures that automatically cease radiative cooling at low temperatures is therefore highly desirable. Existing efforts in self-switching radiative cooling, however, are either purely theoretical (*21*–*24*) or limited to materials characterization with little relevance to practical household thermal regulation (*25*–*28*). Very recently, a smart subambient coating was developed (*29*), focusing on the reduction of solar absorption by fluorescence rather than modulation of thermal emittance by temperature.

We took a different, holistic approach by designing and fabricating a mechanically flexible coating structure to minimize total energy consumption through the entire year. This temperature-adaptive radiative coating (TARC) automatically switches its sky-window emittance to 0.90 from 0.20 when the surface temperature rises above ~22°C, a practical threshold not previously available. Our TARC delivers high radiative cooling power exclusively for the high-temperature condition (Fig. 1A). We also optimized the solar absorptance at ~0.25 (solar reflectance = 0.75) for all-season energy saving in major US cities (fig. S7). Our TARC demonstrates effective surface temperature modulation in an outdoor test environment. We performed extensive simulations based on the device properties and the climate database, which show advantages of TARC over existing roof coating materials in energy savings for most US cities in different climate zones (Fig. 1C). The energy savings by TARC not only bring economic benefits but also contribute to environmental preservation by reducing greenhouse gas emissions.

We developed the TARC based on the well-known metal-insulator transition (MIT) of the strongly correlated electron materials $W_xV_{1-x}O_2$ (*30*–*32*), and the transition temperature (T_{MIT}) is tailored to ~22°C by setting the composition x at 1.5% (*33*). We embedded a lithographically patterned two-dimensional array of thin $W_xV_{1-x}O_2$ blocks in a BaF_2 dielectric layer that sits on top of an Ag film (Fig. 2A). In the insulating (I) state of $W_xV_{1-x}O_2$ at $T < T_{MIT}$, the material is largely transparent to the infrared (IR) radiation in the 8- to 13-μm sky spectral window, so this sky-window IR radiation is reflected by the Ag mirror with little absorption (*34*). By contrast, the $W_xV_{1-x}O_2$ becomes highly absorptive in the sky window when it switches to the metallic (M) state at $T > T_{MIT}$ (*34*). The absorption is further amplified by the designed photonic resonance with adjacent $W_xV_{1-x}O_2$ blocks as well as with the bottom Ag layer through the ¼-wavelength cavity. The ¼-wavelength cavity structure induces Fabry-Perot resonance and was used in previous work to enhance thermal emission (*21*, *23*). According to Kirchhoff's law of radiation (*35*), the sky-window emittance equals the sky-window absorptance and switches from low to high when the temperature exceeds T_{MIT}. Consequently, strong sky-window radiative cooling is turned on in operation exclusively at high temperatures, leaving the system in the solar-heating or keep-warm mode at low temperatures. Details on the fabrication process and structural parameters are found in the supplementary materials (*36*) (fig. S1).

Our fabricated TARC has high flexibility for versatile surface adaption, as well as a microscale structure consistent with the design (Fig. 2B). We examined the emittance switching over the entire sample using a thermal infrared (TIR) camera (Fig. 2C). We imaged the TARC surface together with two reference samples having similar thicknesses but constant low thermal emittance (0.10, copper plate) or constant high thermal emittance (0.95, black tape), respectively. Although the thermal emission of the reference samples appeared to not be strongly temperature sensitive from 20 to 30°C, the TARC showed a marked change,

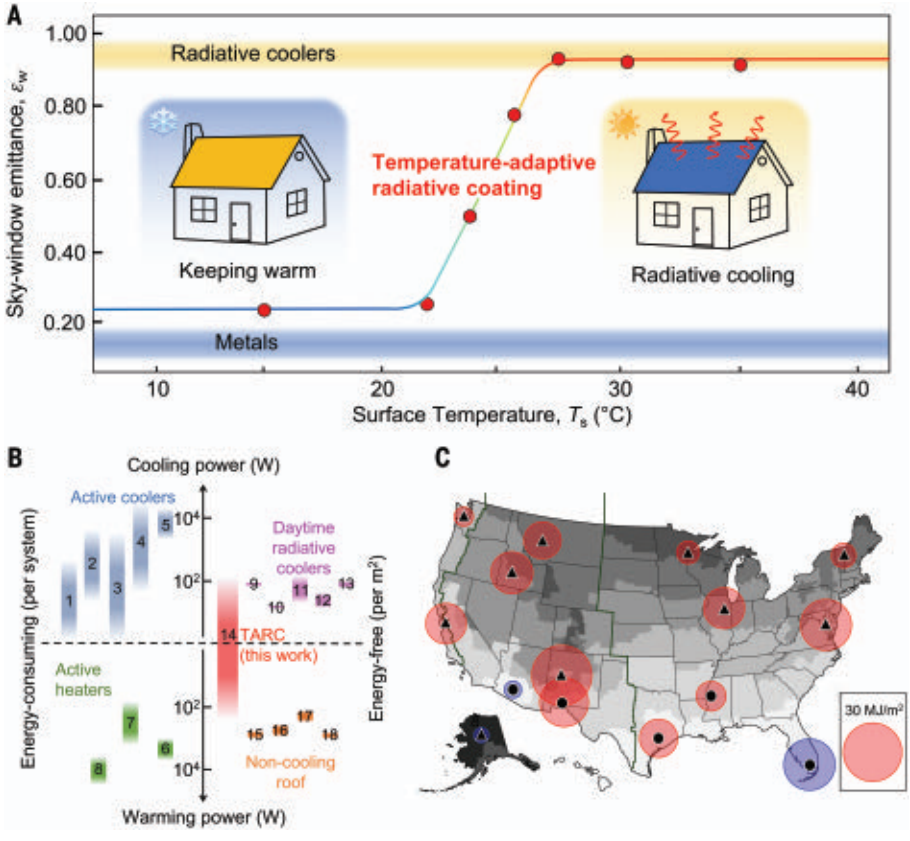


Fig. 1. TARC and its benefits for household thermal regulation. (**A**) Basic property of TARC in sky-window (8 to 13 μm) emittance modulation and schematics for temperature management when used as a household roof coating. The data points are the measured sky-window emittances of a TARC. The two color bands represent the temperature-independent thermal emittance of metals and radiative coolers. (**B**) TARC compared with other thermal regulation systems, highlighting the unique benefit of TARC of being simultaneously energy-free and temperature adaptive (details in table S1). (**C**) $SCSES_{min}$ of TARC compared with other existing roof-coating materials for different cities representing the 15 climate zones in the United States. Red and blue circles indicate positive and negative $SCSES_{min}$ values, respectively. The values are scaled to the area of the circles. Representation of the triangle and circle icons is explained in the materials and methods (subsection, "Projection of energy savings") (*36*).

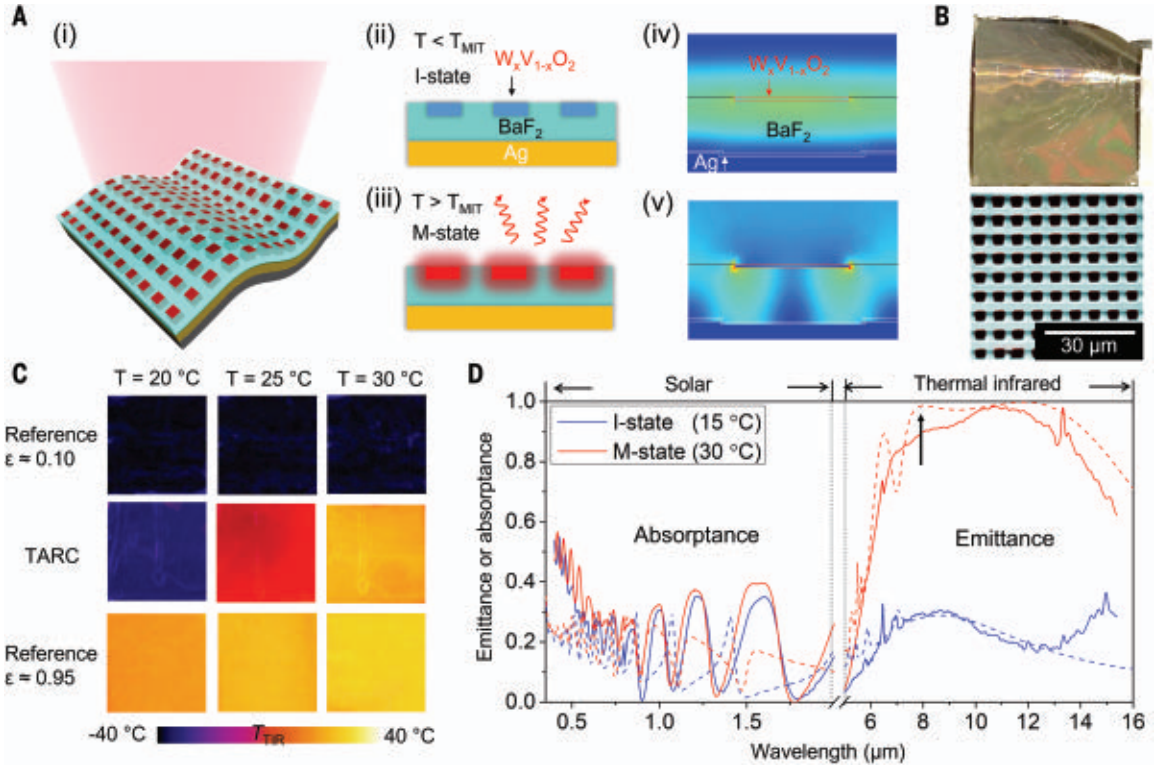


Fig. 2. Basic properties of TARC with experimental characterization. (**A**) Schematics of the structure (i), materials composition and working mechanism (ii and iii) of the TARC. Subpanels (iv) and (v) show the simulated distribution of electric field intensity below and above the transition temperature, respectively, when electromagnetic waves with a wavelength of 7.8 μm were normally incident on the TARC structure. (**B**) Photograph (2 cm × 2 cm) and false-color scanning electron microscope image of TARC showing high flexibility and structural consistency with the design. (**C**) TIR images of TARC compared with those of two conventional materials (references) with constantly low or high thermal emittance showing the temperature-adaptive switching in thermal emittance of TARC. (**D**) Solar spectral absorptance and part of the thermal spectral emittance of TARC at a low temperature and a high temperature, measured by a UV-visible-NIR spectrometer with an integrating sphere and an FTIR spectrometer, respectively. Measurements (solid curves) show consistency with theoretical predictions (dashed curves). The arrow at 7.8 μm denotes the wavelength where the distribution of electric field intensity shown in subpanels (iv) and (v) of (A) are simulated.

corresponding to the switch in sky-window emittance at the MIT around 22°C.

We measured the spectral properties of the TARC by a UV-visible-NIR spectrometer and Fourier transform infrared spectroscopy (FTIR) for the solar and TIR wavelength regimes, respectively (Fig. 2D). The solar absorptance (A, 0.3 to 2.5 μm) is ~0.25, and the sky-window emittance (ε_w, 8 to 13 μm) is ~0.20 in the I state and ~0.90 in the M state, consistent with theoretical simulations and other characterization results (fig. S2 and fig. S3).

The emittance switching of the TARC enables deep modulation of radiative cooling power in response to ambient temperature, which we first measured in vacuum (Fig. 3A). We suspended a heater membrane by thin strings in a vacuum chamber, which was cooled with dry ice to ~−78°C to minimize radiation from the chamber walls. We attached a piece of Al foil with $\varepsilon_{Al} \approx 0.03$ or a TARC of the same size to the top of the heater in two separate measurements. At each stabilized sample temperature T, the heating powers needed for the two coating scenarios are denoted as $P_{Al}(T)$ and $P_{TARC}(T)$, respectively. The cooling flux (power per area $\mathcal{A}$) contributed by the TARC was calculated as $P''_{cool}(T) = [P_{TARC}(T) - P_{Al}(T)]/\mathcal{A}$. We used the Al foil reference to calibrate background heat loss from thermal conduction through the strings. We plotted the calibrated cooling power (Fig. 3B), which shows an abrupt increase in $P''_{cool}(T)$ when T rises above the MIT temperature. $P''_{cool}(T)$ measurements in the I state and M state are well fitted by the Stefan-Boltzmann radiation law, with values of sky-window ε_w extracted to be ~0.20 and ~0.90, respectively, consistent with the spectrally characterized results (Fig. 2D). We considered and corrected the effect of radiation from the chamber wall (~−78°C) for the calibration. We introduced a constant factor of γ (≈ 0.7) to account for the difference between the vacuum and ambient measurement conditions (details in fig. S4) (*36*).

We demonstrated the actual outdoor performance of the TARC (Fig. 4). We recorded the surface temperatures (T_s) of the TARC, together with a dark roof coating product (Behr no. N520, asphalt gray) and a cool (white) roof coating product (GAF RoofShield white acrylic), over 24 hours on a sunny summer day on a rooftop in Berkeley, California, with a careful design of the measurement system to minimize the effects of artifacts (fig. S5).

From 00:00 to 09:00 local daylight time (LDT), when the ambient temperature was below T_{MIT}, the TARC was 2°C warmer than the two reference roof coatings, arising from the low sky-window emittance ($\varepsilon_w = 0.20$) of the TARC in the I state and thus a lower radiative cooling power than the references ($\varepsilon_w = 0.90$). The 2°C temperature elevation is consistent with adiabatic simulation results based on these nominal emittance values and the local weather database [see the supplementary materials (*36*), note A, section I]. From 09:00 to 13:00 LDT, when the samples were in direct sunlight, T_s was dominated by the solar absorption in balance with radiative cooling and air convection, and the differences between the samples agree with the simulated results assuming the solar absorptance A to be 0.15, 0.25, and 0.70 for the white roof coating,

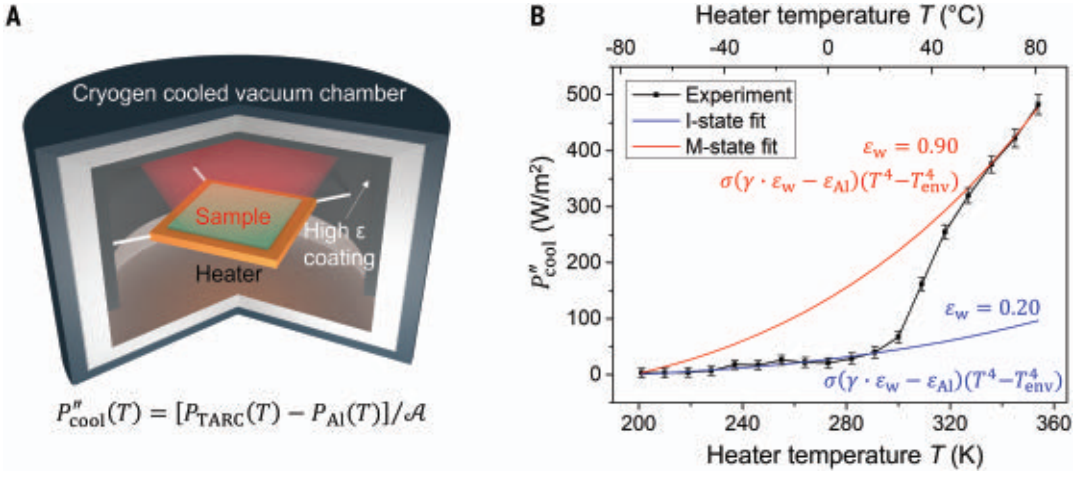


Fig. 3. Characterization of intrinsic radiative cooling power of TARC in a cryogenic vacuum chamber. (**A**) Schematics of the experimental setup showing a thin heater membrane covered by either a TARC or an Al foil and suspended in a cryogen-cooled vacuum chamber. The Al foil reference is used to cancel the effect of thermal loss through conduction. (**B**) Calibrated experimental cooling flux (power/area) of TARC as a function of temperature in vacuum (the black data line). Fitting of $P''_{cool}(T)$ at I and M states by the Stefan-Boltzmann radiation law gives ε_w values of 0.20 and 0.90, respectively.

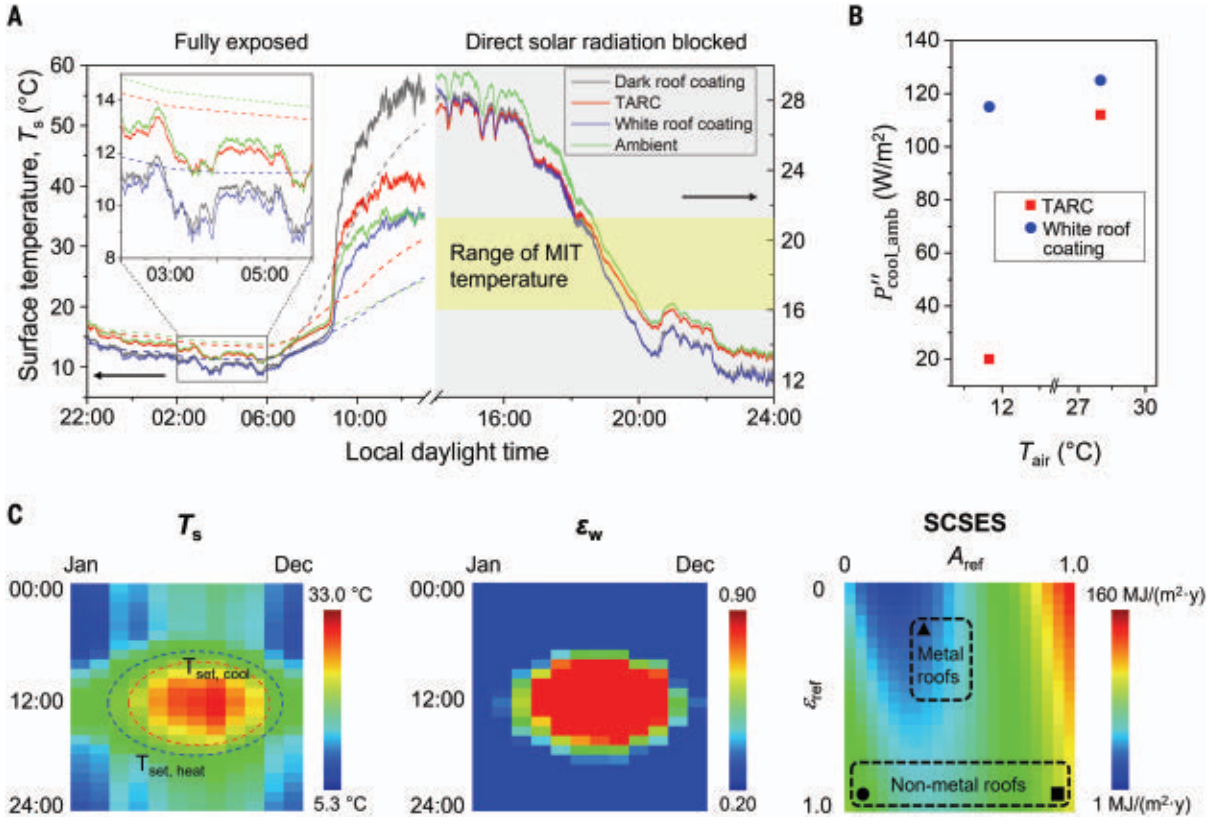


Fig. 4. Characterization of TARC in an outdoor environment. (**A**) Surface temperature of TARC, a commercial dark roof coating (A = 0.70, ε_w = 0.90), and a commercial white roof coating (A = 0.15, ε_w = 0.90) in an open-space outdoor environment recorded over a day-night cycle. The measurement was taken on 5 July 2020, in Berkeley, California (37.91°N, 122.28°W). The solid and dashed curves are experimental data and simulation results based on a local weather database (*37*), respectively. Measurements starting from 14:00 LDT were performed with the direct solar radiation blocked. Temperature observed after sunset show clear signs of the TARC shutting off thermal radiative cooling as its surface ambient temperature falls below T_{MIT}. (**B**) Measured ambient cooling power of TARC and white roof coating with direct solar radiation blocked in the outdoor environment. (**C**) T_s and the corresponding ε_w mapping of TARC over 24 hours and the full year for Berkeley. Also shown are the SCSES of TARC compared with all other materials with fixed solar absorptance (A_{ref}) and fixed thermal emittance (ε_{ref}). The icons in the SCSES map correspond to those used in Fig. 1C, denoting the radiative parameters (A, ε_w) of the strongest rival to TARC in source energy savings for the local climate (*36*).

TARC, and the dark roof coating, respectively. After 13:00 LDT, we erected a shield to intentionally block direct solar radiation to the surface of the samples. This imitates the scenario of a cloud blocking the sun but with the rest of the sky mostly clear. We quickly observed a convergence of the T_s curves for all three samples, an indication that the thermal emittance of the TARC in the M state is close to that of the two references (0.90). This condition persisted for a few hours until T_s started to drop below T_{MIT} = 22°C. After this point, TARC grew warmer than the two references, with a final temperature difference of ~2°C, similar to the 00:00 to 09:00 LDT period. This indicates

that the TARC switched to the low-emittance I state. The 24-hour outdoor experiments demonstrate the emittance switching and resultant temperature regulation by TARC. Although the white roof coating shows an advantage over TARC in thermal management in summer daytime and under solar radiation (Fig. 4A), the TARC regulates the roof temperature closer to the heating and cooling setpoints (22 and 24°C) than the white roof coating for almost all of the other conditions, including daytime in other seasons and all of the nighttime (fig. S6). From an all-year-round perspective, the TARC demonstrates superiority compared with regular roof coatings in terms of source energy saving.

To directly compare their ambient condition cooling fluxes (P''_{cool_amb}), we heated the TARC and the white roof coating to the air temperature with the direct solar radiation blocked. P''_{cool_amb} refers to the net cooling flux from the surface—namely, the thermal radiative heat loss flux minus the absorbed diffuse solar irradiance. We plotted the P''_{cool_amb} values that we obtained at a low and a high air temperature (Fig. 4B). The TARC exhibits a clear switching of P''_{cool_amb} by a factor over five across the MIT. This behavior is in stark contrast to the nearly constant P''_{cool_amb} around 120 W/m^2 for the shaded white roof coating, which is consistent with values (90 to 130 W/m^2) reported in literature for roofs surfaced with daytime radiative cooling materials (*5*, *9*, *10*).

We performed extensive numerical simulations to analyze the performance of TARC in household energy saving for the US cities from an all-season perspective (*36*). We show the simulated results (Fig. 4C) for Berkeley where the measurements (Fig. 4, A and B) were performed. We calculated an hour-month map of T_s using a local weather file (*37*), laying the basis for estimation of energy saving. We assumed heating and cooling setpoints $T_{set,heat}$ = 22°C and $T_{set,cool}$ = 24°C (*38*), and approximated that the building will need heating when $T_s < T_{set,heat}$ and require cooling when $T_s > T_{set,cool}$. We used past simulations of cool-roof energy savings to predict potential space-conditioning source energy savings (SCSES) per unit roof area attainable by using TARC in place of roofing materials that have static values of solar absorptance and thermal emittance (*36*). The figure of merit of TARC is represented by $SCSES_{min}$, the minimum value of SCSES found over all existing conventional roofing materials, which have constant values of A_{ref} and ε_{ref} (Fig. 4C, dashed boxes). We mapped $SCSES_{min}$ for cities representing the 15 US climate zones (Fig. 1C). This figure-of-merit map shows that TARC provides clear, positive annual space-conditioning source energy savings relative to existing roof coating materials in most major cities, except for climates that are constantly cold (such as Fairbanks) or hot (such as Miami) throughout the year. It highlights the advantage of TARC, especially in climate zones with wide temperature variations, day to night or summer to winter. For example, we estimate that for a single-family home in Baltimore, Maryland, built before 1980, modeled with roof assembly thermal insulance 4.3 m^2/(K·W), gas furnace annual fuel utilization efficiency 80%, and air conditioner coefficient of performance 2.64 (*38*), $SCSES_{min}$ is 22.4 MJ/(m^2·y), saving 2.64 GJ/y based on a roof area of 118 m^2. We also calculated the source energy saving of TARC as a function of its solar absorptance (fig. S7), showing that the actual solar absorptance of TARC is close to the optimal value for major US cities.

The TARC could be readily upgraded for heavy-duty outdoor applications by coating it with a thin polyethylene (PE) membrane, which is nontoxic, hydrophobic, and transparent both in the visible and thermal IR regions. While protecting the TARC from contacting the dust and moisture in complex environments, the PE coating has little impact on the thermal modulation performance (fig. S9). Polymer imprinting instead of photolithography could also be used to more easily produce the material for large scale application. By embedding VO_2 particles in layered PE membranes, we estimated the multilayered metamaterial to achieve comparable modulation performance ($\Delta\varepsilon_w > 0.8$) as the TARC we presented and would be producible in a roll-to-roll fashion (figs. S10 and S11). Roll-to-roll manufacturing of PE-based TARC would be beneficial because of its high scalability, low cost (*9*), and the fact that it is free from the liquid evaporation process in fabrication (*39*). The PE layer can be also replaced by other organic or inorganic materials with negligible optical loss in the wavelength ranges of both solar irradiation and IR atmospheric transparency window, so that the TARC technology can be designed specifically to be endurable in different environmental conditions.

We developed a mechanically flexible, energy-free TARC for intelligent regulation of household temperature. Our system features a thermally driven metal-insulator transition in cooperation with photonic resonance, and demonstrates self-switching in sky-window thermal emittance from 0.20 to 0.90 at a desired temperature of ~22°C. These attractive properties enable switching of the system from the radiative cooling mode at high temperatures to the solar-heating or keep-warm mode at low temperatures in an outdoor setting. For most cities in the United States, our simulations indicate the TARC may outperform all conventional roof materials in terms of cutting energy consumption for households.

REFERENCES AND NOTES

1. US Energy Information Administration, "Annual Energy Review 2020" (2020); https://www.eia.gov/totalenergy/data/annual/
2. US Energy Information Administration, "2015 Residential Energy Consumption Survey" (2015); https://www.eia.gov/consumption/residential/
3. A. R. Gentle, G. B. Smith, *Adv. Sci. (Weinh.)* **2**, 1500119 (2015).
4. M. Dong, N. Chen, X. Zhao, S. Fan, Z. Chen, *Opt. Express* **27**, 31587–31598 (2019).
5. B. Orel, M. K. Gunde, A. Krainer, *Sol. Energy* **50**, 477–482 (1993).
6. P. Berdahl, M. Martin, F. Sakkal, *Int. J. Heat Mass Transf.* **26**, 871–880 (1983).
7. N. N. Shi *et al.*, *Science* **349**, 298–301 (2015).
8. J. Kou, Z. Jurado, Z. Chen, S. Fan, A. J. Minnich, *ACS Photonics* **4**, 626–630 (2017).
9. Y. Zhai *et al.*, *Science* **355**, 1062–1066 (2017).
10. J. Mandal *et al.*, *Science* **362**, 315–319 (2018).
11. T. Li *et al.*, *Science* **364**, 760–763 (2019).
12. Z. Li, Q. Chen, Y. Song, B. Zhu, J. Zhu, *Adv. Mater. Technol.* **5**, 1901007 (2020).
13. P. Berdahl, *Appl. Opt.* **23**, 370–372 (1984).
14. X. Li *et al.*, *Cell Rep. Phys. Sci.* **1**, 100221 (2020).
15. G. Ulpiani, G. Ranzi, K. W. Shah, J. Feng, M. Santamouris, *Sol. Energy* **209**, 278–301 (2020).
16. K. Goncharov *et al.*, "1500 W deployable radiator with loop heat pipe" (SAE Technical Paper 2001-01-2194, 2001); https://doi.org/10.4271/2001-01-2194.
17. C. Lashley, S. Krein, P. Barcomb, "Deployable radiators-A multi-discipline approach" (SAE Technical Paper 981691, 1998); https://doi.org/10.4271/981691.
18. H. Zhao, Q. Sun, J. Zhou, X. Deng, J. Cui, *Adv. Mater.* **32**, e2000870 (2020).
19. Z. Xia, Z. Fang, Z. Zhang, K. Shi, Z. Meng, *ACS Appl. Mater. Interfaces* **12**, 27241–27248 (2020).
20. H. Nagano, Y. Nagasaka, A. Ohnishi, *J. Thermophys. Heat Trans.* **20**, 856–864 (2006).
21. M. Ono, K. Chen, W. Li, S. Fan, *Opt. Express* **26**, A777–A787 (2018).
22. M. Chen, A. M. Morsy, M. L. Povinelli, *Opt. Express* **27**, 21787–21793 (2019).
23. S. Taylor, Y. Yang, L. Wang, *J. Quant. Spectrosc. Radiat. Transf.* **197**, 76–83 (2017).
24. W.-W. Zhang, H. Qi, A.-T. Sun, Y.-T. Ren, J.-W. Shi, *Opt. Express* **28**, 20609–20623 (2020).
25. K. Ito, T. Watari, K. Nishikawa, H. Yoshimoto, H. Iizuka, *APL Photonics* **3**, 086101 (2018).
26. K. Nishikawa, K. Yatsugi, Y. Kishida, K. Ito, *Appl. Phys. Lett.* **114**, 211104 (2019).
27. H. Kim *et al.*, *Sci. Rep.* **9**, 11329 (2019).
28. K. Sun *et al.*, *ACS Photonics* **5**, 2280–2286 (2018).
29. X. Xue *et al.*, *Adv. Mater.* **32**, e1906751 (2020).
30. F. J. Morin, *Phys. Rev. Lett.* **3**, 34–36 (1959).
31. T.-L. Wu, L. Whittaker, S. Banerjee, G. Sambandamurthy, *Phys. Rev. B Condens. Matter Mater. Phys.* **83**, 073101 (2011).
32. C. Kim, J. S. Shin, H. Ozaki, *J. Phys. Condens. Matter* **19**, 096007 (2007).
33. S. Lee *et al.*, *Science* **355**, 371–374 (2017).
34. A. S. Barker, H. W. Verleur, H. J. Guggenheim, *Phys. Rev. Lett.* **17**, 1286–1289 (1966).
35. J. Agassi, *Science* **156**, 30–37 (1967).
36. The materials and methods are available as supplementary materials.
37. US Department of Energy, "EnergyPlus Weather Data" (2021); https://energyplus.net/weather.
38. P. J. Rosado, R. Levinson, *Energy Build.* **199**, 588–607 (2019).
39. A. Leroy *et al.*, *Sci. Adv.* **5**, eaat9480 (2019).

ACKNOWLEDGMENTS

Funding: This work was funded by the Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, US Department of Energy, under contract no. DE-AC02-05-CH11231 (EMAT program KC1201). Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, US Department of Energy, under contract no. DE-AC02-05CH11231. J.W. acknowledges support from a Bakar Prize. R.L. acknowledges support from the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the US Department of Energy under contract no. DE-AC02-05CH11231. J.Y. acknowledges support from the National Science Foundation under grant no. 1555336. M.P.G. gratefully acknowledges the National Science Foundation for fellowship support under the National Science Foundation Graduate Research Fellowship Program. **Author contributions:** K.T., K.D. J.L., and J.W. conceived the general idea. K.T. and K.D. designed the device. K.T. fabricated the device. K.T., M.P.G., H.K., Q.W., A.J., J.J.U., and J.Y. contributed to the spectral characterizations. K.T., K.D., J.L. Y.R., and C.P.G. contributed to the solar simulator characterizations. K.T., J.L., and C.-Y.L. performed the vacuum chamber characterizations. K.T., K.D., J.L., and J.W. performed the field experiments. K.D., J.L. and J.Y. performed

the numerical electromagnetic simulations. K.T., F.G.R., and R.L. performed all other simulations. All authors discussed and analyzed the results. K.T., K.D., J.L., and J.W. wrote the manuscript with assistance from other authors. All authors reviewed and revised the manuscript.
Competing interests: R.L. is an unpaid, nonvoting member of the board of directors of the Cool Roof Rating Council (CRRC) and a paid consultant to the CRRC. K.T., K.D., J.L., and J.W. are inventors of a provisional patent application related to this work. The authors declare that they have no competing interests. **Data and materials availability:** All data required to evaluate the conclusions in the manuscript are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf7136
Nomenclature
Materials and Methods
Supplementary Text
Figs. S1 to S19
Tables S1 to S6
References (*40*–*105*)

13 November 2020; resubmitted 5 May 2021
Accepted 26 October 2021
10.1126/science.abf7136

BIOTECHNOLOGY

Multiple rereads of single proteins at single–amino acid resolution using nanopores

Henry Brinkerhoff[1], Albert S. W. Kang[1], Jingqian Liu[2], Aleksei Aksimentiev[2], Cees Dekker[1]*

A proteomics tool capable of identifying single proteins would be important for cell biology research and applications. Here, we demonstrate a nanopore-based single-molecule peptide reader sensitive to single–amino acid substitutions within individual peptides. A DNA-peptide conjugate was pulled through the biological nanopore MspA by the DNA helicase Hel308. Reading the ion current signal through the nanopore enabled discrimination of single–amino acid substitutions in single reads. Molecular dynamics simulations showed these signals to result from size exclusion and pore binding. We also demonstrate the capability to "rewind" peptide reads, obtaining numerous independent reads of the same molecule, yielding an error rate of $<10^{-6}$ in single amino acid variant identification. These proof-of-concept experiments constitute a promising basis for the development of a single-molecule protein fingerprinting and analysis technology.

Genetic sequence is a key source of information about protein primary sequence. However, because they do not directly encode information about protein abundance or about posttranslational modification and splicing of proteins, neither the DNA genome nor the RNA transcriptome fully describe the protein phenotype. A robust method for directly identifying proteins and detecting posttranslational modifications at the single-molecule level would greatly benefit proteomics research (*1*), enabling quantification of low-abundance proteins as well as distributions and correlations of posttranslational modifications, all at a single-cell level. Here, we provide proof-of-concept data for a nanopore-based approach that can discriminate single peptides at single–amino acid sensitivity with high fidelity and potential for high throughput. Although it is not presently capable of de novo protein sequencing, this nanopore peptide reader provides site-specific information about the peptide's primary sequence that may find applications in single-molecule protein fingerprinting and variant identification.

Recently, biological nanopores have been used as the basis of a single-molecule DNA sequencing technology (*2*) that is capable of long reads and detection of epigenetic markers in a portable platform with minimal cost (*3*). In such experiments, single-stranded DNA (ssDNA) is slowly moved step by step through a protein nanopore embedded in a thin membrane, partially blocking an electrical current carried by ions through the nanopore. The DNA stepping is accomplished using a DNA-translocating motor enzyme that moves DNA through the pore in discrete steps, yielding a series of steps in the ion current. Each ion current level characterizes the bases residing in the pore at that step, and the sequence of levels can be decoded into the DNA base sequence.

It has been hypothesized that nanopores can also be used for protein fingerprinting or sequencing (*4*, *5*). Methods in which small peptide fragments freely translocate through a pore have shown sensitivity to single amino acids (*6*–*8*), but we lack a method for determining the order of amino acids and reconstructing the sequence of single proteins. Using a ClpX protein unfoldase to pull a peptide through a nanopore yielded signals that effectively distinguished between different peptides (*9*), but these reads were difficult to interpret, in part because of the irregular stepping behavior of ClpX (*10*). In our study, we instead applied the precise stepwise control of a DNA-translocating motor (*11*–*13*) to pull a peptide through a nanopore, similarly to simultaneous work by Yan *et al.* (*14*) but presenting several key advances: the use of a helicase that pulls the polymer through MspA in smaller, half-nucleotide steps; the ability to identify single amino acid substitutions; and the capability to obtain high-fidelity signals by rereading the same single molecule multiple times.

We developed a system in which a DNA-peptide conjugate was pulled through a biological nanopore by a helicase that was walking on the DNA section (Fig. 1). The conjugate strand consisted of an 80-nucleotide DNA strand that was covalently linked to a 26–amino acid synthetic peptide by a DBCO click linker on the 5′ end of the DNA connecting to an azide modification at the C terminus of the peptide (materials and methods section 1 and fig. S1). A negatively charged peptide sequence of mostly aspartic acid (D) and glutamic acid (E) residues was chosen so that the electrophoretic force assisted in pulling the peptide into the pore. We used the mutant nanopore M2 MspA (*15*) with a cuplike shape that separates the helicase by ~10 nm from the constriction of the pore where the blockage of ion current occurs (*16*). For the DNA-translocating motor enzyme, we used Hel308 DNA helicase because (i) it pulls ssDNA through MspA in half-nucleotide ~0.33-nm observable steps (*13*), which are close to single–amino acid steps; (ii) because it is a stable and processive helicase that tolerates high salt concentrations (*16*); and (iii) its >50 pN pulling force (*16*) is likely to denature any secondary structure in target peptides.

We found that, similar to nanopore reads of DNA, ratcheting a peptide through the nanopore generated a distinct steplike pattern in the ion current (Fig. 1D). Durations of ion current steps varied from read to read, but the sequence of levels was highly reproducible (fig. S2). The progression of ion current steps was accurately identified using custom software (materials and methods section 2 and fig. S3), and further analysis was performed on the sequence of the median ion current values for each step (Fig. 1E).

This sequence of ion current levels first closely tracked the sequence expected for the template strand of DNA, which can be predicted using a DNA sequence–to–ion current map developed previously (*17*, *18*) (materials and methods section 3). After the end of the DNA crossed to the cis side of MspA's constriction, we continued to observe stepping over the linker (a length of ~2 nm, or six Hel308 steps), and subsequently over the peptide. The stepping of the peptide through the MspA

[1]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629 HZ Delft, Netherlands. [2]Center for Biophysics and Quantitative Biology and Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
***Corresponding author. Email: c.dekker@tudelft.nl**

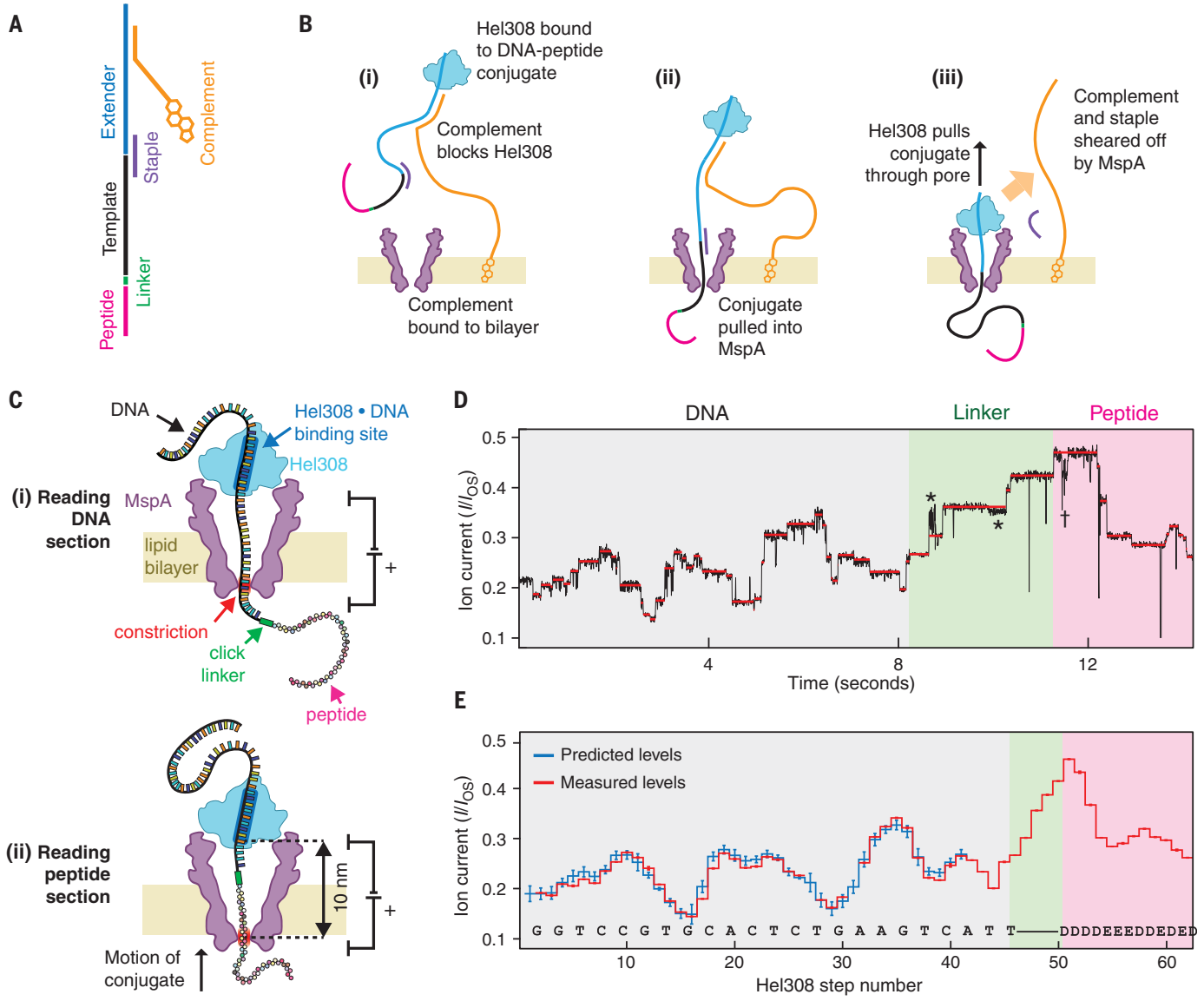


Fig. 1. Reading peptides with a nanopore. (**A**) The DNA-peptide conjugate consists of a peptide (pink) attached via a click linker (green) to an ssDNA strand (black). This DNA-peptide conjugate is extended with a typical nanopore adaptor comprised of an extender that acts as a site for helicase loading (blue) and a complementary oligo with a 3′ cholesterol modification (gold). (**B**) The cholesterol associates with the bilayer as shown in (i), increasing the concentration of analyte near the pore. The complementary oligo blocks the helicase, until it is pulled into the pore (ii), causing the complementary strand to be sheared off (iii), whereupon the helicase starts to step along DNA. (**C**) As the helicase walks along the DNA, it pulls it up through the pore, resulting in (i) a read of the DNA portion followed by (ii) a read of the attached peptide. (**D**) Typical nanopore read of a DNA-peptide conjugate (black), displaying steplike ion currents (identified in red). The asterisks indicate a spurious level not observed in most reads and therefore omitted from further analysis. The dagger symbol indicates a helicase backstep. The ion current is displayed as a fraction of the open pore current I_{OS}. (**E**) Consensus sequence of ion current steps (red), which for the DNA section is closely matched by the predicted DNA sequence (blue). The linker and peptide sections are identified by counting half-nucleotide steps over the known structural length of the linker. Error bars in the measured ion current levels are errors in the mean value, often too small to see. Error bars in the prediction are standard deviations of the ion current levels that were used to build the predictive map in previous work (*18*).

constriction produced distinguishable ion current steps, much like those from DNA, but with a higher average ion current. Because individual reads might contain a varying number of steps owing to helicase backstepping and errors in step segmentation, we identified these features by cross-comparison of several independent reads, producing a "consensus" ion current sequence free of helicase missteps or step-segmentation errors (materials and methods section 4). By counting the steps in these consensus sequence traces, we determined the parts of the traces that corresponded to the linker (the first six steps after the DNA) and the peptide (all steps thereafter) in the MspA constriction. We confirmed this analysis by altering the peptide sequence at a selected site and observing the location of the resulting change in the ion current stepping sequence, as discussed below. We restricted further analysis to reads containing both DNA and peptide sections (materials and methods section 5 and fig. S4).

Our approach allowed us to discriminate peptide variants that differed by only a single amino acid. We obtained reads (N = 211) of three different DNA-peptide conjugates in 19 different pores, where the peptide sequences consisted of a mixture of negatively charged D and E residues, with a single variation—that is,

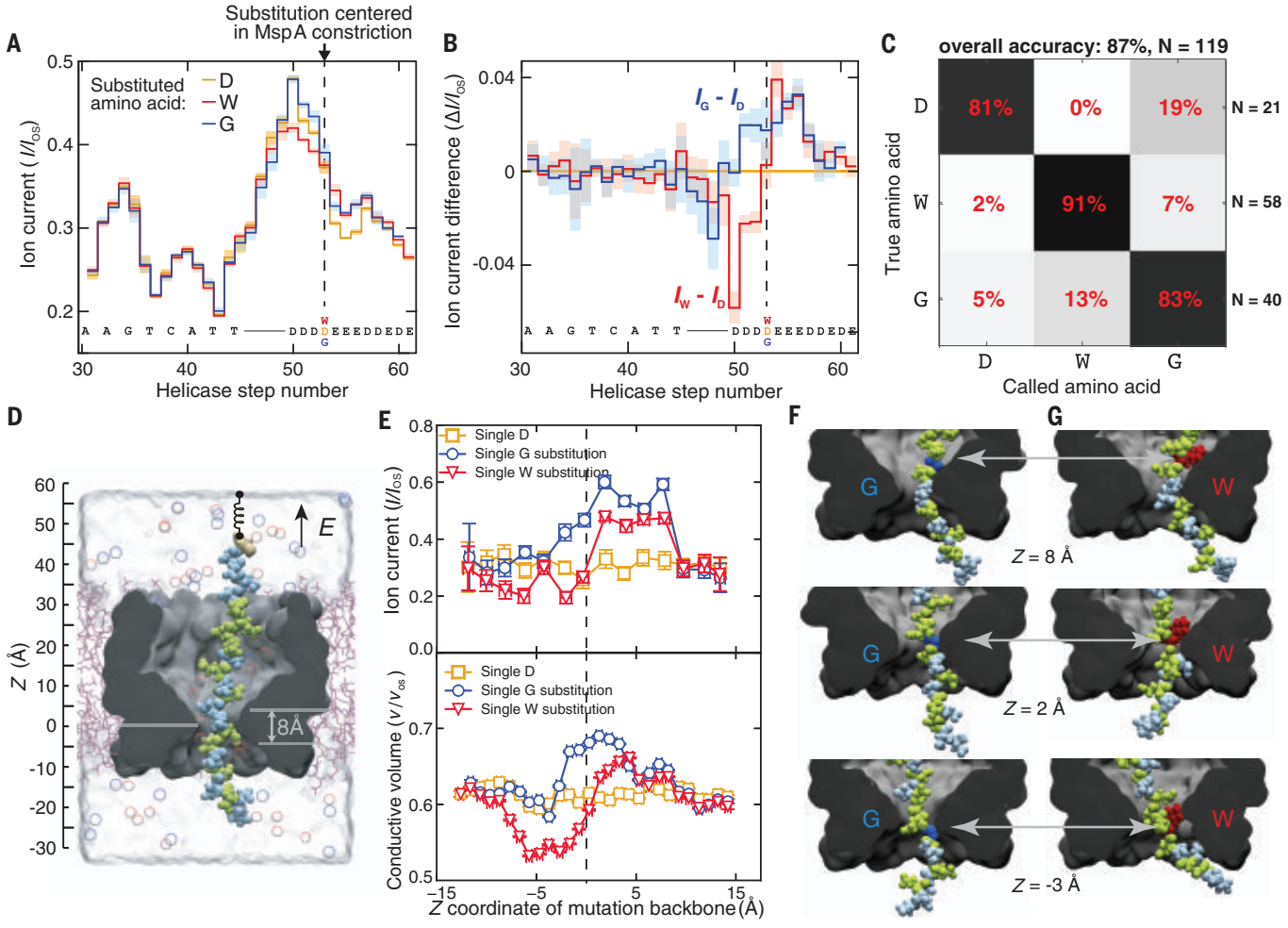


Fig. 2. Detection of single amino acid substitutions in single peptides. (**A**) Consensus ion current sequences for each of the three measured variants (D, gold; W, red; G, blue), which differ significantly at the site of the amino acid substitution. (**B**) Difference in ion current between the W (red) and G (blue) variants and the D variant. Error bars are standard deviations. (**C**) Confusion matrix showing error modes of a blind classifier in identifying variants of reads, demonstrating an 87% single-read accuracy. (**D**) All-atom model where a reduced-length MspA pore (gray) confines a polypeptide chain (Glu, green; Asp, light blue; Cys, beige). The top end of the peptide is anchored using a harmonic spring potential, representing the action of the helicase at the rim of a full-length MspA. Water and ions are shown as semitransparent surface and spheres, respectively. (**E**) (Top) Ionic current in MspA constriction versus *z* coordinate of the mutated residue backbone from MD simulations. (Bottom) Fraction of nanopore construction volume available for ion transport. Vertical and horizontal error bars denote standard errors and standard deviations, respectively. (**F** and **G**) Representative molecular configurations observed in MD simulations of peptide variants. Glycine and tryptophan residues are shown in dark blue and red, respectively. Considerable peptide–pore surface interactions are observed.

D, glycine (G), or tryptophan (W)—placed four amino acids away from the C terminus that connected to the linker (see table S1 for full sequences). The three variants showed a reproducible difference at the site of the substituted amino acid, which could be seen by comparing the consensus sequences of ion current levels (Fig. 2, A and B). As is typical of nanopore experiments, a single-site variation was found to affect several ion current steps, because an "8-mer" of amino acids around the pore constriction of MspA affects the ion current blockage level (*11*, *17*) owing to the finite constriction height and stochastic displacements of the strand up and down through the nanopore (*19*). The center of the differing region in the ion current sequence was at the expected site: ~10 helicase steps away from the end of the DNA section (six half-nucleotide steps for the linker and four more along the peptide to the variant site). The signals varied by several standard deviations over multiple sequential levels, demonstrating that variations as small as a single amino acid substitution could be resolved. The differences of the ion currents for the W- and G-substituted variants from the D-substituted variant (Fig. 2B) showed a notable behavior: when G, which has just a hydrogen atom as a side chain, occupied the nanopore constriction, we saw higher ion current levels, as expected from a smaller amino acid volume. But when the bulky W variant moved through the constriction, the ion current first decreased and then, counterintuitively, increased relative to the medium-sized D variant.

To understand the origin of these patterns, we performed all-atom molecular dynamics (MD) simulations measuring the ion current with peptide variants at varying positions within the MspA constriction. In a typical simulation, a polypeptide chain was threaded through a reduced-length model of the MspA nanopore embedded in a lipid bilayer and surrounded by 0.4 M KCl electrolyte (Fig. 2D). Peptides with either a W or G substitution in a mixed D/E sequence were examined under a +200 mV bias at various locations relative to the MspA constriction (see materials and methods section 6 and figs. S5 to S8 for details). Patterns of ionic current blockades resulted in Fig. 2E (top panel), matching the counterintuitive blockade current patterns that were experimentally measured for G and W substitutions

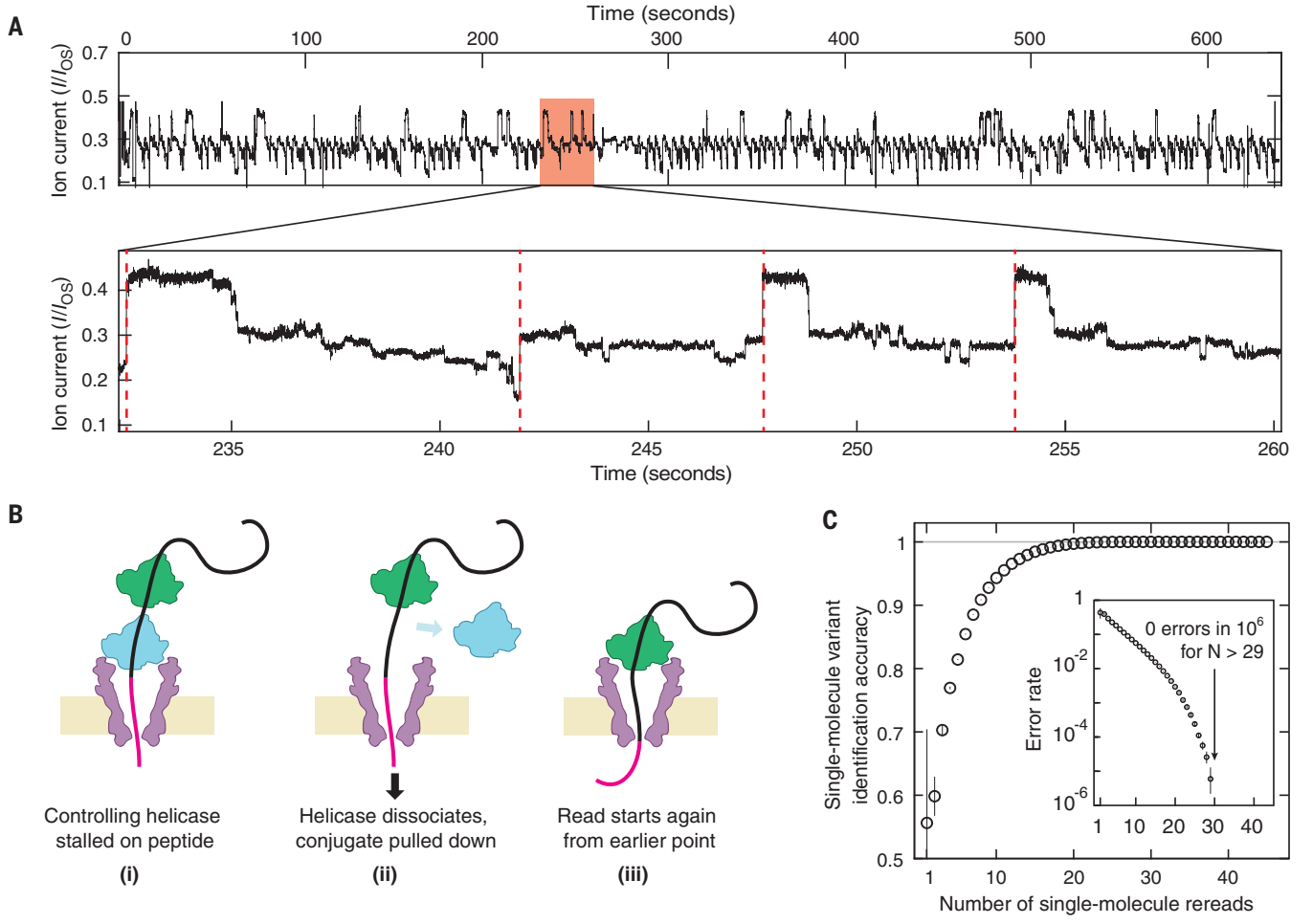


Fig. 3. Rereading of a single peptide. (**A**) Highly repetitive ion current signal corresponding to numerous rereads of the same section of an individual peptide (in this case, the G-substituted variant). The expanded plot (bottom) shows a region that contains four rewinding events (red dashed lines), where the trace jumps back to level 52 ± 2 of the consensus displayed in Fig. 2A. (**B**) Rereading is facilitated by helicase queueing, where (i) a second helicase binds behind the primary helicase that controls the DNA-peptide conjugate, rereading starts when (ii) the primary helicase dissociates, and (iii) the secondary one becomes the primary helicase that drives a new round of reading. (**C**) By using information from multiple rereads of the same peptide, the identification accuracy can be raised to very high levels of fidelity. These results indicate that with sufficient numbers of rereads, random error can be eliminated and single-molecule error rate can be pushed lower than 1 in 10^6 even with poor single-pass accuracy. Inset is a logarithmic plot of the error rate = 1 − accuracy.

(compare Fig. 2, A and B). Furthermore, the ion current correlated with the nanopore constriction volume that was available for ion transport near the pore mouth (Fig. 2E, bottom panel), with the latter quantity being more accurately characterized by the all-atom MD method (*19*). In the case of a G residue, its upward motion was accompanied by an increase of the nanopore volume (Fig. 2E, bottom), which subsided as the residue left the nanopore constriction (Fig. 2F), in sync with the blockade current (Fig. 2E, top). A W residue, however, reduced the nanopore constriction volume when it was located below the constriction (Fig. 2E, top) but increased the volume at and above the constriction. The latter counterintuitive effect could be traced back to a binding of the W side chain to the nanopore surface above the constriction (Fig. 2G). Thus, a glycine substitution merely increases the nanopore volume as the residue passes through the constriction, whereas the tryptophan residue decreases the volume when its side chain enters the constriction and subsequently increases the volume when its side chain binds to the inner nanopore surface (fig. S9).

To quantitatively assess the distinguishability of peptide variants, we computed a so-called confusion matrix (Fig. 2C). Using a hidden Markov model, we quantified the relative likelihoods of the alignments to the three consensus sequences for 119 reads withheld from the consensus sequence generation, finding that we could identify the correct variant with an average accuracy of 87% (materials and methods section 7). This high rate of correct single substitution identification compares favorably to early nanopore experiments, which identified single-nucleotide variants with considerably lower accuracy (*17*). Still, the limited single-read accuracy is an ongoing challenge in developing nanopore sequence analysis approaches, requiring the implementation of strategies to increase sequencing fidelity to acceptable levels (*18*, *20*). The largest error modes in nanopore reads are due to random effects, as enzymes step stochastically both forward and backward and sometimes step too quickly to be clearly resolved, resulting in incorrect step identifications. In DNA sequencers, this random error is typically addressed by obtaining 20× coverage or more, averaging many independent reads of different molecules. However, for a truly single-molecule technology, single-read accuracy is essential.

The identification fidelity of our nanopore protein reader can be greatly increased by obtaining many independent rereadings of the same individual molecule with a succession of controlling helicases, eliminating the random errors that lead to inaccuracies in nanopore reads. At a very high concentration of helicase, on the order of 1 μM, the DNA in

the pore nearly always had a second helicase queued up behind the one controlling its motion (Fig. 3B) (*21*). When the first helicase reached the linker at the end of the DNA section, it could no longer process the molecule and subsequently fell off. The DNA-peptide conjugate was then immediately pulled back into the nanopore such that the queued helicase, which was still bound to the DNA, took control as the new DNA-pulling enzyme. This "rewound" the system and initiated a new and independent read of the peptide. The numbers of rereads on the same single peptide can be very large: Fig. 3A shows an example of a raw data trace with 117 rereads on a single peptide containing the G substitution. This event was purposefully ended by the reversal of voltage to eject the DNA-peptide conjugate from the pore. We observed a typical rewinding distance of ~17 helicase steps, commensurate with a rewinding by a distance of ~17 amino acids, a number that is consistent with the roughly nine DNA bases that are bound within the controlling helicase (*16*). Of the 117 rereads in Fig. 3A, 45 rereads stepped back far enough to provide a reread of the variant site.

We observed significant improvement of the read accuracy with an increasing number of rereads (Fig. 3C). To quantify the increase in the accuracy of the readings as a function of the number of rereadings, we randomly chose subsets of the 45 measured rereads and computed the identification accuracy using N rereads as the fraction of subsets containing N rereads that yielded the correct consensus identification (materials and methods section 8). Even when single reads were limited to as low as ~50% identification accuracy owing to only partial coverage of the variant site, the rereading method allowed single molecules to be identified at high levels of confidence. As the inset in Fig. 3C shows, the error rate decreased with the number of rereads, yielding an undetectably low error rate (<1 in 10^6) when using more than ~30 rereads of an individual peptide. Analysis on reread traces from other variants yielded similar results (fig. S10).

The method described here provides an approach for reading single proteins with sensitivity to single–amino acid changes, which is particularly powerful because of the rereading mode of operation that reduces the stochastic error. Transforming this method into a technology capable of de novo protein sequencing remains a substantial challenge. With any of the 20 amino acids at each position along the protein sequence and a read-head width (*17*) of about eight amino acids, the number of measurements required to build an ion current–to–amino acid map is impractically large. However, many proteomics applications do not require de novo sequencing, instead using other forms of sequence analysis that rely on a priori knowledge of candidate sequences before decoding. These include identifying or "fingerprinting" proteins even in heterogeneous mixtures, mapping posttranslational modifications, and measurements of small samples, all of which involve comparing single-molecule measurements to reference signals of known proteins and interesting variants.

Our methodology has several limitations, but these may be addressed experimentally. Although the pore is capable of translocating heterogeneously charged peptides with neutral polar, nonpolar, negative, and positive amino acids (supplementary text section 1; sample reads shown in fig. S11), highly positively charged peptides may not be efficiently translocated through the pore. Fortunately, analysis of the human proteome reveals that negatively charged stretches of protein sequence are more common than positively charged stretches (*22*), particularly in alkaline pH conditions like those used in our experiments. If needed, the MspA pore can be engineered to provide stronger electro-osmotic forces, which can exceed electrophoretic forces and translocate analytes regardless of charge (*7*, *23*). The read length intrinsic to the technique, ~25 amino acids depending on the length of the DNA-peptide linker, does allow application of this method to many biologically relevant short peptides, such as 8 to 12–amino acid major histocompatibility complex–binding peptides (*24*). Additionally, this finite read length still represents an improvement over the <10 amino acid–long peptide fragments used in mass spectrometry (*25*), and protein fragmentation and shotgun sequencing methods similar to those used in traditional protein sequencing can naturally be applied to this newly developed technique. Technical modifications such as using a variable-voltage control scheme (*18*) have been shown to improve the accuracy of DNA sequencing, and the physical principle that this scheme relies on is equally applicable to peptide sequencing (supplementary text section 2 and fig. S12).

Reads of DNA-peptide conjugates like those presented here could be measured in high throughput with any existing commercially available nanopore sequencing hardware capable of accommodating MspA (e.g., the commercial MinION system) without requiring any reengineering of the device, changing only the sample preparation and data analysis. Furthermore, our methodology retains the features that enabled the success of nanopore DNA sequencing: low overhead cost, physical rather than chemical sensitivity to small changes in single molecules, and the flexibility to be reengineered to target specific applications. Overall, our findings constitute a promising first step toward a low-cost method capable of single-cell proteomics at the ultimate limit of sensitivity to concentration, with a wide range of applications in both fundamental biology and the clinic.

REFERENCES AND NOTES

1. J. A. Alfaro *et al.*, *Nat. Methods* **18**, 604–617 (2021).
2. J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Deamer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13770–13773 (1996).
3. D. Deamer, M. Akeson, D. Branton, *Nat. Biotechnol.* **34**, 518–524 (2016).
4. J. Nivala, D. B. Marks, M. Akeson, *Nat. Biotechnol.* **31**, 247–250 (2013).
5. D. Rodriguez-Larrea, H. Bayley, *Nat. Nanotechnol.* **8**, 288–295 (2013).
6. F. Piguet *et al.*, *Nat. Commun.* **9**, 966 (2018).
7. L. Restrepo-Pérez, C. H. Wong, G. Maglia, C. Dekker, C. Joo, *Nano Lett.* **19**, 7957–7964 (2019).
8. H. Ouldali *et al.*, *Nat. Biotechnol.* **38**, 176–181 (2020).
9. J. Nivala, L. Mulroney, G. Li, J. Schreiber, M. Akeson, *ACS Nano* **8**, 12365–12375 (2014).
10. J. C. Cordova *et al.*, *Cell* **158**, 647–658 (2014).
11. E. A. Manrao *et al.*, *Nat. Biotechnol.* **30**, 349–353 (2012).
12. G. M. Cherf *et al.*, *Nat. Biotechnol.* **30**, 344–348 (2012).
13. I. M. Derrington *et al.*, *Nat. Biotechnol.* **33**, 1073–1075 (2015).
14. S. Yan *et al.*, *Nano Lett.* **21**, 6703–6710 (2021).
15. T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis, J. H. Gundlach, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20647–20652 (2008).
16. J. M. Craig *et al.*, *Nucleic Acids Res.* **47**, 2506–2513 (2019).
17. A. H. Laszlo *et al.*, *Nat. Biotechnol.* **32**, 829–833 (2014).
18. M. T. Noakes *et al.*, *Nat. Biotechnol.* **37**, 651–656 (2019).
19. S. Bhattacharya, J. Yoo, A. Aksimentiev, *ACS Nano* **10**, 4644–4651 (2016).
20. A. D. Tyler *et al.*, *Sci. Rep.* **8**, 10931 (2018).
21. A. C. Rand, thesis, University of California, Santa Cruz, Santa Cruz, CA (2017).
22. R. D. Requião *et al.*, *PLOS Comput. Biol.* **13**, e1005549 (2017).
23. D. P. Hoogerheide, P. A. Gurnev, T. K. Rostovtseva, S. M. Bezrukov, *Nanoscale* **9**, 183–192 (2017).
24. M. Wieczorek *et al.*, *Front. Immunol.* **8**, 292 (2017).
25. J. Cox, N. C. Hubner, M. Mann, *J. Am. Soc. Mass Spectrom.* **19**, 1813–1820 (2008).
26. H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev, C. Dekker, Code and Data for "Multiple re-reads of single proteins at single-amino-acid resolution using nanopores," version 1, Zenodo (2021); https://doi.org/10.5281/zenodo.5506740.

ACKNOWLEDGMENTS

We thank J. Gundlach and his lab members at University of Washington for providing the MspA nanopore and for sharing key pieces of software, and we thank F. Mentzou and X. Shi for assistance with data collection, E. van der Sluis for Hel308 purification, and J. van der Torre for helpful advice on DNA construct preparation. We also acknowledge supercomputer time on the Blue Waters at the University of Illinois at Urbana-Champaign; Expanse at University of California, San Diego; and Frontera at the Texas Advanced Computing Center. **Funding:** Dutch Research Council (NWO) NWO-I680 (SMPS) (C.D.); Dutch Research Council (NWO)/Ministry of Education, Culture and Science (OCW) Gravitation programs NanoFront (C.D.); European Research Council Advanced Grant 883684 (C.D.); European Commission Marie Skłodowska-Curie action Individual Fellowship 897672 (H.B.); European Molecular Biology Organization Short-Term Fellowship 8968 (A.S.W.K); National Institutes of Health grant R21-HG011741 (A.A.); Extreme Science and Engineering Discovery environment allocation MCA05S028 (A.A.); Leadership Resource Allocation MCB20012 on Frontera of the Texas Advanced Computing Center (A.A.). **Author contributions:** H.B. and C.D. conceived of the protein analysis method. H.B. and A.S.W.K. conducted nanopore experiments and analyzed the data. H.B. developed additional analysis code. J.L. and A.A. designed and conducted MD simulations. All authors discussed experimental findings and co-wrote the manuscript. **Competing interests:** TU Delft has filed a patent application (PCT/NL2020/050814) on technologies described herein, with H.B. and C.D. listed as inventors. **Data and materials availability:** All data and custom code used in this paper are available for download online (*26*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4381
Materials and Methods
Supplementary Text
Figs. S1 to S13
Table S1
References (*27*–*40*)
MDAR Reproducibility Checklist

13 July 2021; accepted 24 October 2021
Published online 4 November 2021
10.1126/science.abl4381

FOREST ECOLOGY

Ecosystem fluxes during drought and recovery in an experimental forest

Christiane Werner[1]*†, Laura K. Meredith[2,3,4]†, S. Nemiah Ladd[1]†, Johannes Ingrisch[1,5], Angelika Kübert[1], Joost van Haren[3,6], Michael Bahn[5], Kinzie Bailey[2], Ines Bamberger[1]‡, Matthias Beyer[7], Daniel Blomdahl[8], Joseph Byron[9], Erik Daber[1], Jason Deleeuw[3], Michaela A. Dippold[10,11], Jane Fudyma[12]§, Juliana Gil-Loaiza[2], Linnea K. Honeker[3], Jia Hu[2], Jianbei Huang[13], Thomas Klüpfel[9], Jordan Krechmer[14], Jürgen Kreuzwieser[1], Kathrin Kühnhammer[1,7], Marco M. Lehmann[15], Kathiravan Meeran[5], Pawel K. Misztal[8], Wei-Ren Ng[3], Eva Pfannerstill[9]¶, Giovanni Pugliese[1,9], Gemma Purser[16], Joseph Roscioli[14], Lingling Shi[10,11], Malak Tfaily[4,11,17], Jonathan Williams[9,18]

Severe droughts endanger ecosystem functioning worldwide. We investigated how drought affects carbon and water fluxes as well as soil-plant-atmosphere interactions by tracing $^{13}CO_2$ and deep water $^{2}H_2O$ label pulses and volatile organic compounds (VOCs) in an enclosed experimental rainforest. Ecosystem dynamics were driven by different plant functional group responses to drought. Drought-sensitive canopy trees dominated total fluxes but also exhibited the strongest response to topsoil drying. Although all canopy-forming trees had access to deep water, these reserves were spared until late in the drought. Belowground carbon transport was slowed, yet allocation of fresh carbon to VOCs remained high. Atmospheric VOC composition reflected increasing stress responses and dynamic soil-plant-atmosphere interactions, potentially affecting atmospheric chemistry and climate feedbacks. These interactions and distinct functional group strategies thus modulate drought impacts and ecosystem susceptibility to climate change.

Climate change is increasing the frequency and severity of droughts worldwide, threatening ecosystem functioning (*1*) with the potential to strongly diminish carbon sequestration (*1*–*3*). This is particularly concerning for large tropical forests, which are a major component of the terrestrial carbon sink (*4*). Much of the sink capacity of tropical forests can be lost after severe droughts (*5*), and their overall ability to assimilate and retain carbon is projected to decline further under future warming scenarios (*6*). Moreover, tropical forests represent a major source of atmospheric biogenic volatile organic compounds (VOCs) (*7*), which may be amplified under heat and drought (*8*), causing climate feedbacks through ozone, organic aerosol formation (*9*), and aerosol-radiation interactions (*10*). Although large-scale changes in carbon and water fluxes through ecosystems can be observed through monitoring networks (*11*), many of the mechanisms underpinning these drought-induced dynamics remain unclear (*12*). Specific knowledge gaps include the following: (i) how total ecosystem susceptibility and resilience are shaped by different plant water use strategies (*13*) and access to deep water (*12*); (ii) how drought-stressed vegetation adjusts carbon investments into maintenance and protection (*14*), including the production of VOCs (*8*, *15*); and (iii) how interactions and feedbacks between different components modulate ecosystem dynamics (*16*–*18*).

The Biosphere 2 Tropical Rainforest (*19*) (figs. S1 to S3) is an enclosed experimental ecosystem where interactions among individual ecosystem components in response to environmental changes can be mechanistically studied [e.g., (*20, 21*)]. As part of the Biosphere 2 Water Atmosphere and Life Dynamics (B2WALD) campaign (fig. S4), we imposed a 9.5-week drought on this system (*19*) to determine the mechanisms responsible for the overall changes in ecosystem-scale water and carbon dynamics in response to drought and recovery. Drought propagated dynamically through different forest strata (Fig. 1). Atmospheric drought [vapor pressure deficit, VPD] increased rapidly in the sunlit canopy, whereas the understory was buffered by canopy shading and did not reach maximum values until late into severe drought (Fig. 1A). The soil dried out sequentially, with the topsoil drying rapidly during early drought and the deepest soil layers maintaining high moisture until late into severe drought.

The largest decline in ecosystem water and carbon fluxes coincided with increasing VPD and drying of topsoil during early drought (Fig. 1, B and C). Evapotranspiration (ET), ecosystem respiration (R_{eco}), and gross primary productivity (GPP) declined relative to pre-drought values by 58, 50, and 47%, respectively. Each of these fluxes decreased further during severe drought, but at lower rates (30, 27, and 32% of pre-drought values, respectively). Net ecosystem exchange of carbon (NEE = GPP − R_{eco}) was buffered by concomitant reductions in GPP and R_{eco} during early drought. The forest remained a carbon sink—albeit a small one—despite the 79% reduction in GPP under severe drought.

Daytime concentrations of atmospheric VOCs also changed dynamically in response to drought (Fig. 1E). These changes were not uniform among compounds, and concentrations of distinct VOCs increased sequentially in response to drought: first isoprene, then monoterpenes, and finally hexanal. These VOCs play important roles in leaf stress tolerance and signaling (*22*, *23*), and hexanal has been associated with drought-induced leaf senescence (*24*).

Net uptake of isoprene and monoterpenes by the soil (Fig. 1F) was influenced by both overlying atmospheric concentrations and soil moisture. During drought, the concentration-normalized soil uptake capacity of monoterpenes increased relative to isoprene (fig. S5). This indicates greater persistence of monoterpene scavenging by soils under drought when plant monoterpene emissions were highest. Together, interactions between plants and soil led to distinct patterns in the relative abundance of atmospheric VOC concentrations as the drought progressed (Fig. 1E), serving as a diagnostic indicator of ecosystem drought stress (*24*), with isoprene indicating the onset of ET and GPP reduction and hexanal indicating their final decline under severe drought.

The structured readdition of moisture to the ecosystem—first in groundwater layers and then surface soil—released ecosystem components from drought with different temporal dynamics. The vegetation moderately responded to the addition of deep water, resulting in a slow increase in GPP and ET. Topsoil moisture, soil respiration, and soil VOC dynamics did not respond, indicating that hydraulic lift (*25*) was not sufficient to relieve the system from drought. By contrast, soil respiration and

[1]Ecosystem Physiology, Faculty of Environment and Natural Resources, Albert-Ludwig-University of Freiburg, Freiburg, Germany. [2]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA. [3]Biosphere 2, University of Arizona, Oracle, AZ, USA. [4]BIO5 Institute, The University of Arizona, Tucson, AZ, USA. [5]Department of Ecology, University of Innsbruck, Innsbruck, Austria. [6]Honors College, University of Arizona, Tucson, AZ, USA. [7]Institute of Geoecology - Environmental Geochemistry, Technical University Braunschweig, Braunschweig, Germany. [8]Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX, USA. [9]Department of Atmospheric Chemistry, Max Planck Institute for Chemistry, Mainz, Germany. [10]Biogeochemistry of Agroecosystems, University of Göttingen, Göttingen, Germany. [11]Geo-Biosphere Interactions, University of Tuebingen, Tuebingen, Germany. [12]Department of Environmental Science, University of Arizona, Tucson, AZ, USA. [13]Max Planck Institute for Biogeochemistry, Jena, Germany. [14]Aerodyne Research, Billerica, MA, USA. [15]Forest Dynamics, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland. [16]Centre for Ecology and Hydrology, University of Edinburgh, Edinburgh, UK. [17]Pacific Northwest National Laboratory, Richland, WA, USA. [18]Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus.
***Corresponding author. Email: c.werner@cep.uni-freiburg.de** †These authors contributed equally to this work. ‡Present address: Atmospheric Chemistry Group, University of Bayreuth, Bayreuth (BayCEER), Germany. §Present address: Department of Land, Air and Water Resources, University of California Davis, Davis, CA, USA. ¶Present address: Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, CA, USA.

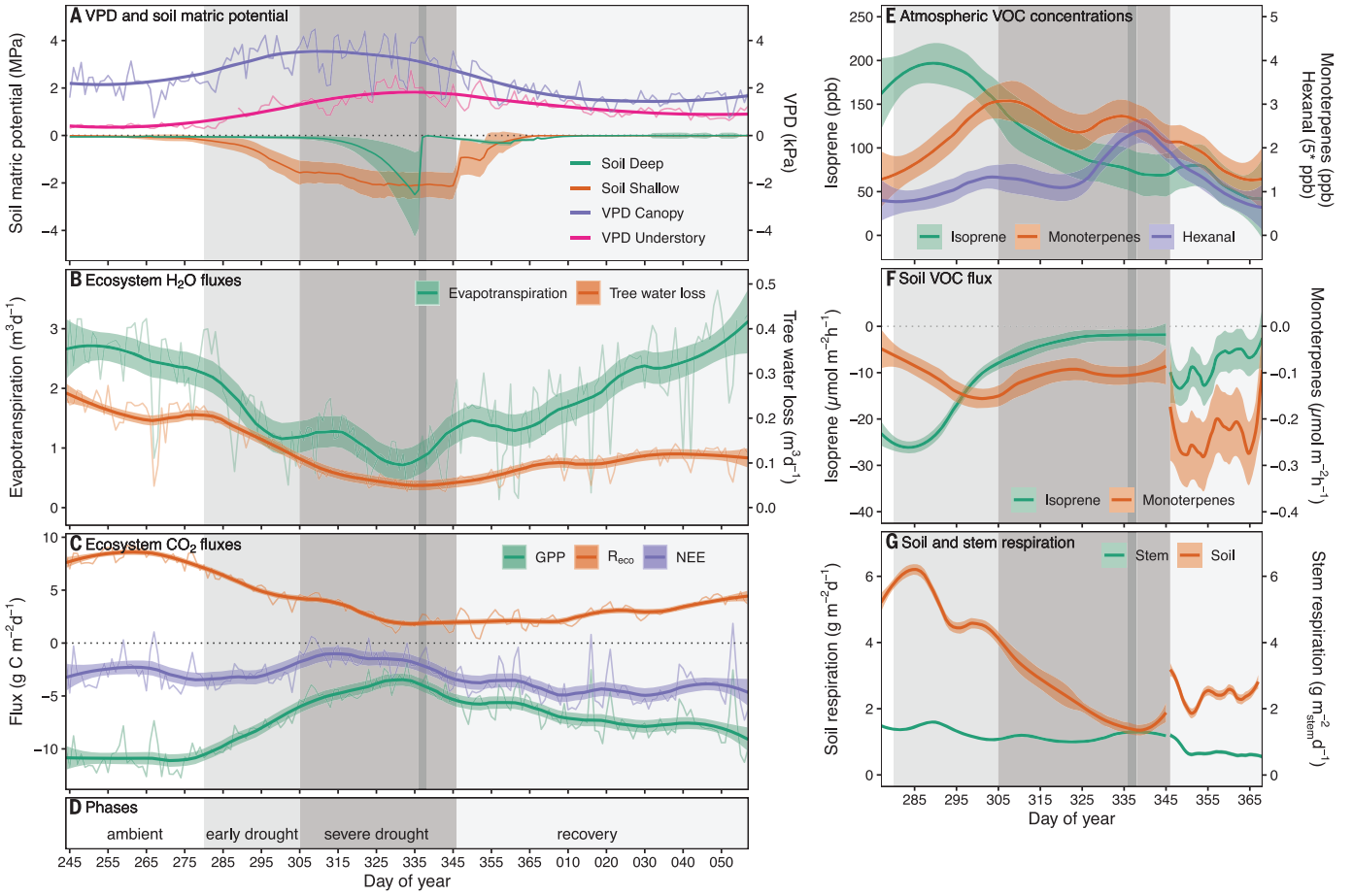


Fig. 1. Ecosystem fluxes of CO_2, H_2O, and VOCs, as well as environmental drivers during drought and recovery. Daily means of (**A**) vapor pressure deficit (VPD) in the tree canopy and understory in addition to soil matric potential in shallow and deep soil; (**B**) ecosystem evapotranspiration and tree water loss; (**C**) ecosystem CO_2 fluxes: gross primary productivity (GPP), ecosystem respiration (R_{eco}), and net ecosystem exchange (NEE); (**D**) drought phases; (**E**) atmospheric VOC concentrations in parts per billion (ppb); (**F**) daily maximum soil uptake rate of isoprene and monoterpenes (n = 12); and (**G**) daily means of stem (n = 12) and soil respiration (n = 12). Background shading in all panels matches phases in (D). The vertical dark gray line indicates deep soil rewetting. Thick lines are smoothing splines with 95% confidence intervals, and thin lines in (A) to (C) are daily mean fluxes.

VOC uptake increased immediately upon rain rewetting (Fig. 1). Within 3 months of drought release, ET values had returned to pre-drought levels, but carbon fluxes displayed pronounced legacy effects, with R_{eco} and GPP recovering to 45 and 90% of pre-drought values, respectively.

The vegetation displayed highly diverse drought responses among and within species, driven by the interplay of species-specific drought adaptations and differences in microclimate conditions. We identified four major functional response types: drought-tolerant and drought-sensitive canopy-forming trees and drought-tolerant and drought-sensitive understory plants (table S1).

Drought-sensitive canopy trees dominated total ecosystem water fluxes under pre-drought conditions but had the largest reductions in sap flow coincident with upper soil drying in early drought (Fig. 2), when leaf senescence and substantial leaf shedding were observed. Leaf aging and shedding occurred in all species to different degrees, but leaf shedding was higher in drought-sensitive canopy trees, particularly in *Clitoria fairchildiana* (~25% leaf loss). By contrast, drought-tolerant canopy trees had significantly lower water fluxes than their drought-sensitive counterparts under non-stressed conditions ($P < 0.001$); they also had more moderate reductions in sap flow with no visible signs of senescence during early drought.

The water fluxes of understory plants were one order of magnitude lower than those of canopy trees (Fig. 2). Drought-sensitive understory plants displayed the strongest decline in pre-dawn and midday leaf water potential, indicating severe stress. Drought-tolerant understory plants showed only modest signs of drought stress. The presence of a robust understory in locations where shading buffered changes in VPD and temperature helped maintain ecosystem functioning at reduced levels during the severe drought. Thus, an intact forest canopy can provide an important beneficial forest microclimate, with a critical role in buffering the long-term effects of climate warming (*26*). Overall, total plant water flux from drought-tolerant plants in both strata became more important during the drought, with their combined contribution increasing from 15 (pre-drought) to 35% (severe drought) (Fig. 2B). By contrast, drought-sensitive canopy trees dominated the total water flux (74 and 57% during pre-drought and severe drought, respectively).

The swift reduction in water use by drought-sensitive canopy trees (Fig. 2A) preceded deep soil drying by several weeks (Fig. 1A). Contrary to the expectation that trees would use all accessible water before reducing leaf area and carbon-fueled activities, ^2H enrichment of transpiration after addition of a ^2H_2O-tracer to the deep soil during severe drought demonstrated that all canopy trees in both the drought-sensitive and drought-tolerant groups had access to deep water (Fig. 3). However, despite accessibility by roots, these deep water

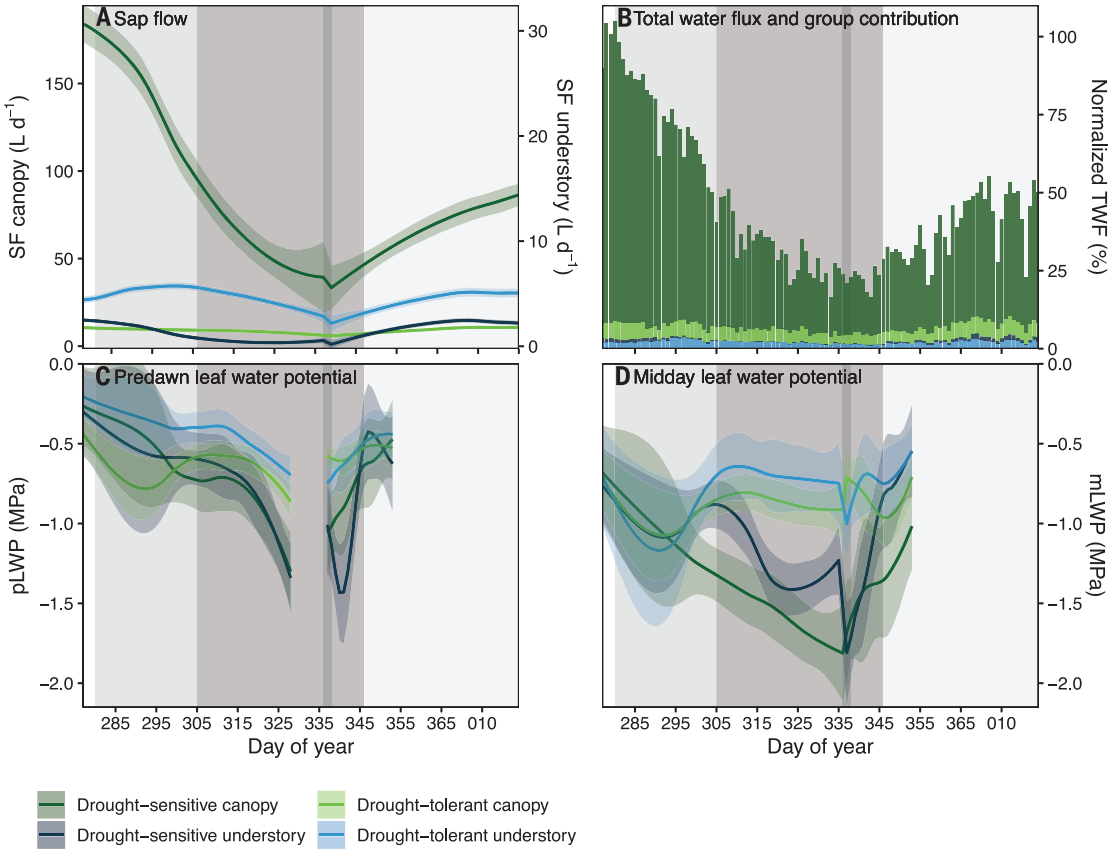


Fig. 2. Physiological responses of different plant functional groups and impact on total water flux. (**A**) Sap flow (SF) (n = 3 to 8 individuals per functional group). (**B**) Normalized total water flux (normalized TWF) and relative functional group contribution (percent). (**C**) Predawn (pLWP) and (**D**) midday (mLWP) leaf water potential (n = 3 to 6). Lines are based on locally estimated scatterplot smoothing. Background shading is as in Fig. 1. For information on plant groupings, see table S1.

reserves were largely spared until severe drought (Fig. 1A). Additionally, despite strong drought stress, there was a pronounced lag between uptake and transpiration of deep water (Fig. 3A). The peak of ^{2}H-enriched transpiration occurred only after the deep soil had dried after addition of ^{2}H-labeled water, and when unlabeled precipitation was once again available in the forest. The several-week delay of maximum transpired δ^{2}H values indicates that stem water residence time was unexpectedly long, with the quickest response in drought-sensitive canopy trees and the greatest delay for drought-tolerant canopy trees. This is consistent with low sap flow velocities of drought-tolerant trees (Fig. 2) and with the recovery of tree water content in stems in drought-sensitive species (Fig. 3B).

Although early drought responses by sensitive plants reduced water loss and increased ecosystem drought resistance, these adaptations delayed recovery after the return of rain by slowing the flux of water through the ecosystem. Even after 7 weeks, the total canopy water flux had recovered to only 54% of pre-drought values (Fig. 2B). Drought-sensitive species contributed to a rapid resumption of ecosystem function upon rewetting, but at reduced rates (Fig. 2A). Specifically, legacy effects were caused by persistent structural changes in drought-sensitive trees (loss of hydraulic conductivity and leaves). By contrast, drought-tolerant species did not exhibit these structural changes and became relatively more important to ecosystem function during drought, contributing to overall ecosystem resistance. However, their generally slower responses limited the contribution of drought-tolerant trees during recovery.

In addition to reducing water and CO_2 fluxes through the ecosystem, the rate at which assimilated carbon moves through ecosystems can decline under drought (*14*, *27*, 28). We traced carbon allocation from leaves to stems and soils with whole-ecosystem $^{13}CO_2$ pulse-chase experiments (Fig. 4) (19). Reduced carbon assimilation under drought resulted in lower ecosystem ^{13}C uptake (37% of pre-drought values; Fig. 4H) despite addition of twice as much ^{13}C label to the atmosphere under drought (Fig. 4A). Although ecosystem assimilation declined under drought, bulk leaf material became more ^{13}C enriched (Fig. 4, B and C, and table S2), with a greater increase for drought-sensitive plants. Likewise, δ^{13}C of leaf respiration was higher after the drought $^{13}CO_2$ pulse than pre-drought, with greater increases for canopy trees relative to understory plants and for drought-sensitive plants relative to drought-tolerant plants (Fig. 4, B and C, and table S3). Although metabolic activity was reduced, the fresh carbon was preferentially used as respiratory substrate.

Belowground ^{13}C transport was slower under drought, as indicated by δ^{13}C values of stem respiration, whereas the ^{13}C residence time in stems increased (Fig. 4, D and I, and tables S3 and S4). ^{13}C enrichment of fine roots was ultimately similar after both label pulses (Fig. 4E). However, soil respiration exhibited less ^{13}C enrichment during the drought, suggesting that ^{13}C was preferentially used for building root biomass rather than for root respiration or exudates.

Despite reduced ecosystem carbon uptake and total VOC emissions, plants continued to allocate a similar proportion of fresh carbon to de novo VOC synthesis, as incorporation of ^{13}C into both isoprene and monoterpenes remained high (Fig. 4, F and G). Maintaining carbon allocation into VOC synthesis demonstrates the fundamental role of these compounds in protecting plants from heat stress and photooxidative damage that can be caused by reduced stomatal conductance and C metabolism under drought (*8*).

The B2WALD experiment demonstrates the importance of plant functional groups for understanding ecosystem responses to drought (fig. S7) (*12*, *29*–*30*). Different hydraulic strategies of distinct plant functional groups and

their changing proportional contributions determine the overall magnitude of reduction in ecosystem fluxes as drought progresses (Figs. 2 to 4). Moreover, plant adaptations determine the velocity and residence time of water and carbon in different plant organs such as stems and leaves (Figs. 2 to 4). Therefore, distinct plant hydraulic strategies can increase resistance to drought: The presence of drought-tolerant trees buffered the drought decline in ecosystem fluxes, whereas the rapid restriction of water use in drought-sensitive trees lowered the risk of detrimental depletion of soil water sources, which improves the likelihood of withstanding prolonged droughts. The degree to which ecosystems are able to resist extreme droughts will affect their functioning upon return to more favorable hydrologic regimes. In our study, the drought's legacy was visible as the carbon sink capacity of the forest remained suppressed months after the cessation of drought (Fig. 1), partly as a result of the physiological and structural responses of drought-sensitive canopy trees. This has important implications for coupled climate-vegetation Earth system models, which will miss critical processes if distinct plant hydraulic responses are not considered (*12*, *31*).

Notably, the forest flux dynamics were tightly coupled to shallow soil processes, as the largest decrease in ecosystem fluxes corresponded to early dry-down of topsoil moisture (Fig. 1, S8), despite access to deep water reserves by all canopy trees (Fig. 3). Plants can dynamically adjust their root water uptake depth through active regulation of conductivity in the rooting zone (*32*); however, most roots are allocated to the upper soil zone, where they mediate important soil-plant interactions in the rhizosphere through carbon and nutrient exchange (*18*). Drying of this zone induced significant down-regulation of transpiration—particularly in drought-sensitive canopy trees—indicating that changes in critical zones trigger drought responses, rather than changes in the total water accessible to plants. Reductions in transpiration and assimilation were associated with delayed transport of carbon belowground (Fig. 4) and with a likely reduction in fresh carbon allocated to soil respiration (Figs. 1 and 4). Although phloem transport in these trees was delayed by the drought, its magnitude was not reduced (*14*). Additionally, the rate at which water moved from deep soils to the leaves of drought-stressed, deep-rooted trees was unexpectedly slow (*33*), with especially long residence times for stem water in drought-tolerant canopy trees (>55 days; Fig. 3). Incorporation of different plant hydraulic strategies and dynamics in response to sequential soil drying in models may be critical for accurately simulating the magnitude and timing of vegetation-driven changes in ecosystem fluxes and land-atmosphere feedbacks under climate change (*16*, *31*).

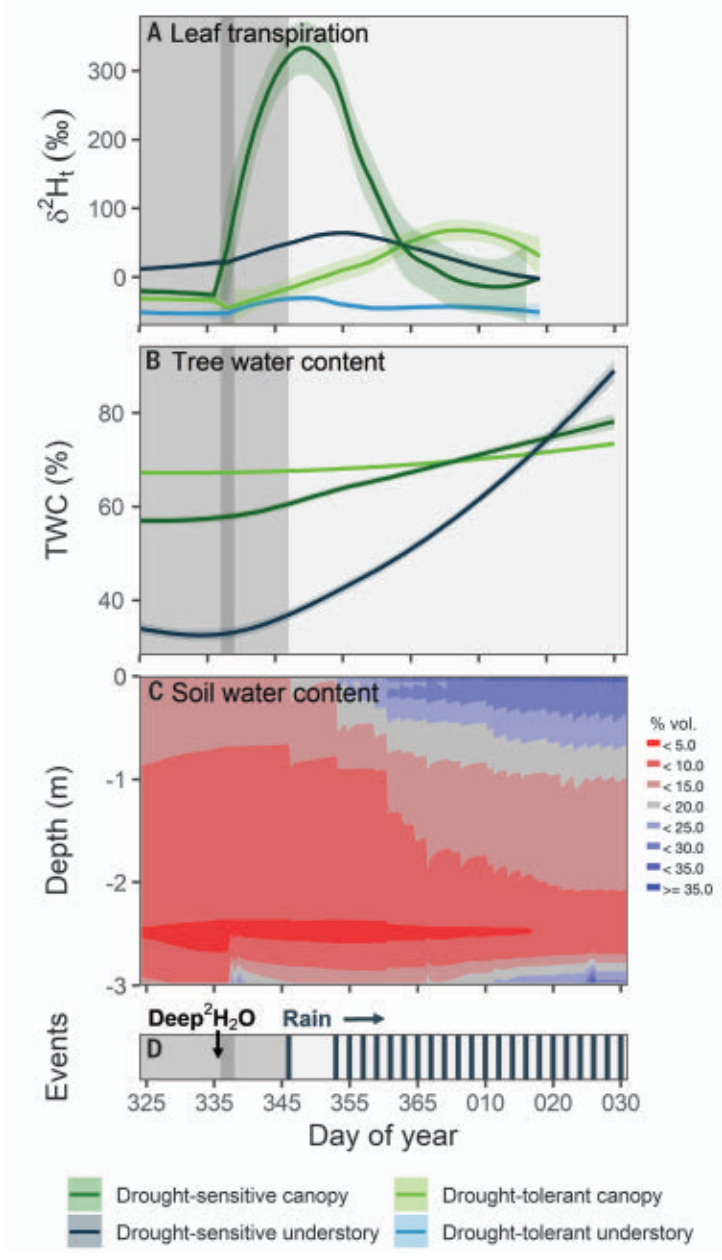


Fig. 3. Deep water uptake and recovery of plant functional groups. A ^2H_2O-pulse was applied at depth on day 336 (gray line) 10 days before rain. (**A**) δ^2H value of leaf transpiration (= δ^2H_t = $(^2H/^1H)_{transpired\ water}$/$(^2H/^1H)_{VSMOW}$ − 1, per mil) (n = 5 to 10 per functional group), (**B**) tree water content (TWC) (n = 2 to 6; full time series shown in fig. S6), (**C**) soil water content per depth (percent volume), and (**D**) rain events after drought. Lines in (A) to (C) are based on locally estimated scatterplot smoothing. No data were available for the TWC of drought-tolerant understory plants. For information on plant groupings, see table S1.

Additionally, Earth system models need to account for coupled plant-soil interactions, which not only change the movement of carbon and water through ecosystems but can also impact atmospheric chemistry. In the B2WALD drought experiment, plant-soil interactions led to a distinct pattern in atmospheric VOC concentrations, which tracked the increase in drought severity (Fig. 1). Moreover, the relative increase in monoterpene concentrations in early drought can increase atmospheric reactivity, which could potentially promote secondary aerosol formation—including cloud condensation nuclei (*34*)—representing a feedback that could favor rain and help relieve some of the forest's drought stress (*35*). Our findings suggest that the goal of understanding and predicting ecosystem function in response to global climate change will be best accomplished by incorporating plant functional

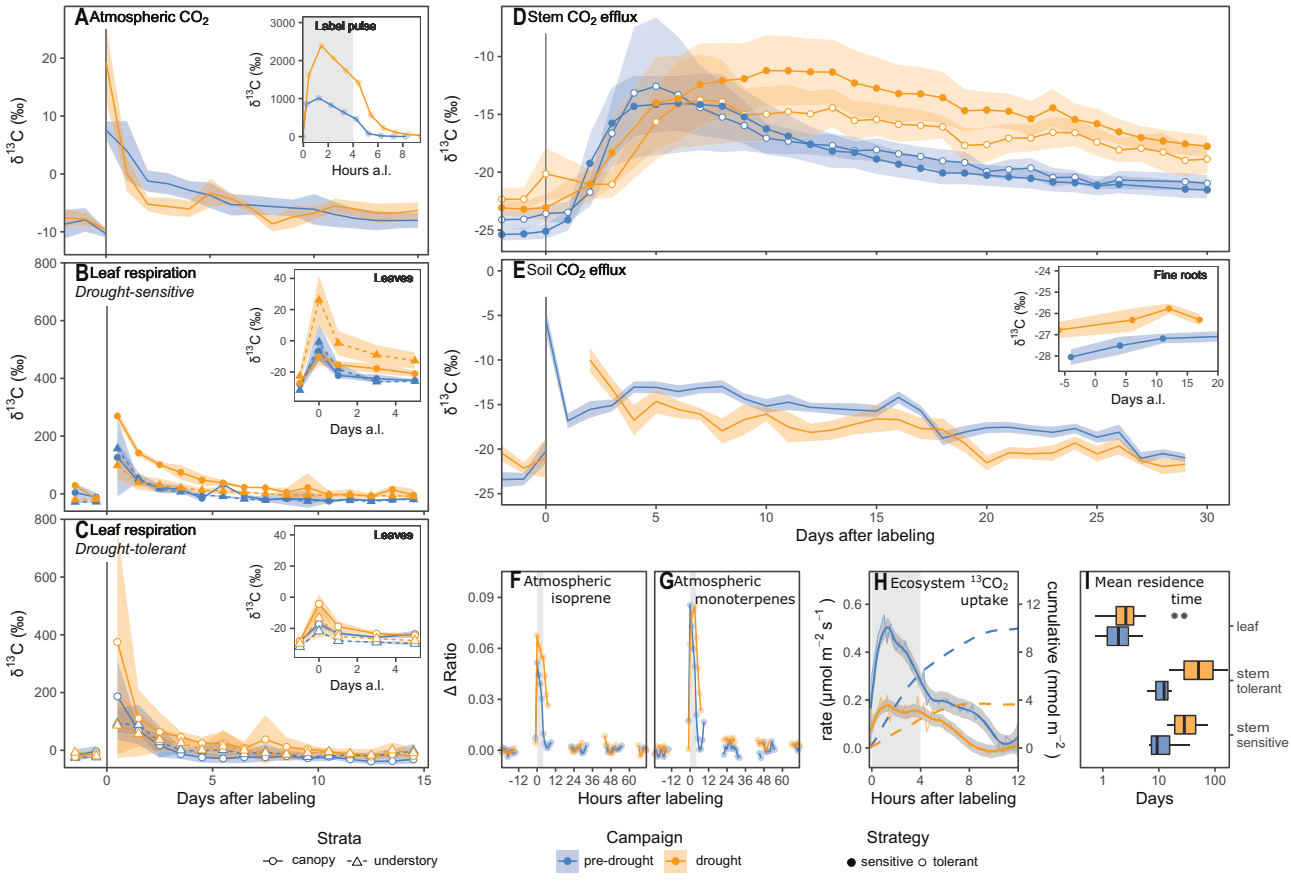


Fig. 4. ^{13}C dynamics in fluxes and pools of ecosystem compartments after pulse labeling. Isotopic composition in (**A**) atmospheric CO_2, (**B** and **C**) leaf respiration (n = 3 to 6 per functional group), (**D**) stem respiration (n = 3 to 6), (**E**) soil CO_2-efflux (n = 12), and (**F** and **G**) atmospheric isoprene and monoterpenes, respectively, relative to the time of $^{13}CO_2$-pulse labeling (gray shading) during pre-drought (blue) and drought (orange) conditions. (**H**) Uptake rate (left axis) and cumulative uptake (right axis) of $^{13}CO_2$ during and in the initial hours after each labeling pulse (a.l.), derived from ecosystem GPP and $\delta^{13}C$ of atmospheric CO_2 ($\delta^{13}C = (^{13}C/^{12}C)_{sample}/(^{13}C/^{12}C)_{VPDB} - 1$, per mil). (**I**) Mean residence time of ^{13}C in leaf- and stem-respired CO_2 (n = 3 to 6). Open and closed symbols in (B) to (D) denote drought-tolerant and drought-sensitive functional groups, respectively. Circles and triangles indicate canopy and understory, respectively. The ΔRatio in (F) and (G) is the ^{13}C enrichment above the natural abundance of ^{13}C isotopologues. Insets in (A) to (C) and (E) are as follows: (A) strength of atmospheric $^{13}CO_2$ pulse; (B) and (C) $\delta^{13}C$ of leaf samples; and (E) $\delta^{13}C$ of root samples (n = 2 to 7). Symbols show mean values ± standard error.

groups and their interactions with soils and atmosphere into both observational and modeling studies.

REFERENCES AND NOTES

1. G. B. Bonan, *Science* **320**, 1444–1449 (2008).
2. M. Reichstein *et al.*, *Nature* **500**, 287–295 (2013).
3. A. Bastos *et al.*, *Sci. Adv.* **6**, eaba2724 (2020).
4. Y. Pan *et al.*, *Science* **333**, 988–993 (2011).
5. J.-P. Wigneron *et al.*, *Sci. Adv.* **6**, eaay4603 (2020).
6. W. Hubau *et al.*, *Nature* **579**, 80–87 (2020).
7. A. B. Guenther *et al.*, *Geosci. Model Dev.* **5**, 1471–1492 (2012).
8. F. Loreto, J.-P. Schnitzler, *Trends Plant Sci.* **15**, 154–166 (2010).
9. A. Arneth *et al.*, *Nat. Geosci.* **3**, 525–532 (2010).
10. N. Unger, *Geophys. Res. Lett.* **41**, 8563–8569 (2014).
11. H. Chu, D. D. Baldocchi, R. John, S. Wolf, M. Reichstein, *J. Geophys. Res. Biogeosci.* **122**, 289–307 (2017).
12. W. R. L. Anderegg *et al.*, *Nature* **561**, 538–541 (2018).
13. T. J. Brodribb, J. Powers, H. Cochard, B. Choat, *Science* **368**, 261–266 (2020).
14. H. Hartmann, M. Bahn, M. Carbone, A. D. Richardson, *New Phytol.* **227**, 981–988 (2020).
15. J. K. Holopainen, J. Gershenzon, *Trends Plant Sci.* **15**, 176–184 (2010).
16. N. M. Levine *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 793–797 (2016).
17. J. Joseph *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24885–24892 (2020).
18. F. I. Pugnaire *et al.*, *Sci. Adv.* **5**, eaaz1834 (2019),
19. See supplementary materials for ecosystem fluxes during drought and recovery in an experimental forest.
20. U. Rascher *et al.*, *Plant Cell Environ.* **27**, 1239–1256 (2004).
21. M. N. Smith *et al.*, *Nat. Plants* **6**, 1225–1230 (2020).
22. R. K. Monson, S. M. Weraduwage, M. Rosenkranz, J.-P. Schnitzler, T. D. Sharkey, *Oecologia* **197**, 885–902 (2021).
23. C. Werner, L. Fasbender, K. M. Romek, A. M. Yáñez-Serrano, J. Kreuzwieser, *Front. Plant Sci.* **11**, 1242 (2020).
24. K. J. Jardine *et al.*, *Plants* **4**, 678–690 (2015).
25. J. Aguirre-Gutiérrez *et al.*, *Ecol. Lett.* **22**, 855–865 (2019).
26. F. Zellweger *et al.*, *Science* **368**, 772–775 (2020).
27. D. Epron *et al.*, *Tree Physiol.* **32**, 776–798 (2012).
28. N. K. Ruehr *et al.*, *New Phytol.* **184**, 950–961 (2009).
29. J. Aguirre-Gutiérrez *et al.*, *Nat. Commun.* **11**, 3346 (2020).
30. S. Haberstroh *et al.*, *New Phytol.* **231**, 1784–1797 (2021).
31. W. R. L. Anderegg, A. T. Trugman, D. R. Bowling, G. Salvucci, S. E. Tuttle, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14071–14076 (2019).
32. M. Dubbert, C. Werner, *New Phytol.* **221**, 1754–1763 (2019).
33. M. Sprenger *et al.*, *Rev. Geophys.* **57**, 800–834 (2019).
34. E. Y. Pfannerstill *et al.*, *Front. For. Glob. Change* **1**, 12 (2018).
35. U. Pöschl *et al.*, *Science* **329**, 1513–1516 (2010).
36. L. Meredith *et al.*, B2WALD campaign team and contributions, Version 2, University of Arizona Research Data Repository (2021).

ACKNOWLEDGMENTS

We acknowledge valuable support from all members of the B2WALD team, as detailed in the B2WALD contribution list (*36*). **Funding:** This work was funded by the European Research Council [ERC consolidator grant 647008 (VOCO2) to C.W.] and financial support from the Philecology Foundation to Biosphere 2 to L.K.M., with in-kind support detailed elsewhere (*36*). **Author contributions:** C.W., L.K.M., and S.N.L. conceived and designed the study; all authors collected and/or analyzed data; C.W. wrote the first draft with input from L.K.M. and S.N.L. Data analysis was led by J.I., A.K., J.v.H., D.B., J.K., L.K.M., S.N.L., and C.W., with input from all authors. All authors revised the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data used in this manuscript are publicly available (DOI: 10.25422/azu.data.14632593).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj6789
Materials and Methods
Figs. S1 to S9
Tables S1 to S8
References (*37*–*61*)

30 May 2021; accepted 13 October 2021
10.1126/science.abj6789

Meet BIT

—We sincerely invite global talents to join Beijing Institute of Technology.

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Founded in 1940, BIT has been one of the key universities in China and the first batch of universities which has been listed in the national "211 Project", "985 Project" and the "Top A World-class Universities".

Global Recruitment System

Beijing Institute of Technology implements an internationally adopted tenure system which is in line with the university strategy. It provides an academic path from tenure-track professor to tenured professor to dedicated professor for our faculties. Now we are recruiting leading academics as well as potential young scholars globally.

Dynamic Faculty Development Mechanism

Based on the layout of our academic disciplines, and following the paths of talent growth, BIT has a number of innovation research centers and young scientists' studios on the way. They will enjoy enough autonomy and long-term and stable preferential terms including but not limited to human resources, financial support, and equipment allocation. (BIT now has 12 national research experimental platforms, 65 provincial and ministerial level research platforms, 23 school-level public platforms, and 14 peripheral scientific research platforms.)

Attractive Welfare and Benefits

1. BIT provides sufficient start-up research grants for newly-recruited talents according to their corresponding positions. BIT gives priority to support them in working conditions, team building, and doctoral / Master candidate enrollment in consideration with their actual needs in teaching and research.

2. BIT adheres to high standards in terms of recruiting talents. The university also provides public institution staffing quotas and internationally competitive salary and welfare with long-term pay progression.

3. BIT provides housing subsidies and assists new faculty members in finding temporary housing.

4. BIT provides assistance in school or kindergarten access for children of the new faculty members.

Current Opportunities

Dedicated positions:

Including Outstanding Professors, Chair Professors and Distinguished Professors.

- Academic leaders with outstanding academic attainments.
- Significant achievements recognized by international peers in their research fields.
- Capabilities to catch up with or surpass the international academic frontier, and to lead an academic team in key academic areas.

Tenured positions:

Including Tenured Professors and Tenured Associate Professors.

- Recognized as a leading academic in his/her own academic fields;
- Significant achievements and influence among peers;
- Great potential for innovation and development.

Tenure-track positions:

Including Tenure-track Professors, Tenure-track Associate Professors and Tenure-track Assistant Professors.

- With doctoral degree from world-leading universities; with a good academic background andpotential.

Postdoctoral researcher:

- With a doctoral degree obtained during the past 3 years; with a good academic background.

To apply & contact us

I: Contact us

Academic Positions	Contact information
Tenure-track Professor/Tenured Professor/Dedicated Professor	Contact: Li Chuandong, Ma Kai Tel: +86 10 68918117/4243 Email: bitzhaopin@bit.edu.cn
Postdoctoral researcher	Contact: Huang Mei, Zhang Ju Tel: +86 10 68914539 Email: bitpostdoctor@bit.edu.cn

BIT HR Office

BIT Recruitment

II: Apply online: http://rszhaopin.bit.edu.cn/zp.html#/

III: Please scan the QR codes for further information

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

ScienceCareers online @sciencecareers.org
OPPORTUNITIES IN CHINA

New heights of academic brilliance

ShanghaiTech University is building new infrastructure and creating incentives for research and teaching innovations.

As China looks to become carbon-neutral by 2060, new opportunities arise in targeting the largest population in the world for cross-disciplinary research and teaching, according to Yin Jie, vice president and provost of ShanghaiTech University. Jointly founded by the Chinese Academy of Sciences and Shanghai Municipal People's Government in 2013, ShanghaiTech has an embedded culture of research innovation to address global priorities in environmental and public health.

"Shanghai is home to the best ranked universities in China defined by scope and scale. At ShanghaiTech, success is about being the best at where our strengths are. These are built upon our dedication to an open environment to promote independent research of real impact, and not just measured by the number and comprehensiveness of our programmes," says Yin.

Photon science, for example, is one of the subjects with significant potential. Yin explains that the tradition of optics research in China runs deep, but the development of its branch in photonics is relatively recent. "Developing photonic infrastructure could take a long time, and may not yield a sharp increase in number of papers, but such studies could shed light on green energy, serving national goals and the needs of the international scientific community," he says.

A growing cluster for photon science

At a global photon science research centre at Zhangjiang Comprehensive National Science Center in Shanghai, new facilities have been designed to study the internal structure and dynamic changes of live cells, and catalytic reactions at the femtosecond level, a quadrillionth of a second, to glean clues about the underlying mechanisms. Supported by national funding and state laboratories, ShanghaiTech has been a main player at each step, from creating the blueprint for large-scale facilities, to coordinating their assembly.

A materials scientist by training, Yin explains that Shanghai-XFEL Beamline Project (SBP) of the Shanghai Soft X-ray Free Electron Laser Facility (SXEFL) could allow researchers to view catalysts in continuous motion at the molecular level.

"Many catalysts work at high-temperature, high-pressure conditions that require substantial energy expenditure. With SXEFL, we re-examined catalyst design to reduce by-products and reaction requirements, thereby minimizing environmental pollution," says Yin. "One example is perovskite, a popular catalyst that is attracting much attention for its low production cost and high conversion efficiency from solar energy to electrical energy."

SXEFL, working at wavelengths as short as 2.0 nm, also holds great promise for viewing biological specimens and

ADVERTISEMENT

organic molecules in their natural aqueous environment, such as in live cell imaging. In June 2021, ShanghaiTech announced SXEFL's full coverage of the "water window", an electromagnetic spectrum where oxygen and water are transparent to soft X-rays, which are absorbed however by other building blocks of life such as carbon and nitrogen.

This series of infrastructure developments, with plans extending to 2025 and beyond, include Shanghai Superintense Ultrafast Laser Facility (SULF), Shanghai HIgh repetitioN rate XFEL and Extreme light facility (SHINE), and Shanghai Synchrotron Radiation Facility (SSRF). SSRF, for example, hosts the Nano Angle Resolved Photoelectron Spectroscopy (NanoARPES), the first of its kind in China, designed and developed by ShanghaiTech. Operationally ready since October 2021, the station will enable X-ray photons to focus on an extremely small solid sample via synchrotron radiation, thus helping to detect electronic structure and advance studies into low dimensional and quantum materials.

"These facilities welcome not only ShanghaiTech staff and students, but also researchers from other parts of China and the world," adds Yin.

Innovations from inside out

Along with the new teaching and research opportunities designed around these new infrastructures, Yin points to national support for land acquisition, campus expansion, as well as other financial investments. There is academia-business synergy in its Zhangjiang neighbourhood that has encouraged a growing number of start-ups and multinationals specializing in emerging fields and crossroads of science.

The faculty structure at ShanghaiTech reflects this need for increasing collaboration, such as in materials science, information engineering, and public health, defying traditional categorization to flexibly tackle translational complexities.

"While the impact of our materials science innovations sees potential to reach other branches of basic science including physics and chemistry, similarly, an increasing number of our projects in immunology and biomedical engineering demonstrate the potential for clinical studies," says Yin. "These could benefit from our new infrastructure and network of national laboratories across Shanghai." Furthermore, there is also a push for young researchers and students to take centre stage, which for Yin, could mean taking a step back to reform the traditional top-down approach towards academic planning and resource allocation.

"Younger researchers may lack access to funding and other support compared to their senior peers, and China is not alone in this problem," he says. "But we have devoted greater attention to an inclusive environment and a support system to encourage long-term personal growth, together with a shift of emphasis from knowledge transfer to skill building."
Securing the funding and having the infrastructure in place are essential, but he also sees the value of effective institutional structure and incentives. Yin wants to ensure an open environment with enough room for young academics to explore their own research interests, while taking responsibility for their own research and career path. Students are encouraged to take initiative in starting their own projects and working together on international competitions.

After all, the greatest strength of ShanghaiTech, he believes, stems not only from its world-class facilities, but its talent. "We're here to support them all the way by offering them world-class resources and all the guidance they need".

ADVERTISEMENT

OPPORTUNITIES IN CHINA
Science Careers online @sciencecareers.org

Donghua University Global Recruiting Program

DHU (Donghua University) is the first textile institution of higher learning in the People's Republic of China. With having "Virtue, Erudition, Aspiration & Earnestness" as a motto, DHU is working to become a unique multidisciplinary, high-level institution with a focus on textile, material, as well as design. The university has trained more than 260,000 talents for the country. A large number of scientific research achievements are widely used in aerospace, major construction projects, environmental protection and other fields, making constructions to "Shenzhou", "Tiangong", "Beidou", "Chang'e".

DHU adheres to the core strategy of strengthening the university with talents, continuously optimizes the structure of the talent team, and strengthens the construction of the faculty. There are more than 2,300 faculty members and staff and more than 1,400 full-time teachers, including more than 900 teachers with senior professional titles. It has gradually formed a group of teachers that includes high-level talents, such as academicians, recipients of National Science Fund for Distinguished Young Scholars, recipients of National Science Fund for Excellent Young Scientists, Chief Scientist of the National 973 Programme, Chief Scientist of National Key Research and Development program, and other high-level talents as well as outstanding young and middle-aged academic elites.

The 7th Donghua University Shangshi Forum for International Young Scholars will be held online on 24th Dec. 2021. Welcome young scholars to exchange latest research trends. More information please contact us by email rcb@dhu.edu.cn or by telephone (+86)021-67792042.

Colleges (Departments, Centers)

College of Textile
College of Fashion and Design
Glorious Sun School of Business and Management
College of Mechanical Engineering
School of Information Sciences and Technology
School of Computer Sciences and Technology
College of Chemistry, Chemical Engineering and Biotechnology
College of Materials Science and Engineering
College of Environmental Science and Engineering
College of Humanities
College of Marxism Studies
College of Science
College of Foreign Languages
Department of Physical Education
Continuing Education College
International Cultural Exchange School
Shanghai International College of Fashion and Innovation
Innovation Center for Textile Science and Technology
Center for Civil Aviation Composites
Center for Advanced Low-dimension Materials
Shanghai International Fashion Innovation Center
Institute of Functional Materials
Institute of Artificial Intelligence

Listed into the Shanghai Universities First-Class Disciplines Project

Textile Science and Engineering
Materials Science and Engineering
Design
Mechanical Engineering
Control Science and Engineering
Environmental Science and Engineering
Chemistry

Recruitment Positions

Distinguished Young Professor:
1. Under the age of 35 for researchers in natural science and engineering, or under the age of 40 for researchers in humanities and social sciences.
2. Applicant should get PhD degree and have postdoctoral experience or obtained assistant professorship or above in overseas universities; or professors in domestic universities or institutions.

ECNU Seeking Global Talents

Founded in 1951, and based in Shanghai, East China Normal University (ECNU) is one of the top research universities in China. Directly under the Ministry of Education and sponsored by the national key university programs—"Project 211" and "Project 985"—our university is renowned for its teacher education and high-level research programs in both basic and applied sciences. In 2017, ECNU was selected into Class A of the "Double First-Class Program" (First Class University and First-Class Academic Discipline in the World), an education initiative launched by the Ministry of Education aiming at developing elite Chinese universities into world-class institutions by the end of 2050. Adhering to the strategy of pursuing interdisciplinary development, internationalization, and informatization, ECNU has made great achievements in talent training, scientific research, community service and international exchanges in the past few decades. With all these efforts, ECNU intends to fulfill its vision of "Creativity, Character and Community" and becomes a more prestigious university with higher visions. Join ECNU and grow up with us together!

Job Opportunities and Qualifications

ECNU invites applications for tenure-track faculty positions in multiple disciplines in all ranks from assistant researchers to professors. The positions of research scientists and postdoctoral fellowships are also open for application.

Applicants must possess a doctorate degree in a closely related field and show evidence of relevant teaching and research experience, including peer-reviewed publications. They must have a professional record that proves their external funding history or potential.

How to Apply

If you are interested, please contact our HR office for more details regarding the applying process and application materials

Contact: Mr. Wang

E-mail: cwang@bio.ecnu.edu.cn

Office Tel: 86-21-62232452/86-21-54836171

Personnel Department: http://hr.ecnu.edu.cn

ECNU Recruitment Announcement

ECNU Talent Recruitment Platform

ScienceCareers online @sciencecareers.org

DEPUTY DIRECTOR FOR INTRAMURAL RESEARCH

National Institutes of Health Department of Health and Human Services

THIS POSITION IS SUBJECT TO THE COVID-19
VACCINE MANDATE, AS A CONDITION OF EMPLOYMENT

Do you have a vision for the future of intramural research? Are you a first-rate Scientific Leader seeking a career at one of the preeminent biomedical research institutes in the nation and the world? The Deputy Director for Intramural Research (DDIR), NIH, serves as the principal advisor on all matters relating to the quality and effectiveness of NIH intramural science and research policies, and participate in the overall planning for the research programs of the NIH.

We are looking for applicants with a commitment to scientific excellence and the energy, enthusiasm, and innovative thinking necessary to lead within a dynamic and diverse organization. Applicants must possess a Ph.D. and/or an M.D. and have demonstrated senior-level research experience and knowledge in the broad field of biomedical research and extensive experience in the management and administration of programs of research. For the full list of qualifications, please see the announcement located at: **http://www.jobs.nih.gov/ vacancies/executive.htm**.

The successful candidate for this position will be appointed at a salary commensurate with experience and accomplishments, and full Federal benefits, including leave, health and life insurance, retirement and savings plan (401K equivalent) will be provided. A recruitment or relocation bonus may be available, and relocation expenses may be paid. In accordance with Executive Order 14043, Federal employees are required to be fully vaccinated against COVID-19.

If you are ready for an exciting leadership opportunity, please see the detailed vacancy announcement at **http://www.jobs.nih.gov/vacancies/executive.htm**

Applications are due by 11:59 p.m. (EST), **February 1, 2022**. Please contact Mariela Light at **SeniorRe@od.nih.gov** for questions and/or additional information.

DHHS, and NIH ARE EQUAL OPPORTUNITY EMPLOYERS

TENURE TRACK FACULTY POSITIONS

The Department of Biochemistry and Molecular Biology in the Medical College of Georgia (Georgia's only public academic medical center) at Augusta University is expanding its research programs and is seeking tenure-track faculty at the rank of **Assistant Professor, Associate Professor or Professor**. Successful applicants will have research interests in the areas of benign, malignant and age-related diseases, molecular signaling, inflammation/immunology, gene/epigenetic regulation, biomarkers, and targeted-therapy or in novel areas with translational research appeal. The successful candidate will contribute to the department's robust extramural grant portfolio, as well as, contribute to the department's education mission and graduate and post-graduate training programs. The candidate will have access to departmental and institutional state-of-the-art cores and equipment, biotelemetry, genome editing of mice/rats, next generation and 10x single cell sequencing, bioinformatics, and advanced microscopy/imaging, flow-cytometry, mass spectrometry, Seahorse XFEs, small-animal imaging and MRI. Candidates with established extramural research programs are desirable. Candidates applying at the Associate Professor or Professor level should have extramurally funded research programs. Highly competitive salaries and excellent start-up packages are available. With ambient weather, affordable living, and the home of the US Cyber Command, and the Masters Golf Tournament, Augusta, Georgia and the greater CSRA is a fast-growing community, with nature walks along the Savannah River, a ballpark across the river, convenient commute, and some of the best schools in Georgia.

Minimum Requirement
Doctoral degree (PhD or MD/PhD in Life Sciences, or equivalent) and a minimum of 3 years of postdoctoral experience

Shift /Salary
Academic rank, tenure status and salary to be commensurate with qualifications and experience of the candidate.

How to Apply
To be considered an applicant for this position you must apply online at Augusta University Faculty Careers (**http://www.augusta.edu/hr/jobs/faculty**) **Job ID # 235527**. Please upload your Curriculum vitae, research statement, professional references, etc. as one document.

Augusta University is an Equal Opportunity and Equal Access Institution

TENURE-TRACK ASSISTANT PROFESSOR EVOLUTIONARY BIOLOGY

Position Description Summary
The Department of Biological Sciences at The George Washington University (GWU) invites applicants for a tenure-track faculty position in Evolutionary Biology at the rank of **Assistant Professor**, to begin in the Fall 2022 Semester. The successful candidate will complement existing strengths in integrative, interdisciplinary research in the department, and will be expected to establish and maintain an externally funded research program comprising graduate and undergraduate students. Teaching duties involve a core undergraduate course in Evolution as well as an upper-division course in an area related to the successful candidate's specialty. We particularly welcome candidates from groups that are historically underrepresented in biology and/or have demonstrated leadership toward building an equitable and inclusive scholarly environment.

To be considered, please complete an online faculty application (**https://www.gwu.jobs/postings/88591**). Review of applications will begin on **January 10th, 2022**, and continue until the position is filled, pending final budgetary approval. Employment offers are contingent on the satisfactory outcome of a standard background screening.

The university is an Equal Employment Opportunity/Affirmative Action employer that does not unlawfully discriminate in any of its programs or activities on the basis of race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity or expression, or on any other basis prohibited by applicable law.

POSTDOCTORAL FELLOW POSITIONS

EDWARD A. DOISY DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
SAINT LOUIS UNIVERSITY SCHOOL OF MEDICINE

Saint Louis University, a Catholic Jesuit institution dedicated to education, research, health care, and service, is seeking outstanding applicants for **Postdoctoral Fellow positions** in structural enzymology of coagulation factors in the lab of Dr. Enrico Di Cera (**https://biochem.slu.edu/faculty/dicerawp/**) in the Department of Biochemistry and Molecular Biology (**https://biochem.slu.edu/**). Documented experience in biophysical methods (rapid kinetics, smFRET) and/or other structural biology techniques (X-ray, NMR, cryo-EM) is required.

Please submit a cover letter, curriculum vitae, and addresses of three references to **https://slu.wd5.myworkdayjobs.com/Careers**, Job #**2020-01666**.

Saint Louis University is an equal opportunity/affirmative action employer. All qualified candidates will receive consideration for the position applied for without regard to race, color, religion, sex, age, national origin, disability, marital status, sexual orientation, military/veteran status, gender identity, or other non-merit factors. We welcome and encourage applications from minorities, women, protected veterans, and individuals with disabilities (including disabled veterans). If accommodations are needed for completing the application and/or with the interviewing process, please contact Human Resources at 314-977-5847.

The 2022 Louisa Gross Horwitz Prize for Biology or Biochemistry

The Louisa Gross Horwitz Prize was established under the will of the late S. Gross Horwitz through a bequest to Columbia University and is named to honor the donor's mother. Louisa Gross Horwitz was the daughter of Dr. Samuel David Gross (1805–1889), a prominent surgeon of Philadelphia and author of the outstanding *Systems of Surgery* who served as president of the American Medical Association.

Each year since its inception in 1967, the Louisa Gross Horwitz Prize has been awarded by Columbia University for outstanding basic research in the fields of biology or biochemistry. The purpose of this award is to honor a scientific investigator or group of investigators whose contributions to knowledge in either of these fields are deemed worthy of special recognition.

The prize consists of an honorarium and a citation, which are awarded at a special presentation event. Unless otherwise recommended by the Prize Committee, the prize is awarded annually. Katalin Karikó, PhD, University of Pennsylvania, and Drew Weissman, MD, PhD, University of Pennsylvania, are the 2021 awardees.

QUALIFICATIONS FOR THE AWARD

The purpose of this prize is to reward scientists that have made recently transformative discoveries not yet recognized by high-visibility international awards. The Prize Committee recognizes no geographical limitations. The prize may be awarded to an individual or a group. When the prize is awarded to a group, the honorarium will be divided among the recipients, but each member will receive a citation.

NOMINATIONS

All materials must be written in the English language and submitted electronically at:
http://www.cumc.columbia.edu/research/horwitz-prize

Deadline date: January 26, 2022, 5:00 p.m. (EST)
Renominations are by invitation only.
Self-nominations are not permitted.

NOMINATIONS SHOULD INCLUDE:

1. A summary of the research on which this nomination is based (no more than 500 words).
2. A summary of the significance of this research in the fields of biology or biochemistry (no more than 500 words).
3. A brief biographical sketch of the nominee, including positions held and awards received by the nominee.
4. A copy of the nominee's curriculum vitae.
5. A key publication list of up to ten of the nominee's most significant publications relating to the research noted under item 1.

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

BOSTON COLLEGE
SCHILLER INSTITUTE FOR
INTEGRATED SCIENCE AND SOCIETY

SENIOR ENERGY AND CLIMATE CHANGE FACULTY POSITIONS

At Boston College, the newly established Schiller Institute for Integrated Science and Society enhances multi-disciplinary, collaborative research and pedagogies that address critical societal issues in the areas of energy, health and the environment.

As the start of a multi-year faculty hiring initiative, the Institute seeks four faculty at the rank of **Associate** or **Full Professor** in two broad areas: a) climate change science, mitigation and adaptation, and b) the energy transition. Three faculty positions are grounded in the natural sciences (physics, chemistry, earth and environmental sciences, biology, materials science, etc.); a fourth is grounded in the social sciences (economics, political science, sociology, etc.).

The Schiller Institute stands at the crossroads of the Boston College community. Now in its nascent and dynamic growth phase, the Institute seeks four visionary scholar-leaders for its unique and interdisciplinary inaugural cohort. These faculty will strengthen the research profile of the university; stimulate collaborative, problem-focused research; develop, design, and participate in interdisciplinary curricula and co-curricular activities; build partnerships with leading institutions and organizations; and advance the results of BC applied research into the public sphere.

In the area of climate change impact, mitigation and adaptation, the Institute seeks faculty with well-developed research programs in areas such as climate science, earth system modeling, socio-environmental systems modeling, climate justice, adaptation through emerging technologies or nature-based solutions, societal determinants and consequences of climate change, and impact-based analysis and forecasting. In the area of energy transition, the Institute seeks faculty with significant research programs in science and technologies for renewable energy harvesting, conversion, and utilization, energy storage, and improving energy efficiency; energy poverty; or the societal transition to cleaner/more efficient technologies. Individual faculty and interdisciplinary research teams are invited to apply.

The Schiller Institute is committed to advancing Boston College's commitments to equity, inclusion, and belonging. Our excellence can only be fully realized by faculty, students, and staff who share our commitment to these values. Successful candidates for our faculty positions will demonstrate a commitment to advancing equity and inclusion in their research, teaching and service.

Applications should include a letter describing the applicant's interest in the position and relevant experience, a curriculum vitae, and a 2-3-page research statement. Applications should be submitted through Interfolio: **http://apply.interfolio.com/99440**

Nominations and inquiries can be made to Professor Laura J. Steinberg, Seidner Family Executive Director of the Schiller Institute for Integrated Science and Society, at **schiller.institute@bc.edu**.

Review of applications will begin on **January 15, 2022** and will continue until the positions are filled.

Equal Employment Opportunity Statement: Boston College is an affirmative action, equal opportunity employer. In concert with our Jesuit, Catholic mission, Boston College is dedicated to the goal of building a culturally diverse and pluralistic faculty and staff committed to teaching and working in a multicultural environment and strongly encourages applications for women, minorities, individuals with disabilities, and covered veterans.

ScienceCareers online @sciencecareers.org

By **Angela Hessler**

A path to independence

Six paragraphs into the corporate newsletter, I learned that my job at a major energy company, like those of more than 100 other earth scientists, was moving halfway across the country. It would be a win for us, we were told—a chance to boost our careers by interacting with new teams. I had just come off maternity leave and didn't see it that way. Our young family loved where we lived, and after 5 years at the company I had settled into a productive work routine. Plus, my job had solved our family's "two body problem." I knew I'd need to find a new path.

Before moving to industry, I'd held an assistant professor position for 4 years. My teaching load had left little time for research, and the faculty job was far from my spouse's. Working for the company offered a solution. As a younger geologist I had not dreamed about a career in industry, but I enjoyed getting back to research, tackling applied problems, and interacting with my industry colleagues.

I realized I should not get too comfortable, however. Not long after my start date, it became clear that the company expected us to demonstrate a warriorlike enthusiasm for change. One manager recommended I write in my annual review that I wanted to transfer to other offices every 2 years, even though I did not want that at all. So the relocation announcement wasn't a complete surprise.

"Could I keep doing research, I started to wonder, on my own?"

Once I'd made the decision to quit, I found myself at a fork in the road. As I saw it, I could apply for the few geology jobs in the area, none of which was a good fit—or I could leave my scientific career behind. Both options made me sad. I loved geology, and I was trained in something I knew was of value. On the other hand, I didn't want to take a job that didn't interest me. As I stared at these two paths, up on my periphery came a third. Could I keep doing research, I started to wonder, on my own?

I knew of groups that received grants to carry out research outside academia. I contemplated whether I could compete for some of that money, too—a path that would enable me to continue to do what I love without any pressure to move.

I decided to give it a try. But before I could apply for funding, I knew I had to start publishing again, something I'd largely set aside during my time in industry. I started with a low-hanging fruit, getting an old data set ready for publication. I soon ran up against the stigma of being "unaffiliated." During the submission process, my generic email address was rejected. I set up a nonprofit organization and resubmitted using a ".org" email address—that time without issue.

After the paper came out, one of my former industry colleagues pointed me to a government program that invests energy revenue into research, which is how I got my first grant. Soon I was back to the hands-on study of rocks, examining them under a microscope I'd borrowed from the state geological bureau and sending samples to a commercial lab for geochemical analysis. For the first time in a long while, I was being paid to answer my own questions.

Setting out as an independent researcher was tough at first, with no guarantee of income. It has also been challenging to navigate the grant proposal requirements of many private foundations and government entities, which can exclude small nonprofits like mine. But over the years, I've been able to grow my research portfolio and continue publishing. I'm lucky that in my field, I can outsource some lab work and otherwise carry out research without institutional resources. I've also been fortunate to collaborate with researchers in industry and academia who shared funding and insights.

On the home front, the flexibility of my position has had its benefits. Our family relocated, on our own terms, taking my work with us. And with three school-age kids, I have appreciated the freedom to scale back my work during periods of sickness or during summer vacations.

The disruption at my industry job, unwelcome at first, propelled me in a new direction that has turned out to be the most creative period of my career. There is always uncertainty, and I might never declare victory, but at least I am not running (and losing) someone else's race. ■

Angela Hessler is the founder of the Deep Time Institute.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

AAAS.ORG/COMMUNITY

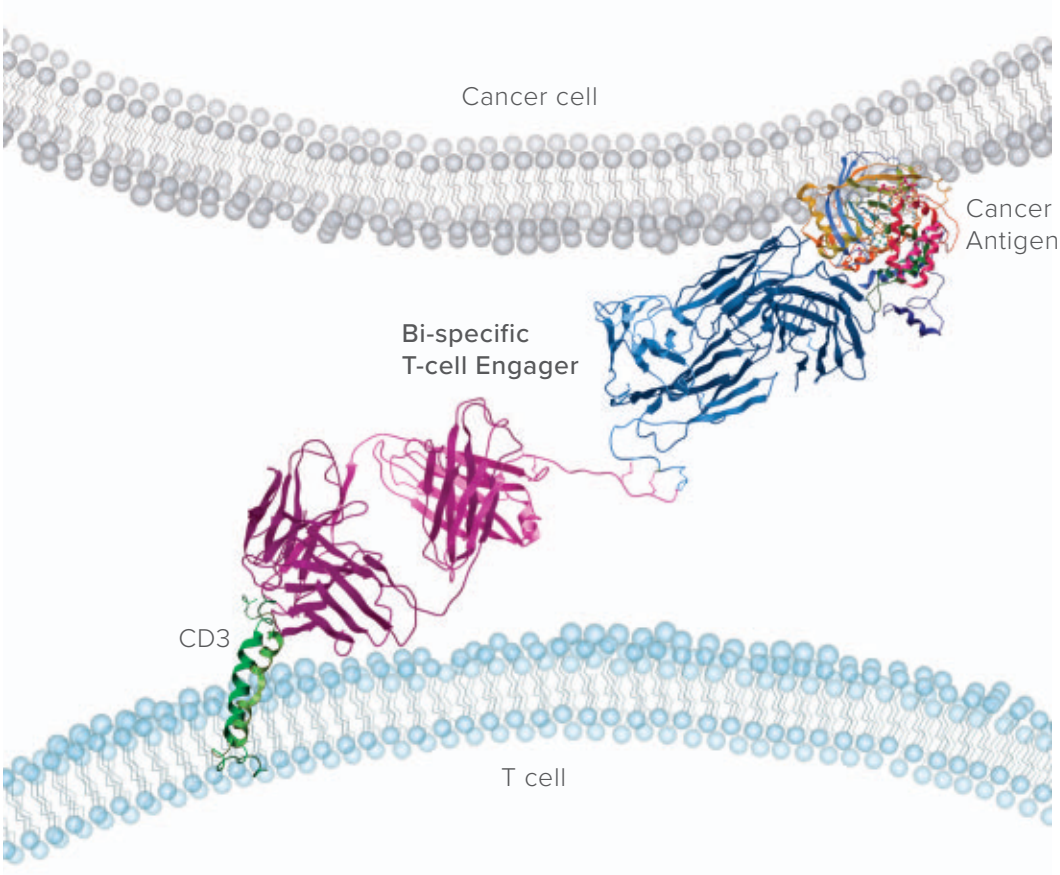
AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Bispecific Antibody Development

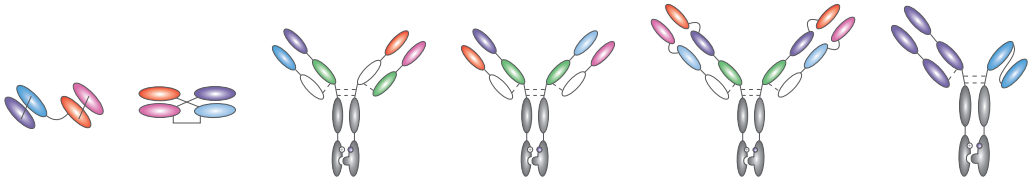
Custom Bispecific Antibodies for Research Use



Learn how ProMab's services can help you establish a proof-of-concept for your bispecific antibody and accelerate your antibody therapeutics research.

Personalized Antibody Solutions

- Bispecific Antibody Design
- Bispecific Antibody Development
- Bispecific Antibody Validation
- Monoclonal Antibody Development
- Antibody Humanization
- Antibody Producing Stable Cell Line



Stylized overview of the numerous designs used to generate bispecific antibodies.

All products are for research only

Discover more | www.promab.com

ProMab Biotechnologies, Inc.
2600 Hilltop Dr, Building B, Suite C320, Richmond, CA 94806
1.866.339.0871 | info@promab.com

SCAN ME